STATISTICAL ANALYSIS PLAN	
Study Code	D5272C00001 (Legacy #3151-201-008)
Edition Number	3
Date	13-Dec-2022

A 54-Week, Multicenter, Randomized, Double-blind, Placebo-controlled, Parallel-group Phase 2 Study to Assess the Efficacy and Safety of Brazikumab in Participants with Moderately to Severely Active Ulcerative Colitis (Expedition Lead-in)

Redacted For Public Disclosure

TABLE OF CONTENTS

TITLE PAGE 1
TABLE OF CONTENTS 2
TABLE OF TABLES 6
LIST OF ABBREVIATIONS 7
AMENDMENT HISTORY 10
1 INTRODUCTION 14
2 CHANGES TO PROTOCOL PLANNED ANALYSES 16
3 DATA ANALYSIS CONSIDERATIONS 16
3.1 Timing of Analyses 16
3.2 Analysis Populations 17
3.3 General Considerations 18
3.3.1 General Study Level Definitions 19
3.3.1.1 Study Day 1 19
3.3.1.2 Baseline 19
3.3.1.3 Study Periods 19
3.3.2 Visit Window 20
3.3.2.1 Study day 20
3.3.2.2 Conventions for the visit window 20
3.3.2.3 Early discontinuation visits 21
3.3.2.4 Multiple observations 21
3.3.3 Handling of Unscheduled Visits 22
3.3.4 Multiplicity/Multiple Comparisons 22
3.3.5 Handling of Protocol Deviations in Study Analysis 23
4 STATISTICAL ANALYSIS 23
4.1 Study Population 23
4.1.1 Participant Disposition and Completion Status 23
4.1.1.1 Definitions and Derivations 23
4.1.1.2 Presentation 24
4.1.2 Analysis Sets 25
4.1.2.1 Definitions and Derivations 25
4.1.2.2 Presentation 26
4.1.3 Protocol Deviations 26
4.1.3.1 Definitions and Derivations 26
4.1.3.2 Presentation 26
4.1.4 Demographics 27
4.1.4.1 Definitions and Derivations 27
4.1.4.2 Presentation 27
4.1.5 Baseline Characteristics 27

4.1.5.1 Definitions and Derivations 27
4.1.5.2 Presentation 28
4.1.6 Disease Characteristics 28
4.1.6.1 Definitions and Derivations 28
4.1.6.2 Presentation 29
4.1.7 Medical History and Concomitant Disease 30
4.1.7.1 Definitions and Derivations 30
4.1.7.2 Presentation 30
4.1.8 Prior and Concomitant Medications 30
4.1.8.1 Definitions and Derivations 30
4.1.8.2 Presentation 32
4.1.9 Study Intervention Compliance 33
4.1.9.1 Definitions and Derivations 33
4.1.9.2 Presentation 33
4.2 Endpoint Analyses 34
4.2.1 Primary Endpoint 40
4.2.1.1 Definition 40
4.2.1.2 Derivations 40
4.2.1.3 Handling of Dropouts and Missing Data 42
4.2.1.4 Primary Analysis of Primary Endpoint 42
4.2.1.5 Sensitivity Analyses of the Primary Endpoint 44
4.2.1.6 Supplementary Analyses of the Primary Endpoint 45
4.2.1.7 Subgroup Analyses 45
4.2.2 Key Secondary Endpoint 46
4.2.3 Secondary Endpoints: Binary Endpoints 46
4.2.3.1 Definition 46
4.2.3.2 Derivations 47
4.2.3.3 Handling of Dropouts and Missing Data 47
4.2.3.4 Primary Analysis of Secondary Binary Endpoints 47
4.2.3.5 Supplemental Analyses of the Secondary Binary Endpoints 47
4.2.3.6 Subgroup Analyses 48
4.2.4 Secondary Endpoint: exposure-response model 48
4.2.5 Exploratory Endpoint: IL-22 cutoff determination 48
4.2.5.1 Definition 48
4.2.5.2 Derivations 48
4.2.5.3 Handling of Dropouts and Missing Data 48
4.2.5.4 Primary Analysis of Exploratory Endpoint 49
4.2.6 Exploratory Endpoint: Binary Endpoints 49
4.2.6.1 Definition 49
4.2.6.2 Derivation 50
4.2.6.3 Handling of Dropouts and Missing Data 50
4.2.6.4 Primary Analysis of Exploratory Endpoints 50
4.2.6.5 Supplementary analysis 50
4.2.7 Exploratory Endpoint: Change from Baseline in mMS 51

4.2.7.1 Definition 51
4.2.7.2 Derivation 51
4.2.7.3 Handling of Dropouts and Missing Data 51
4.2.7.4 Primary Analysis of Exploratory Endpoints 51
4.2.8 Exploratory Endpoint: Change from Baseline in CS use 52
4.2.8.1 Definition 52
4.2.8.2 Derivation 52
4.2.8.3 Handling of Dropouts and Missing Data 52
4.2.8.4 Primary Analysis of Exploratory Endpoints 52
4.2.9 Exploratory Endpoint: Mucosal Inflammation through Week 54 52
4.2.9.1 Definition 52
4.2.9.2 Derivation 54
4.2.9.3 Handling of Dropouts and Missing Data 54
4.2.9.4 Primary Analysis of Exploratory Endpoints 54
4.2.9.5 Supplementary analysis 56
4.2.9.6 Subgroup analysis 56
4.2.10 Exploratory Endpoints: Change in Signs and Symptoms (BMD, Evening Diary, FACIT-F) 56
4.2.10.1 Definition 56
4.2.10.2 Derivation 57
4.2.10.3 Handling of Dropouts and Missing Data 58
4.2.10.4 Primary Analysis of Exploratory Endpoints 59
4.2.11 Exploratory Endpoints: Health Related Quality of Life (HRQoL) (IBDQ, SF-36, EQ-5D-5L) 59
4.2.11.1 Definition 59
4.2.11.2 Derivation 60
4.2.11.3 Handling of Dropouts and Missing Data 61
4.2.11.4 Primary Analysis of Exploratory Endpoints 61
4.2.11.5 Subgroup analysis 61
4.3 Pharmacodynamic Endpoints 62
4.3.1 Definition 62
4.3.2 Handling of Dropouts and Missing Data 62
4.3.3 Primary Analysis of Exploratory Endpoints 62
4.3.3.1 Presentation 62
4.3.4 Subgroup analysis 62
4.4 Pharmacokinetics 62
4.5 Immunogenicity 63
4.5.1 Derivations 63
4.5.2 Primary analysis of ADA 64
4.6 Safety Analyses 64
4.6.1 Exposure 65
4.6.1.1 Definitions and Derivations 65
4.6.1.2 Presentation 65
4.6.2 Adverse Events 65

4.6.2.1 Definitions and Derivations 65
4.6.2.2 Presentation 66
4.6.3 Clinical Laboratory, Blood Sample 67
4.6.3.1 Definitions and Derivations 67
4.6.3.2 Presentations 71
4.6.4 Clinical Laboratory, Urinalysis 71
4.6.4.1 Definitions and Derivations 71
4.6.4.2 Presentations 72
4.6.5 Other Laboratory Evaluations 72
4.6.5.1 Definitions and Derivations 72
4.6.5.2 Presentations 72
4.6.6 Physical Examinations 72
4.6.6.1 Definitions and Derivations 72
4.6.7 Vital Signs 72
4.6.7.1 Definitions and Derivations 72
4.6.7.2 Presentations 73
4.6.8 Electrocardiogram 74
4.6.8.1 Definitions and Derivations 74
4.6.8.2 Presentations 74
4.6.9 Other Safety Assessments 75
4.6.9.1 Definitions and Derivations 75
4.6.9.2 Presentations 75
5 INTERIM/PRIMARY ANALYSIS 75
6 REFERENCES 76
7 APPENDIX 78
7.1 Modified Mayo Score 78
7.2 Bowel Movement Diary 79
7.3 Evening eDiary 80
7.4 FACIT-F Scale 81
7.5 Missing start and stop dates 83
7.5.1 Incomplete Start Date 83
7.5.2 Incomplete Stop Date 83
7.6 Differential effect method 85

TABLE OF TABLES

Table 3-1 Populations for Analysis 17
Table 3-2 Pooling strategy for pre-Amendment 5 treatment groups 18
Table 3-3 Analysis Windows 20
Table 4-1 Disease characteristics to be summarized at Baseline 28
Table 4-2 Drug names and ATC codes 29
Table 4-3 Objectives and endpoints 35
Table 4-4 Components of clinical remission calculation 41
Table 4-5 GHI grading and conversion to RHI (Geboes, et al, 2000). 52
Table 4-6 RHI calculation 53
Table 4-7 IBDQ subscores 60
Table 4-8 Protocol-Required Safety Laboratory Assessments 68
Table 4-9 Normal ranges for laboratory values (Central Laboratory Services Manual) 68
Table 4-10 Criteria for Potentially Clinically Significant Laboratory Values 70
Table 4-11 Other laboratory testing 72
Table 4-12 Criteria for Potentially Clinically Significant Vital Signs 73
Table 4-13 Criteria for Potentially Clinically Significant Electrocardiograms 74
Table 7-1 Items Included in mMS 78
Table 7-2 Bowel movement diary 79
Table 7-3 Daily recall items 80

LIST OF ABBREVIATIONS

Abbreviation or Specialized Term	Definition
AE	Adverse event
AESI	Adverse Event of Special Interest
ALP	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AP	Abdominal Pain
AST	Aspartate Aminotransferase
BM	Biomarker
BMD	Bowel Movement Diary
BMI	Body mass index
BSFS	Bristol Stool Form Scale
CDL	Clinical Database Lock
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CRF	Case Report Form
CRP	C-Reactive Protein
CS	Corticosteroid
CSP	Clinical Study Protocol
CSR	Clinical Study Report
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
ED	Early discontinuation of study intervention
EIM	Extra-Intestinal Manifestation
EQ-5D-5L	5-level European Quality of Life - 5 Dimensions
ET	Early termination
FACIT-F	Functional Assessment of Chronic Illness Therapy – Fatigue Scale
FAS	Full Analysis Set
FCP	Fecal Calprotectin
GHI	Geboe's Histological Index
HRQoL	Health-Related Quality of Life
IBD	Inflammatory Bowel Disease
IBDQ	Inflammatory Bowel Disease Questionnaire

ICF	Informed Consent Form
ICH	International Council for Harmonization
IL	Interleukin
IP	Investigational Product
IPD	Important Protocol Deviation
ITT	Intent-to-Treat
IV	Intravenous
JAK	Janus kinase
LLN	Lower Limit of Normal
LOCF	Last Observation Carried Forward
LS	Least Squares
LSF	Loose Stool Frequency
LSLV	Last Subject Last Visit
MAR	Missing at Random
Max	Maximum
MCS	Mental Component Score
MedDRA	Medical Dictionary for Regulatory Activities
Min	Minimum
MMRM	Mixed Models for Repeated Measures
mMS	modified Mayo Score
MPP	Maintenance Period Population
MS	Mayo score
NRI	Non-responder imputation
NRS	Numerical Rating Scale
NSAID	Non-steroidal anti-inflammatory drug
PCS	Potentially Clinically Significant
PD	Pharmacodynamic
PF	Physical Functioning scale
PGA	Physician's Global Assessment
PH	Physical Health scale
PK	Pharmacokinetics
PRO	Patient Reported Outcome
PT	Preferred Term
Q1	1st Quartile
Q3	3rd Quartile

RHI	Robart's Histopathology Index
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SAS	Statistical Analysis System
SC	Subcutaneous
SD	Standard Deviation
SE	Standard Error
SF-36	Short-Form 36 Health Survey
SI	International System of Units
SOC	System Organ Class
TBL	Total Bilirubin
TESAE	Treatment Emergent Serious Adverse Event
TNF	Tumor Necrosis Factor
UC	Ulcerative Colitis
ULN	Upper Limit of Normal
VAS	Visual Analog Scale
WHO	World Health Organization

AMENDMENT HISTORY

CATEGORY Change refers to:	*Date*	*Description of change*	*In line with CSP?*	*Rationale*
N/A		Initial approved SAP	N/A	N/A
Entire document	9 Nov 2020	Original SAP as transferred from Allergan with change of sponsor to AstraZeneca and change of study number to AstraZeneca study number.	N/A	N/A
Entire document	18 Jan 2022 (Version 2.0)	Updated SAP to change to AstraZeneca standard format and to reflect Protocol Version 6 Amendment 5 changes	Yes, Version 6	N/A
Entire document	13 Dec 2022 (Version 3.0)	Updated SAP to align with Protocol Version 7 Amendment 6 changes	Yes, Version 7	To align with latest protocol
Entire document	13 Dec 2022	Minor editorial changes	N/A	Revisions made for clarity
Entire document	13 Dec 2022	Renamed “adapted mMS” to “mMS”	Yes, Version 7	To align with the new FDA guidance and CSP
Entire document	13 Dec 2022	Remove presentation by BM+/BM- for all study population and safety tables	Yes, Version 7	Aligns with the CSP. Only efficacy is to be explored by BM+/BM-
Entire document	13 Dec 2022	Remove presentations for the MPP by BM+ and BM-	Yes, Version 7	This would be difficult to interpret due to the limited number of participants expected in each group
Entire document	13 Dec 2022	Updated wording of PROs (Stool Frequency and Rectal Bleeding)	Yes, Version 7	Further clarity on PRO score calculation
Entire document	13 Dec 2022	Added “brazikumab total” to the description of all study population and safety tables.	N/A	Revisions made to align with AZ standard TLFs
Primary endpoint	13 Dec 2022	Added Week 10 endoscopy window of 14 days and endoscopy no more than 7 days after Week 10 IP dose	Yes, Version 7	14-day window increases the number of analysable endoscopies at Week 10 and KSE input confirmed no clinically significant difference seen with a 14-day window
Primary endpoint	13 Dec 2022	Added supplementary analysis for clinical response defined as Reduction in total MS $\geq$ 3 points from Baseline and $\geq$ 30% decrease from Baseline and a decrease in the rectal bleeding score $\geq$ 1 point from Baseline or a score of 0 or 1	Yes, Version 7	Added a supplementary definition of clinical response to explore the impact of different definitions of the primary endpoint.

CATEGORY Change refers to:	***Date***	***Description of change***	***In line with CSP?***	***Rationale***
Derivation of primary or secondary endpoint	13 Dec 2022	Added a sentence to cover patients who were randomized who never received study drug	Yes, Version 7	Confirming those who were randomized and never received study drug in legacy CSP will not be included in the FAS population, This aligns with decisions made and documented prior to the Inform data lock
Derivation of primary or secondary endpoint	13 Dec 2022	Clarified that NRI for intercurrent events will be calculated before NRI for missing data; minor wording changes.	N/A	Added for clarity
Derivation of primary or secondary endpoint	13 Dec 2022	Changed the missing data algorithm for PRO components from a sliding window to an extended window	Yes, Version 7	An extended window will keep the PRO diary days closer to the endoscopy day
Derivation of primary or secondary endpoint	13 Dec 2022	The Maintenance Period Population (MPP) has been updated such that data for participants will be summarized according to the randomized intervention	Yes (aligned with FAS), Version 7	Corrected the wording to be consistent with the FAS, and to preserve the initial randomization for all efficacy analysis sets
Derivation of primary or secondary endpoints	13 Dec 2022	Change CS-free remission for the FAS to the subpopulation with CS at baseline	Yes, Version 7	Revisions made for clarity
Data Presentations	13 Dec 2022	Added all timepoints in presentations for changes from Baseline in signs and symptoms	Yes, Version 7	Revisions made for clarity
Data Presentations	13 Dec 2022	Changed reporting of responder endpoints (removed estimated response rate and 95% confidence interval for each treatment)	Yes, Version 7	Updated to align with shells
Other	13 Dec 2022	Definition of “Completer” simplified and aligned to modified Study Periods	N/A	Revisions made for clarity
Other	13 Dec 2022	Added table of ATC codes for medications	N/A	Added for clarity
Other	13 Dec 2022	Clarify that all time points included in analysis and NRI due to missing data is included as a separate category	N/A	Added for clarity
Other	13 Dec 2022	Added language that a ‘BY’ statement will be used in SAS to create the subgroups; that the subgroup variable will not be included as a covariate; added reference to forest plot	N/A	Added for clarity

CATEGORY Change refers to:	*Date*	*Description of change*	*In line with CSP?*	*Rationale*
Other	13 Dec 2022	Removed reference to BM+/BM- analysis of secondary endpoints	Yes, Version 7	To align with the exploratory objectives in the CSP
Other	13 Dec 2022	Add exploratory endpoint for clinical remission at Week 54	Yes, Version 7	This data is collected at Week 54. This addition is to ensure intended use is appropriately reflected in the study objectives
Other	13 Dec 2022	Add symptomatic remission at Week 10 and at Week 54 (rectal bleeding = 0 and stool frequency= 0 or 1, and at least a 1-point decrease from baseline)	Yes, Version 7	This data is collected at Week 10 and Week 54. This addition is to ensure intended use is appropriately reflected in the study objectives
Other	13 Dec 2022	Add exploratory endpoint for endoscopic improvement at Week 54	Yes, Version 7	This data is collected at Week 54. This addition is to ensure intended use is appropriately reflected in the study objectives
Other	13 Dec 2022	Add exploratory endpoint for endoscopic remission at Week 54 (centrally read endoscopy subscore of 0)	Yes, Version 7	This data is collected at Week 54. This addition is to ensure intended use is appropriately reflected in the study objectives
Other	13 Dec 2022	Remove exploratory clinical response at Week 10 and 54 by BM	Yes, Version 7	Aligns with CSP
Other	13 Dec 2022	Remove PGA defined for clinical response	N/A	Aligns with CSP and latest FDA guidance
Other	13 Dec 2022	Remove Maintenance treatment group as a covariate for the Week 54 analysis	Yes, Version 7	Revisions made for clarity
Other	13 Dec 2022	Added scoring details for the FACIT-F	Yes, Version 7	Revisions made for clarity
Other	13 Dec 2022	Clarified that EQ-5D-5L VAS, 0 is worst health and 100 is best; SF-36 has two subscores instead of a total score	Yes, Version 7	Revisions made for clarity
Other	13 Dec 2022	Add MMRM analysis of PD endpoints	Yes, Version 7	To align with the presentation in the TLFs
Other	13 Dec 2022	Add MMRM subgroup analysis by BM status	Yes, Version 7	To align with the presentation in the TLFs

CATEGORY Change refers to:	***Date***	***Description of change***	***In line with CSP?***	***Rationale***
Other	13 Dec 2022	Remove "A listing of serum concentration data will be presented by treatment arm. The actual time, deviation and percent deviation from nominal time will also be listed. Cmax and tmax will be determined from serum concentration data. The results will be listed and summarized using descriptive statistics."	Yes, Version 7	To align with CSP
Other	13 Dec 2022	Specified 18 weeks for maintenance period tables; replace "event rate "with "adverse events per 100 subject years"	Yes, Version 7	To align with AZ standards
Other	13 Dec 2022	Specified that >5% criteria does not apply to the 2100mg group	Yes, Version 7	Revision made for clarity
Other	13 Dec 2022	Updated definition of potential Hy's law	Yes, Version 7	To align with AZ standards
Other	13 Dec 2022	Patient global impression of change for ulcerative colitis (PGIC-UC) and Physician global impression of severity for ulcerative colitis (PGIS-UC) were removed from the exploratory endpoints	Yes, Version 7	PGIC-UC and PGIS-UC will be used as anchors in the PRO analyses (described in the PRO Analysis Plan and reported outside of the CSR)
Other	13 Dec 2022	Updated analysis for CS use to display descriptive statistics only	Yes, Version 7	Previous planned analysis not aligned with the definition of the endpoint
Other	13 Dec 2022	Updated handling of missing information for AE possibly related to study drug	Yes, Version 7	Revision made for clarity

1 INTRODUCTION

Study D5272C00001 (Legacy # 3151-201-008) is a Phase 2, global, multicenter, randomized, double-blind, placebo-controlled, parallel-group study in participants 18 to 80 years of age (inclusive) who have been diagnosed with ulcerative colitis (UC). The severity of UC will be assessed based on the modified Mayo Score (mMS). Participants with moderately to severely active UC will be included in this study (defined as both stool frequency and rectal bleeding subscores ≥ 1, and modified mayo endoscopic subscore of ≥ 2 based on full colonoscopy within 14 days prior to randomization).

This Statistical Analysis Plan (SAP) provides a technical and detailed elaboration of the statistical analyses of the efficacy and safety data as outlined and/or specified in the most recent Clinical Study Protocol (CSP) (Amendment 6 version 7, dated 10 March 2022). Specifications of tables, figures, and data listings are contained in a separate document. The advanced analysis of pharmacokinetic data may be specified in a separate analysis plan and reported outside of the Clinical Study Report (CSR).

There will be a single interim analysis, also known as the primary analysis. The interim (primary) analyses will be performed after the last participant completes Week 10 or has discontinued early. Thus, Week 10 is the primary timepoint in the study as opposed to the final timepoint at Week 54. Accordingly, the pre-specified primary analyses fulfil the ICH E9 definition of an interim analysis. The study will be unblinded for the interim (primary) analysis, and the results from the primary analysis will be part of the final results in the CSR. AstraZeneca personnel who are directly involved with the conduct of the study, study site personnel, and participants will remain blinded to the treatment assignment for individual participants until the completion of the study. Therefore, there will be 2 Clinical Database Locks (CDL) in the study, at the time of the interim (primary) analyses when all participants complete Week 10 (or discontinue early in the Induction Period) and after the last participant completes the second safety follow-up visit, Last Subject Last Visit (LSLV) or discontinues early.

The interim (primary) analyses based on Week 10 data and final analysis based on Week 54 data will be described in this SAP. All available data from participants who have completed Week 54 or would have completed Week 54 if they hadn't discontinued, will be analysed in the interim (primary) analysis. A Study Integrity Plan has been developed to describe the strategy for protecting the integrity of the study until final database lock, for example by maintaining confidentiality of the unblinded interim (primary) data until final CDL of Study D5272C00001 at Week 54, and document study roles and access level to both aggregated and participant individual level data.

Approximately 255 participants (75 per treatment group (225) plus those in treatment groups prior to CSP Amendment 5 that have been removed and will not be included in the efficacy analysis) are planned for enrolment (as per CSP Amendment 5) in a 1:1:1 ratio to receive

study intervention (720 mg brazikumab IV, 1440 mg brazikumab IV, or placebo IV) during the Induction Period (on Week 0 (Day 1), Week 2, and Week 6). Participants will be stratified for randomization by prior treatment experience, biological-naïve vs prior biological use (failed or were intolerant to prior biological treatment) ("prior biologic use"). Following the Induction Period, participants receiving brazikumab will be re-randomized 1:1 (stratified according to prior biologic use and by induction dose) at Week 10 to receive either 120 mg or 240 mg brazikumab SC every 4 weeks. Participants randomized into the Induction Period placebo treatment arm will continue to receive placebo SC every 4 weeks in the Maintenance Period.

Following a screening period of up to 5 weeks, the length of this study will be up to 68 weeks, consisting of a 54-week Double-Blind Treatment Period (10-week Induction Period and 44-week Maintenance Period combined), and an 18-week Safety Follow-up Period in which no study intervention will be administered. The 18-week Safety Follow-up Period starts after the last subcutaneous administration of brazikumab or placebo in the Maintenance Period (at Week 50) (4 weeks before the end of the Maintenance Period). Therefore, there is an overlap of 4 weeks between the 54-week Double-Blind Treatment Period and 18-week Safety Follow-up Period. For participants who qualify and enrol into the open-label extension study, the Safety Follow-up Period will not be applicable.

There are 3 treatment groups in the Induction Period and 5 treatment groups in the Maintenance Period. The Induction Period groups include:

- Brazikumab low dose

  IV brazikumab 720 mg at Week 0 (Day 1), Week 2, and Week 6
- Brazikumab high dose

  IV brazikumab 1440 mg at Week 0 (Day 1), Week 2, and Week 6
- Placebo

  IV placebo at Week 0 (Day 1), Week 2, and Week 6

The treatment groups for the Induction Period will be referred to as the Induction treatment groups, while the combination of the treatment groups from the Induction Period and the Maintenance Period will be referred to as the Induction/Maintenance treatment groups:

- IV brazikumab 1440 mg followed by SC 240 mg (IV-high/SC-high)
- IV brazikumab 1440 mg followed by SC 120 mg (IV-high/SC-low)
- IV brazikumab 720 mg followed by SC 240 mg (IV-low/SC-high)
- IV brazikumab 720 mg followed by SC 120 mg (IV-low/SC-low)
- Placebo

Prior to implementation of CSP Amendment 5, there were an additional two treatment groups in the Induction Period and three additional treatment groups in the Maintenance Period:

- Vedolizumab
  - IV vedolizumab 300 mg at Visit 2 (Day 1), Visit 3 (Day 15), and Visit 5 (Day 43) followed by IV vedolizumab 300 mg every 8 weeks beginning at Visit 7 (Day 99)
- Brazikumab high dose 2100 mg
  - IV brazikumab 2100 mg at Visit 2 (Day 1), Visit 3 (Day 15), and Visit 5 (Day 43) followed by SC brazikumab 210 mg or 105 mg every 4 weeks beginning at Visit 6 (Day 71)

Additionally, changes to dose have been made from CSP Amendment 4 to Amendment 5:

- The 700 mg brazikumab IV treatment group was changed to 720 mg IV; the 1400 mg brazikumab IV treatment arm was changed to 1440 mg IV.
- The 105 mg brazikumab SC treatment group was changed to 120 mg SC; the 210 mg brazikumab SC treatment arm was changed to 240 mg SC.

Patients randomized per the previous protocol version (D5272C00001 [Legacy #3151-201-008] Amendment 4) will remain in the treatment group assigned at the time of randomization. Pre-Amendment 5 participants will be analysed and reported together within their similar treatment group; for example, participants on 700 mg IV will be included in the 720 mg IV group, and participants on 105 mg SC will be included in the 120 mg SC group (see Section 3.3).

The schedule of assessments can be found in the CSP Section 1.3, Tables 1, 2, 3, and 4.

2 CHANGES TO PROTOCOL PLANNED ANALYSES

This SAP has been authored to account for the different protocol versions, assuring that all data collected will be presented in an appropriate fashion. The Maintenance Period Population (MPP) has been updated such that data for participants will be summarized according to the randomized intervention.

3 DATA ANALYSIS CONSIDERATIONS

3.1 Timing of Analyses

An interim (primary) analysis will be performed after the last participant completes the Week 10 visit or discontinued early in the Induction Period. An additional (final) analysis will be undertaken after the last participant has completed the Week 54 visit or discontinued early. All analyses described in this SAP will be included in the CSR.

3.2 Analysis Populations

The following populations are defined in Table 3-1.

Table 3-1 Populations for Analysis

Population/Analysis set	Description
All Patients Analysis Set	This includes all participants screened for the study.
Full Analysis Set	This analysis set includes all participants who are randomized, excluding participants randomized to active comparator (vedolizumab) or 2100 mg brazikumab. Participants will be summarized according to the randomized study intervention.
Safety population	This includes all participants who have received at least 1 dose of study intervention.[1] Erroneously treated subjects (e.g., those randomized to treatment A but actually given treatment B) are accounted for in the treatment group of the treatment they actually received. A participant who has on one or several occasions received active study intervention is classified as active and is accounted for in the active study intervention treatment group. A participant who has received both active dose regimens is classified as the higher active dose regimen.[2]
PK population	The PK population includes all participants who receive at least 1 dose of study intervention and have at least 1 PK sample containing detectable brazikumab concentrations.
Maintenance Period population	This includes all participants who received at least 1 dose at or after Week 10, excluding participants randomized to active comparator (vedolizumab) or 2100 mg brazikumab. Data for participants will be summarized according to the randomized study intervention.
ADA Evaluable Analysis Set	This includes Safety population participants having non-missing baseline ADA result and at least one non-missing post baseline ADA result; participants will be summarized according to the actual study treatment group.

ADA = Anti-drug antibodies; PK = pharmacokinetic

[1] Participants who received vedolizumab will only be included in the listings. Participants who received 2100 mg brazikumab will be included in the 'total' brazikumab group but not in a separate treatment group.

[2] "Higher active dose" may refer to either higher IV or SC dose.

It should be noted that the Full Analysis Set (FAS) includes all participants enrolled prior to protocol amendment 5, excluding participants randomized to active comparator (vedolizumab) or 2100 mg brazikumab, which are only in the INFORM legacy database. This means that INFORM data (in addition to RAVE data) will be used for efficacy analyses. It should be noted that 3 patients PPD were randomized but did not receive any treatment. Although these patients should be included into the FAS as per definition in Table 3-1, it was agreed to exclude them from the FAS.

3.3 General Considerations

Statistical analysis will be performed using the SAS Software, Version 9.4 or higher.
Efficacy data for participants who were enrolled prior to Protocol Amendment 5 will be included in the analysis, for treatment groups that are still included post Protocol Amendment 5. That is, the following pre-Amendment 5 treatment groups will be pooled with the respective post-Amendment 5 treatment groups as below:

Table 3-2 Pooling strategy for pre-Amendment 5 treatment groups

Pre-Amendment 5 treatment group	Post-Amendment 5 treatment group
IV 700 mg	IV 720 mg
IV 1400 mg	IV 1440 mg
SC 105 mg	SC 120 mg
SC 210 mg	SC 240 mg

Efficacy data for participants randomized to the active comparator (vedolizumab) and to the 2100 mg brazikumab (treatment groups removed in Amendment 5) will be listed only. Pre-Amendment 5 safety data will be summarized together with the safety data collected after Protocol Amendment 5 as above. Participants in the highest dose of brazikumab (2100 mg) will be included in the total (grouped) brazikumab summary but not summarized separately. Safety data for participants receiving vedolizumab will be listed only. Study-related raw data for enrolled participants, including derived data, will be presented in data listings. All listings will be sorted for presentation in order of Induction treatment group, study site, participant number, date of event, and parameter name (if applicable).

Descriptive statistics for continuous variables include the number of participants (n), mean, standard deviation (SD), median, 1st (Q1) and 3rd (Q3) quartiles, minimum (Min), and maximum (Max). Summary statistics for categorical variables include the number of participants (n) and percent (%). Minimum, and maximum will be presented to the same number of decimal places as in the recorded data. Mean, median, Q1, Q3, and CI will be displayed with 1 more decimal place than the measured value. SD and SE will be displayed with 2 decimal places more than the measured value. Percentages will be displayed with 1

decimal place, except 0 and 100%, which will be displayed without any decimal places. P-values will be reported to 3 decimal places with values less than 0.001 displayed as <0.001. All efficacy and safety data will be analysed based on the analysis visits. The details of analysis visit windowing rules are described in Section 3.3.2.2. Tabular summaries will be presented by Induction or Induction/Maintenance treatment group, as appropriate.

In general, unless otherwise stated, efficacy analyses will be conducted with a two-sided test at a significance level of $\alpha = 0.05$. All confidence intervals will be 2-sided 95% confidence intervals, unless stated otherwise. There will be no multiplicity control across endpoints; nominal p-values will be provided as descriptive measures to aid in interpretation of the analysis results and should not be interpreted as statistical significance.

3.3.1 General Study Level Definitions

3.3.1.1 Study Day 1

Day 1 will be defined as the date of first dose of study intervention. If a participant is randomized but discontinues participation without receiving any study intervention, then their Study Day 1 will be undefined. It should be noted that 3 participants PPD PPD enrolled prior to protocol amendment 5, were randomized but did not receive any treatment and are excluded from all analyses.

3.3.1.2 Baseline

Baseline is defined as the last non-missing assessment before the first dose of study intervention administration, with the following exceptions for mMS:

- The Baseline Patient Reported Outcomes (PRO) items (including stool frequency and rectal bleeding) will be based on the participant's average evening diary data completed over the most recent 3 days (consecutive or non-consecutive) within the 7 consecutive days of the Screening colonoscopy exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep).
- If the Physician's Global Assessment (PGA) is not possible at the same time as the colonoscopy visit, the PGA score from Visit 1 may be used.
- The mMS assessed at Screening will be used as Baseline.

Change from Baseline is defined as (Value at Visit X – Value at Baseline).

Percent Change from Baseline is defined as (Value at Visit X – Value at Baseline)/(Value at Baseline)×100.

3.3.1.3 Study Periods

The study periods defined in this section are different from the safety reporting periods for Adverse Events (AE) defined in Section 4.6.2.1.

Screening Period: -35 to -1 days (Screening Visit 1 and Colonoscopy)

Induction Period: Day 1 to 84, or day of early termination prior to Day 84

Maintenance Period: Day 85 to analysis window Week 54, or day of early termination on or later than Day 85 or completed Week 54 and rolling over to OLE.

Double-Blind Treatment Period: Induction through Maintenance Periods

Safety Follow-up Period: 18 weeks after the last dose of brazikumab or placebo (Follow-up Visits 1 (8 weeks post last dose) and 2 (18 weeks post last dose))

3.3.2 Visit Window

3.3.2.1 Study day

For assignment of data to analysis windows, study days will be defined in reference to Study Day 1 as defined in Section 3.3.1.1.

For any analysis visit time windows, if the assessment date is on or after the date of first dose of study intervention, the study day is calculated by:

assessment date – date of first dose of study intervention + 1.

If the assessment date is before the date of first dose of study intervention, the study day is calculated by:

assessment date – date of first dose of study intervention.

Therefore, a negative day indicates a day before the start dose of the study intervention.

If the assessment date is unavailable, the visit date will be used instead.

3.3.2.2 Conventions for the visit window

Table 3-3 presents the analysis windows assigned for efficacy (except daily data parameters) and safety analyses and the corresponding range of treatment days (window) during which data for a particular nominal visit may be collected. The analysis windows will be calculated from the date of first dose for study intervention. These are analysis windows that contain, but are not the same as, the actual visit windows. Each analysis window is equidistant between the neighboring Scheduled Visit Days. For example, Week 2 = Day 15 and Week 4 = Day 29 (Table 3-3). The center of the analysis window is calculated as the average between the two Visit Days: (15+29)/2= 22. Therefore, the end of the Week 2 analysis window is Day 21 and the beginning of the analysis window for Week 4 is Day 22.

Table 3-3 Analysis Windows

Derived Visit	***Scheduled Visit Day***	***Window***
Screening Visit	-35	[-35,-1]
Week 0	Day 1[a]	Days = 1
Week 2	Day 15	Days [2,21]

Week 4	Day 29	Days [22,35]
Week 6	Day 43	Days [36,56]
Week 10	Day 71	Days [57,84]
Week 14	Day 99	Days [85,112]
Week 18	Day 127	Days [113,140]
Week 22	Day 155	Days [141,168]
Week 26	Day 183	Days [169,196]
Week 30	Day 211	Days [197,224]
Week 34	Day 239	Days [225,252]
Week 38	Day 267	Days [253,280]
Week 42	Day 295	Days [281,308]
Week 46	Day 323	Days [309,336]
Week 50	Day 351	Days [337,364]
Week 54	Day 379	Day ≥ 365 to last dose + 6 weeks
Follow-up Visit 1[c]	(Last dose + 6 weeks+1) to (last dose +13 weeks)	
Follow-up Visit 2[c]	(Last dose + 13 weeks +1) to End of Study	

[a] Relative to the date of the first dose of study intervention in the Induction Period. Day 1 = the date of the first dose of study intervention in the Induction Period. There is no Day 0.
[b] Presented in analysis tables for safety parameters, including but not limited to electrocardiograms, clinical laboratory values, and vital signs.
[c] Follow-up Visits will only be carried out for participants who discontinue early or do not join the open label extension study.

3.3.2.3 Early discontinuation visits

Early discontinuation visits for either discontinuation of study intervention or early discontinuation from study will be mapped to analysis windows following the conventions in Table 3-3. Early discontinuation visits will only be used in analysis if the nominal scheduled visit result is missing.

3.3.2.4 Multiple observations

All observations will be assigned to their corresponding analysis window following the conventions in Table 3-3. If multiple assessments are recorded in a single analysis window, the following rules will be followed unless otherwise specified:

- If there are 2 or more observations within the same analysis window, then the observation closest to the Visit where study intervention was given will be used in the analysis.
- If there are two Visits with study intervention, the observation closest to the planned protocol Scheduled study day Visit will be used. The other observation(s) will not be used for analysis but will be included in listings.

- If 2 observations are equidistant from the study day when the study intervention was given, the observation with the earlier collection date will be used in the analysis.
- At Week 4 and Week 54, where study intervention is not planned, then the observation closest to the Scheduled Visit Day will be used in the analysis.
- At Week 10, assessments related to the Induction Period will not be used in analysis if they take place more than 7 days after receipt of maintenance study intervention.

For mMS, apply the same rules as above using assessments that are prior to bowel preparation before any endoscopy done during the analysis window. If there are multiple assessments during the Screening Period, the assessment date that is closest to the first dose of study intervention, and also prior to bowel preparation or the Screening colonoscopy, will be used. Once the assessment date is determined, the most recent 3 days (consecutive or non-consecutive) during the previous seven consecutive day window of the diary (PRO) data will be used to calculate the mMS score (excluding days of bowel prep and colonoscopy).

3.3.3 Handling of Unscheduled Visits

If the study intervention cannot be administered during a scheduled study visit, the rescheduled dose will be administered as an unscheduled visit.

Post-baseline unscheduled visits (including early termination visits) will be mapped to a visit based on the analysis windows in Table 3-3. Both unscheduled and scheduled visits will be used in case of multiple observations within an analysis visit window and same rules as in Section 3.3.2.4 will be applied.

All data from unscheduled visits will be presented in data listings.

3.3.4 Multiplicity/Multiple Comparisons

The null hypothesis is that the clinical remission rate derived from mMS (primary efficacy endpoint) for each of the IV brazikumab 720 mg and IV brazikumab 1440 mg treatment groups is not different from that for the IV placebo treatment group. The alternative hypothesis for each of the brazikumab treatment groups is that the clinical remission rate associated with a brazikumab treatment group is different from that for the placebo treatment group, i.e.:

For the high dose,

$$H_0: p_{high} = p_{placebo} \ vs. H_1: p_{high} \neq p_{placebo}$$

For the low dose,

$$H_0: p_{low} = p_{placebo} \ vs. H_1: p_{low} \neq p_{placebo}$$

All testing will be performed with nominal p-values; because this is not a confirmatory study, no multiplicity adjustments will be performed and secondary/exploratory tests should not be interpreted in terms of statistical significance.

3.3.5 Handling of Protocol Deviations in Study Analysis

A per-protocol (PP) analysis is not planned for this study. Important protocol deviations (IPD) will be listed and tabulated in the CSR. Important protocol deviations are a subset of protocol deviations that may significantly impact the completeness, accuracy, or reliability of the study data or that may significantly affect a participant's rights, safety, or well-being.

IPDs will be documented prior to unblinding the study data for the Induction Period and will be reported for the Induction Period at the Week 10 interim (primary) analysis. The final list of IPD for the Maintenance Treatment Period will be reported at the Week 54 final analysis.

The protocol deviations, definitions, process for identification and assessment are detailed in a separate protocol deviation plan.

4 STATISTICAL ANALYSIS

4.1 Study Population

The study population covers participant disposition, analysis sets, protocol deviations, demographics, baseline characteristics medical history, prior and concomitant medication, and study drug compliance.

4.1.1 Participant Disposition and Completion Status

4.1.1.1 Definitions and Derivations

Enrolled

A participant is enrolled if they, or their legally acceptable representative, sign the informed consent form.

Screen failure

Screen failures are defined as participants who consent to participate in the clinical study but are not subsequently randomly assigned to study intervention.

Randomized

Randomized participants are defined as those who undergo randomization to the Induction Period and receive a randomization number.

Study discontinuation

A study discontinuation will occur if a participant who signs the Informed Consent Form (ICF) and is dosed ceases participation in the study, regardless of circumstances, before the completion of the protocol-defined study procedures.

Treatment discontinuation

A treatment discontinuation will occur if a participant who has taken at least one dose of the study intervention is discontinued from the study intervention before the completion of protocol-defined study procedures.

Completer

A study intervention completer for Induction Period and/or Maintenance Period is defined as follows:

- If the participant has received the first dose of IV brazikumab and the dose at Week 6, the participant is considered to be a study intervention completer for the Induction Period.
- If the participant has received the first dose of SC brazikumab, and the dose at Week 50, and has not missed more than one dose, the participant is considered to be a study intervention completer for the Maintenance Period.

4.1.1.2 Presentation

Listings and summaries will be presented for the All Patients Analysis Set. The listings will include but not be limited to disposition event, time of event, and exposure. The summary will include the number and percentage of participants in the All Patients Analysis Set who:

- Were screened
- Failed screening, and reasons
- Were randomized
- Randomized, did not receive study intervention
- Started Induction Period treatment
- Completed Induction Period treatment
- Discontinued Induction Period treatment, and reasons
- Discontinue study prior to Week 10, and reasons
- Received Week 10 endoscopy
- Started Maintenance Period treatment
- Ongoing in the Maintenance Period (IA only)
- Completed Maintenance Period treatment
- Discontinued Maintenance Period treatment, and reasons
- Completed Week 54
- Withdrew from the study prior to Week 54, and reasons
- Rolled over to the OLE study
- Started the Safety follow-up period
- Ongoing in the Safety follow-up period (IA only)
- Completed the Safety follow-up
- Discontinued during the Safety follow-up period, and reasons

Participant recruitment will be summarized overall and by brazikumab total and Induction treatment group by country and study site. Brazikumab 2100 mg and vedolizumab groups (pre-Amendment 5) will be included. The stratification factor (prior biologic use) will be summarized according to the IXRS vs. according to the eCRF, by Induction treatment group and brazikumab total.

Percentages are based on the number of subjects who started treatment in the induction period or in the maintenance period.

All participants who prematurely discontinue will be listed by period, i.e., the Induction Period, the Maintenance Period, and the Safety Follow-up Period with discontinuation reason for the All Patients Analysis Set.

Summary of COVID-19 study disruptions

COVID-19 disruptions will be summarized for the FAS by Induction treatment, to include:

- Number of participants with at least one disruption due to COVID-19 pandemic
  - Visit impacted
  - Study intervention impacted
  - Concomitant medication impacted
  - Discontinued treatment due to COVID-19
  - Withdrew from the study due to COVID-19

4.1.2 Analysis Sets

4.1.2.1 Definitions and Derivations

There are 6 analysis populations:

All Patients Analysis Set

This analysis set includes all participants screened/enrolled for the study. Screened participants are defined as those who sign informed consent.

The Full Analysis Set (FAS)

This analysis set includes all participants who are randomized to a treatment group (as per the Intent-To-Treat (ITT) principle), excluding the 2100 mg dose group of brazikumab and the vedolizumab group. The FAS will be used for all efficacy analyses unless otherwise indicated. Participants will be summarized according to their randomized study intervention.

As described in Section 3.2, 3 patients were randomized but did not receive any treatment. Although these patients should be included into the FAS as per definition in Table 3-1, it was agreed to exclude them from the FAS.

The Safety (SAF) Analysis Set

This includes all participants who receive ≥ 1 administration of study intervention. Participants who received vedolizumab will only be included in the listings. Participants randomized to 2100 mg brazikumab will be included in the 'total' brazikumab group but will not have a separated reporting category. Participants will be summarized according to the actual study intervention received.

PK Population

This includes all participants who receive at least 1 dose of study intervention and have at least 1 PK sample containing detectable brazikumab concentrations.

Maintenance Period (MPP) Analysis Set

This includes all participants who received at least 1 dose of study intervention at or after Week 10, excluding participants randomized to active comparator (vedolizumab) or 2100 mg brazikumab. Participants will be summarized according to their randomized study intervention.

ADA Evaluable Analysis Set

This includes Safety analysis set participants having non-missing baseline ADA result and at least one non-missing post-baseline ADA result; participants will be summarized according to the actual study treatment group.

4.1.2.2 Presentation

The number of participants in each analysis set will be listed and summarized by Induction treatment group, brazikumab total, including reasons excluded from the Safety set, the FAS, and the PK set. A similar summary will be presented for the Maintenance Period, and will include reasons excluded from the Safety set, MPP, and the PK set.

4.1.3 Protocol Deviations

4.1.3.1 Definitions and Derivations

The definition and derivation of IPDs is described in a separate Protocol Deviation plan. These important protocol deviations will be reviewed and documented for the Induction Period before the database lock for interim (primary) analysis when all participants have completed Week 10 or discontinued early, and again for the Maintenance Treatment Period at the Week 54 final analysis.

4.1.3.2 Presentation

Participants with important protocol deviations in the study will be listed and summarized overall and by Induction treatment group for the FAS.

IPDs summarized include:

- Violations of inclusion criteria
- Violations of exclusion criteria

- Discontinuation criteria for investigational product are met but the participant was not withdrawn from study intervention
- Deviations related to study procedures
- Excluded medication taken
- Investigational product deviation
- Other important protocol deviations

4.1.4 Demographics

4.1.4.1 Definitions and Derivations

Demographic variables will include age, sex, region, country, race and ethnicity. Patient age is entered in years. Age will be further categorized in 3 groups (<40, 40-65, and >65). Participants who reported multiple races will be counted in the category "multiple".

4.1.4.2 Presentation

Demographic variables will be listed and summarized by Induction treatment group, brazikumab total and overall for the FAS and include:

- Age
- Age group [<40, 40-65, and >65]
- Sex
- Race
- Ethnicity
- Country
- Region [US and Non-US]

4.1.5 Baseline Characteristics

4.1.5.1 Definitions and Derivations

Baseline characteristics include

- Height
- Weight
- BMI
- BMI category (<18.5, 18.5 – 24.9, 25-29.9, ≥ 30)

Baseline

Baseline and Change from Baseline are defined in Section 3.3.1.2.

BMI

Body Mass Index (BMI) is calculated from height and weight by dividing (weight in kg) by (height in m)^2. Units are kg/m^2.

4.1.5.2 Presentation

The Baseline characteristics will be listed for the Safety Population.

4.1.6 Disease Characteristics

4.1.6.1 Definitions and Derivations

Disease characteristics to be collected at Baseline are listed in Table 4-1.

Table 4-1 Disease characteristics to be summarized at Baseline

Disease Characteristics at Baseline	Disease duration of UC (years) Current immunomodulator use Baseline disease location (descending colon, pancolitis, other) Extra-intestinal manifestations Prior use of JAK inhibitors Prior biologic use at randomization (stratification factor) Number of different mechanisms of action of biologics Inadequate responders on prior biologics C-Reactive Protein Fecal Calprotectin
mMS	Total score (0-9) Findings at endoscopy Stool frequency score Rectal bleeding score

Disease duration will be calculated as the (Date of First Study Intervention - Date of Initial Diagnosis of Ulcerative Colitis)/365.25 years from the form Ulcerative Colitis History.

Current immunomodulator use: Immunomodulators include azathioprine, 6-mercaptopurine, methotrexate, and other. The "other" category could include cyclosporine A, tacrolimus, and others as identified by the study physician.

Baseline disease location will have options "Descending colon", "Pancolitis", and "Other".

Extra-intestinal manifestation (EIM) responses include
"Extra-intestinal Manifestations" =

- Arthropathy: peripheral, axial arthropathy
- iritis/uveitis
- Skin problems: erythema nodosum, pyoderma gangrenosum, aphthous stomatitis/ aphthous ulcers, anal fissures, anal abscess, non-anal abscess
- other

JAK inhibitors include prior use of tofacitinib or upadacitinib.

Prior biologic use at randomization: prior biologic use = 'Yes' if any biologic use was recorded prior to randomization. See Table 4-2 for examples of biologics.

Inadequate responders on prior biologics at baseline: For each prior biologic used, 'inadequate responder' will be defined as "Intolerance" or "Lack of efficacy".

Number of different mechanism of actions of prior biologics at randomization: each of the biologics defined as Ulcerative Colitis Medication, **and** are defined as inadequate responders as defined above, will be categorized as anti-TNF, anti-integrin, or IL12/23 inhibitor. The number of different mechanisms will be 0-3.

C-Reactive Protein (CRP) and Fecal Calprotectin (FCP) results will be taken from the lab results.

The ATC codes and preferred terms for the referenced medications are listed in Table 4-2:

Table 4-2 Drug names and ATC codes

Category	*Drug Name*	*ATC5 code*	*ATC4 code*
Immunomodulator	Azathioprine	L04AX01	L04AX
	6-mercaptopurine	L01BB02	L01BB
	Methotrexate	L01BA01/L04AX03	L01BA/L04AX
	Cyclosporine A	L04AD01	L04AD
	Tacrolimus	L04AD02	L04AD
	Other		
JAK inhibitors	Tofacitinib	L04AA29	L04AA
	Upadacitinib	L04AA44	L04AA
BIOLOGICS			
Anti-TNF	Etanercept	L04AB01	L04AB
	Infliximab	L04AB02	L04AB
	Adalimumab	L04AB04	L04AB
	Certolizumab pegol	L04AB05	L04AB
	Golimumab	L04AB06	L04AB
Anti-Integrin	Vedolizumab	L04AA33	L04AA
	Natalizumab	L04AA23	L04AA
IL-12 and IL-23 Inhibitor	Ustekinumab	L04AC05	L04AC

mMS is defined in Appendix 7.1 and mMS Baseline in Section 3.3.1.2.

4.1.6.2 Presentation

The Disease characteristics will be listed and summarized descriptively by Induction treatment group, brazikumab total, and overall for the FAS Population. mMS total score will be descriptively summarized with n, mean, SD, min, Q1, median, Q3, and max. mMS subscores will be summarized as categorical variables (n, %).

4.1.7 Medical History and Concomitant Disease

4.1.7.1 Definitions and Derivations

Ulcerative Colitis Medical History

Surgical history, chest x-ray, and family history of colorectal cancer will be collected on the form Associated Persons Medical History. Tuberculosis history is found on the Medical History/Tuberculosis Worksheet, and vaccination history is found on the Concomitant/Prior Medications: Vaccination History form.

General Medical and Surgical History

Verbatim terms will be mapped to system organ class (SOC) and preferred term (PT).

Participants' medical and surgical histories and concurrent procedures will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), version 24.0 or newer.

4.1.7.2 Presentation

Ulcerative Colitis medical history as described in Section 4.1.7.1 will be listed for the Safety analysis set. Chest x-ray, vaccination history, family history of colorectal cancer, and tuberculosis history will be presented in the listing.

General medical and surgical histories will be summarized applying MedDRA terms. Each SOC and PT will be summarized by Induction treatment group, brazikumab total, and overall for the FAS Population. In addition, the number and percentage of participants with concurrent procedures in each SOC and PT will be summarized by Induction treatment group, brazikumab total, and overall for the FAS. Concurrent procedures will also be presented in a listing.

4.1.8 Prior and Concomitant Medications

4.1.8.1 Definitions and Derivations

The World Health Organization (WHO) Drug Dictionary Enhanced, March 2022 or newer, will be used to classify prior and concomitant medications by therapeutic class and generic drug name.

Prior medication

Prior medication is defined as any medication taken before the first dose of the study intervention.

Concomitant medication

Any medication taken by the participant at any time between the date of the first dose (including the date of the first dose) of IP up to 4 weeks after the last dose of IP, inclusive, is considered concomitant medication. Any medications taken by the subject prior to the first dose date of IP that continues after the first dose date of IP are considered both prior medications and concomitant medications. Any medications taken prior to the first dose date of IP whose dose has changed after the date of first dose of IP will be captured with a separate

data entry with the date of dose change, which will be used to determine if the medication is prior or concomitant.

Any medications started later than 4 weeks after the last dose of study intervention date will not be presented in the summary tables but will be included in the participant data listings.

Permitted UC medication

Permitted UC medication should be recorded as concomitant medications. Refer to Table 8 in Section 6.5.1 and Table 9 in Section 6.5.2 of the CSP for details.

Prohibited UC intervention

Prohibited UC interventions should not be taken during the study. The interventions listed below are considered exclusionary and are not permitted through Week 54 of the study. If prohibited interventions are taken, the study intervention must be discontinued. Use of prohibited intervention after randomization is considered an intercurrent event for the primary endpoint (see Section 4.2.1.2).

- Natalizumab
- Anti-TNFα agents
- Vedolizumab
- Any commercially available or experimental biologic agent (e.g., risankizumab, mirikizumab, guselkumab)
- Tofacitinib
- Upadactinib
- Calcineurin inhibitors (e.g., cyclosporine and tacrolimus)
- Mycophenolate mofetil
- Sirolimus (rapamycin)
- Thalidomide
- Live attenuated vaccine
- Intra-abdominal surgery
- Fecal microbiota transplantation
- The use of alternative or complementary treatments must be discussed with the Study Physician/designee. Treatments such as Chinese herbal therapies are considered prohibited interventions.
- Any experimental product or device as specified by the study Inclusion and Exclusion Criteria.

Note: Topical (rectal) aminosalicylic acid (e.g., mesalamine) or topical (rectal) steroids are prohibited from 2 weeks prior to the Screening colonoscopy through Week 54 and will be considered as rescue therapy if taken during this time period. Participants will continue to receive study intervention. The use of other immunosuppressant therapies not listed above

will require discussion with the medical monitor to determine whether study intervention must be discontinued.

Rescue treatment

Permitted rescue medications are allowed as per investigator judgment. Participants will continue to receive study intervention unless a prohibited medication is used.

The definition of a medication as rescue treatment will be guided by CSP Amendment 5, regardless of which amendment a participant was enrolled under. Any new concomitant medication, or any increase in dose of a Baseline medication required to treat new or unresolved UC symptoms, except for antidiarrheal medications already used at Baseline, will be considered rescue treatment. Any **new** initiation of medications or interventions will be considered as rescue treatment:

- 5-aminosalicylates
- Parenteral, oral, or rectal CS
- Oral CS doses above Day 1 dose
- Immunomodulators (azathioprine, 6-mercaptopurine, methotrexate)
- IV immunoglobin

CS doses up to the levels taken on Day 1 are not considered rescue treatment. Other investigational products are not permitted to be used as rescue treatment. Use of rescue medication after randomization is considered an intercurrent event for all endpoints listed in Table 4-3 (see Section 4.2.1.1).

The criteria for use of rescue treatment are found in the CSP Section 6.5.5.

4.1.8.2 Presentation

The following summaries will be provided by treatment group, brazikumab total and overall for the FAS, within WHO drug class and generic drug name:

- Prior medications
- Prior corticosteroid use for UC
- Prior immunomodulator use for UC
- Prohibited concomitant medication use in the Induction Period
- Prohibited concomitant medication use in the Maintenance Period
- Concomitant rescue medication use for UC
- Permitted concomitant medication use
- Concomitant corticosteroid use
- Concomitant immunomodulator use

Prohibited concomitant medication use will be summarized by Induction treatment group, brazikumab total and overall for the Induction Period for the FAS, and in each Induction/ Maintenance treatment group, brazikumab total and overall for the Maintenance Period for the MPP. If a participant took a specific medication multiple times or took multiple medications within a specific therapeutic class, that participant would be counted only once for the coded generic drug name or therapeutic class. If a participant took a specific prohibited concomitant medication in both the Induction and the Maintenance Periods, they will be counted in each summary. Any prior and concomitant medications will be included in listings, including corticosteroid medications.

4.1.9 Study Intervention Compliance

4.1.9.1 Definitions and Derivations

Participants are dosed at the site and will receive study intervention directly from qualified blinded staff under medical supervision. The date and time of dose administered in the clinic will be recorded in the source documents. Study intervention compliance will measure the number of doses received and will be calculated as the total number of doses taken during that period divided by the number of expected doses, multiplied by 100.

4.1.9.2 Presentation

Study intervention compliance will be listed and summarized descriptively as number and percentage receiving expected IP for each week (Weeks 0, 2, and 6) for the Induction Period by Induction treatment group and for brazikumab total using the FAS, and for the Maintenance Period and Double-Blind Treatment Period by Induction/Maintenance treatment group using the MPP. The mean (SD), median, and range of the number of doses received will be presented for the Maintenance and Double-Blind Periods by treatment group. Administration of investigational product will be presented in a listing.

4.2 Endpoint Analyses

This section covers details related to the endpoint analyses such as primary, secondary, other endpoints, including sensitivity and supportive analyses. Whenever parametric models are described, the data will undergo appropriate tests of the normality assumptions and if necessary, data will be transformed or non-parametric substitutes will be used for analysis (for instance, ANCOVA could be replaced by rank ANCOVA).

The purpose of this study is to evaluate the efficacy and safety of brazikumab versus placebo to induce clinical remission based on endoscopic improvement, rectal bleeding, and stool frequency scores in participants with moderate to severe UC who have failed or are intolerant to conventional therapy (5-aminosalicylates, corticosteroids or immunomodulators; 6-mercaptopurine, azathioprine, methotrexate). The study population will also include participants who are biological intervention naïve, have demonstrated a successful response to prior biological intervention, who have failed or were intolerant to biological intervention, or who are corticosteroid dependent. See Table 4-3 for a complete list of objectives and endpoints.

Table 4-3 Objectives and endpoints

Statistical category	*Endpoint*	*Analysis Population*	*Intercurrent event strategy*	*Population level summary (analysis)*	*Details in section*
Objective 1: To compare the efficacy of brazikumab with that of placebo to achieve clinical remission at Week 10					
Primary	Clinical remission at Week 10: mMS meets ALL of the following criteria at Week 10: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS	NRI before Week 10	Percentage of participants achieving the endpoint	4.2.1
Objective 2: To compare the efficacy of brazikumab with that of placebo to achieve sustained clinical remission					
Secondary	Sustained clinical remission requires meeting ALL of the following criteria at both Week 10 and Week 54: mMS at both Week 10 and Week 54: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS	NRI before Week 54	Same as primary	4.2.3
Objective 3: To compare the efficacy of brazikumab with that of placebo to achieve CS-free clinical remission at Week 54					
Secondary	CS-free clinical remission at Week 54 requires meeting ALL the following criteria at Week 54 and who are CS-free for at least 12 weeks before the assessment at Week 54: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS who took CS at Baseline	NRI before Week 54	Same as primary	4.2.3
Objective 4: To compare the efficacy of brazikumab with that of placebo to achieve clinical response at Week 10					
Secondary	Clinical response at Week 10 requires meeting the following criterion at Week 10:	FAS	NRI before Week 10	Same as primary	4.2.3

Statistical category	*Endpoint*	*Analysis Population*	*Intercurrent event strategy*	*Population level summary (analysis)*	*Details in section*
	• Reduction in mMS ≥ 2 points from Baseline AND ≥ 30% from Baseline AND a decrease in the rectal bleeding score ≥ 1-point from Baseline or a score of 0 or 1				
Objective 5: To compare the efficacy of brazikumab with that of placebo to achieve endoscopic improvement at Week 10					
Secondary	Endoscopic improvement improvement at Week 10 requires meeting the following criterion at Week 10: • Endoscopy subscore ≤ 1 at Week 10	FAS	NRI before Week 10	Same as primary	4.2.3
Objective 6: To evaluate the PK and immunogenicity of brazikumab					
Secondary	Population PK model of serum concentrations of brazikumab and analysis for serum anti-brazikumab antibodies	PK		PK model to be developed in the PK analysis plan and reported outside the CSR.	4.3
Objective 7: To characterize the exposure-response relationships of brazikumab					
Secondary	Exposure-response model linking primary endpoint to metrics of model-predicted individual brazikumab exposures	PK		Analysis to be developed in the PK SAP	4.2.4
Objective 8: To evaluate the safety and tolerability of brazikumab					
Safety	AEs, clinical laboratory values, vital signs, physical exams, ECGs	Safety		Summary tables	4.6
Objective 9: To establish the serum IL-22 concentration baseline clinical cutoff for its value in predicting the efficacy of brazikumab					
Exploratory	Exploration of the relationship of baseline serum IL-22 concentration with efficacy of brazikumab at Week 10 and establishment of the serum IL-22 concentration clinical cutoff through clinical remission at Week 10	FAS	NRI before Week 10	Differential effect method	4.2.4
Objective 10: To compare the efficacy of brazikumab with that of placebo to achieve clinical remission, by BM status					
Exploratory	Clinical remission overall and by BM status defined as: mMS at Week 10: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS	NRI before Week 10	Same as primary	4.2.6

Statistical category	*Endpoint*	*Analysis Population*	*Intercurrent event strategy*	*Population level summary (analysis)*	*Details in section*
Exploratory	Clinical remission overall and by BM status defined as: mMS at Week 54: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS	NRI before Week 54	Same as primary	4.2.6
Objective 11: To compare the efficacy of brazikumab with that of placebo to achieve sustained clinical remission, by BM status					
Exploratory	Sustained clinical remission as mMS at both Week 10 and Week 54: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS	NRI before Week 10 and 54	Same as primary	4.2.6
Objective 12: To compare the efficacy of brazikumab with that of placebo to achieve CS-free clinical remission, by BM status					
Exploratory	CS-free clinical remission defined as: mMS at Week 54 for participants who are CS-free for at least the last 12 weeks before the assessment at Week 54: • Endoscopy subscore = 0 or 1, AND • Rectal bleeding subscore = 0, AND • Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline	FAS who took CS at Baseline	NRI before Week 54	Same as primary	4.2.3
Objective 13: To compare the efficacy of brazikumab with that of placebo to achieve clinical response, by BM status					
Exploratory	Clinical response is defined as • Reduction in mMS ≥ 2 points from baseline AND ≥ 30% from baseline AND a decrease in the rectal bleeding score ≥ 1-point from Baseline or a score of 0 or 1 at Week 10	FAS	NRI before Week 10	Same as primary	4.2.6
Objective 14: To compare the efficacy of brazikumab with that of placebo to achieve endoscopic improvement, by BM status					

Statistical category	*Endpoint*	*Analysis Population*	*Intercurrent event strategy*	*Population level summary (analysis)*	*Details in section*
Exploratory	Endoscopic improvement at Week 10 is defined as • Endoscopy subscore ≤ 1 at Week 10	FAS	NRI before Week 10	Same as primary	4.2.3
Objective 15: To compare the efficacy of brazikumab with that of placebo to achieve other measures of improvement					
Exploratory	• Symptomatic remission at Week 10	FAS	NRI before Week 10	Same as primary	4.2.6
Exploratory	• Symptomatic remission at Week 54 • Endoscopic improvement at Week 54 • Endoscopic remission at Week 54	FAS	NRI before Week 54	Same as primary	4.2.6
Objective 16: To compare the efficacy of brazikumab with that of placebo in Changes from Baseline in mMS					
Exploratory	Change from Baseline at Weeks 10 and 54 in mMS total score	FAS	Treatment Policy	LS mean and 95% CI from an ANCOVA test	4.2.7
Objective 17: To compare the impact of brazikumab with that of placebo in Changes from Baseline in CS use					
Exploratory	Change from Baseline in CS use status (Y/N) to Weeks 2, 4, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, and 54	FAS	Treatment Policy	Descriptive statistics	4.2.8
Objective 18: To evaluate the impact of brazikumab on the mucosal inflammation of UC through Week 54					
Exploratory	Change from Baseline for GHI and RHI at Weeks 10 and 54	FAS	Accounted for by the mixed model (assumes MAR)	LS mean and 95% CI from a mixed model for repeated measures	4.2.9
Objective 19: To evaluate the impact of brazikumab on patient's signs and symptoms of UC					
Exploratory	Change from Baseline in signs and symptom scores (e.g., stool frequency, rectal bleeding, urgency, fatigue) derived from the BMD, evening diary; and FACIT-F at Weeks 6, 10, 18, 30, 42, and 54	FAS	Accounted for by the mixed model (assumes MAR)	LS mean and 95% CI from a mixed model for repeated measures	4.2.10
Objective 20: To evaluate the impact of brazikumab on HRQoL at Week 6 through Week 54					

Statistical category	*Endpoint*	*Analysis Population*	*Intercurrent event strategy*	*Population level summary (analysis)*	*Details in section*
Exploratory	Change from Baseline at Weeks 6, 10, 18, 30, 42, and 54 for IBDQ, and Weeks 10, 30, and 54 for SF-36, and EQ-5D-5L	FAS	Accounted for by the mixed model (assumes MAR)	LS mean and 95% CI from a mixed model for repeated measures	4.2.11
Objective 21: To explore transcriptional, histological, protein, and microbiome BMs in study participants and their relationship with study intervention, disease features, clinical outcomes, and patient characteristics					
Exploratory	• Serum plasma, fecal, or gut tissue proteins • Whole blood or gut tissue transcriptional changes • Histological and microbiome assessments	FAS	Exploratory analysis plans, including intercurrent event strategy and summary statistics, to be detailed in a separate exploratory analysis plan and reported outside of the CSR		

AE = adverse event; BM+ = biomarker serum IL-22 concentrations at or above a pre-established cutoff; BMD= Bowel Movement Diary; BSFS = Bristol Stool Form Scale; CS = corticosteroid; ECG = electrocardiogram; EQ-5D-5L = 5-level European Quality of Life - 5 Dimensions; FACIT-F = Functional Assessment of Chronic Illness Therapy – Fatigue Scale; HRQoL = health-related quality of life; IBDQ = Inflammatory Bowel Disease Questionnaire; IL-22 = interleukin-22; LS = least squares; GHI = Geboes Histological Index; IL = interleukin; MAR= missing at random; mMS = modified Mayo Score; NRI = non-responder imputation; NRS = numeric rating scale; PK = pharmacokinetic(s); RHI=Robarts Histopathology Index; SF-36 = Short-Form 36 Health Survey; UC = Ulcerative Colitis.

4.2.1 Primary Endpoint

4.2.1.1 Definition

The primary study objective is to compare the treatment group level response probability (measured as percentage of participants with clinical remission) between Induction brazikumab groups versus placebo group in the FAS Population. The four components of the estimand related to this primary objective are defined as follows:

A. The target population is defined through the inclusion/exclusion criteria for the study to reflect the patient population with moderately to severely active Ulcerative Colitis under investigation.
B. The variable is the same as the primary efficacy parameter, which is binary, indicating clinical remission if
   mMS at Week 10 meets:
   - Endoscopy subscore = 0 or 1, AND
   - Rectal bleeding subscore = 0, AND
   - Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline.
C. Study intercurrent events are captured in the variable definition as a participant who
   - discontinues the study intervention prematurely before Week 10
   - takes rescue treatment (specified in Section 4.1.8.1) before Week 10
   - uses prohibited medication (specified in Section 6.5.4 of the CSP) before Week 10

   Participants who experience any of these intercurrent events are considered as being unsuccessfully treated (Composite Strategy – ICH E9 [R1]) and will be imputed as non-responders (NRI).
D. The population-level summary measure is the percentage of participants with clinical remission.

The estimand defined above is considered the primary estimand. In this case, the clinical question is the treatment effect on clinical remission had no additional medication been available. Clinical remission values after rescue therapy or prohibited medications are not directly relevant since those values also reflect the impact of that additional medication and are considered as failure.

The null hypothesis is described in Section 3.3.4.

4.2.1.2 Derivations

The mMS consists of rectal bleeding, stool frequency, and modified endoscopy subscores. The mMS is captured on the eCRF page and the electronic Clinical Outcome Assessments (eCOA) device, but the mMS will be derived from the raw data of the individual components

for the statistical analysis. Clinical remission is calculated from the individual components of the mMS, as defined in Section 4.2.1.1.

The endoscopy must be performed within 14 days of the Week 10 scheduled visit, and not more than 7 days after administration of an SC maintenance injection. The PRO components (stool frequency and rectal bleeding) will be captured via the evening eDiary. The calculation of the stool frequency and rectal bleeding scores will be based on the participant's average evening eDiary data recorded over the most recent 3 days (consecutive or non-consecutive), within 7 days of endoscopy for the Week 10 endoscopy exclusive of days of bowel prep and endoscopy (i.e., 5 days prior to bowel prep). The average will be rounded up or down to get a whole number to define the subscore; for example, an average 1.49 stool frequency score rounds down to 1 and an average 1.5 score rounds up to 2. Baseline stool frequency is defined in Section 3.3.1.2.

A participant will be defined as meeting clinical remission if all 3 of the components of their mMS subscores (endoscopy, rectal bleeding, and stool frequency) meet the definition for clinical remission (defined in Section 4.2.1.1) at Week 10. The PGA subscore is not included in the definition of clinical remission.

See Appendix 7.1 for more details and Section 4.2.1.3 for handling of missing data.

Table 4-4 Components of clinical remission calculation

Item
Stool Frequency[a] 0 = Normal number of bowel movements for this participant 1 = 1 to 2 bowel movements more than normal 2 = 3 to 4 bowel movements more than normal 3 = 5 or more bowel movements more than normal
Rectal Bleeding[b] 0 = No blood seen 1 = Streaks of blood with stool less than half the time 2 = Obvious blood with stool most of the time 3 = Blood alone passed
Findings of Endoscopy[c] 0 = Normal or inactive disease 1 = Mild disease (erythema, decreased vascular pattern, no friability) 2 = Moderate disease (marked erythema, absent vascular pattern, friability, erosions) 3 = Severe disease (spontaneous bleeding, ulceration)
[a] Each participant serves as his or her own control to establish the degree of abnormality of the stool frequency. [b] The daily bleeding score represented the most severe bleeding of the day. [c] Subscore of 1 (mild disease) modified to reflect no friability for modified Mayo Score as described above.

4.2.1.3 Handling of Dropouts and Missing Data

Calculation of clinical remission requires endoscopy score, rectal bleeding score, and stool frequency score at both Baseline and Week 10. Missing data rules are as follows:

Missing mMS components at Baseline

PRO items are not expected to be missing at Baseline in the FAS. If the PRO item stool frequency or rectal bleeding has been recorded for fewer than 3 days (consecutive or non-consecutive) within 7 days of endoscopy exclusive of the days of endoscopy and bowel prep (i.e., during the 5 days prior to the bowel prep), the Baseline subscore will be set to missing and the participant will be a screen failure.

Missing mMS components at Week 10

1. PRO components missing

If there are not 3 days (consecutive or non-consecutive) with data within 7 days of endoscopy, exclusive of the days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep), data from the week prior to the 7-day scoring window will be used to generate the weekly score. For example, if the period of interest is Day -7 to Day -1, and there are fewer than 3 days of data completed, then include Day -8. If there are still fewer than 3 days of data completed then include Day -9, and so on until 3 days are observed. This principle can only be applied within the 14 days of endoscopy exclusive of the days of endoscopy and bowel prep (i.e., during the 12 days prior to the bowel prep) and up until the day of the previous dose. If the PRO components are still missing, the score will be set to be missing and the participant will be imputed as a non-responder.

2. Missing due to endoscopy component missing

Any endoscopy performed outside the 14-day window around the Week 10 scheduled visit or more than 7 days after administration of SC maintenance injection will be considered missing. If the missing mMS score is due to missing endoscopy subscore at Week 10, the endoscopy score will be set to missing and the participant will be imputed as a non-responder.

4.2.1.4 Primary Analysis of Primary Endpoint

The null hypothesis is described in Section 3.3.4. Statistical comparisons will be explored for each brazikumab dose group versus placebo.

For the primary efficacy endpoint, the observed response proportions will be calculated by Induction treatment group in the FAS. The efficacy endpoint will be analysed using a continuity-corrected Cochran-Mantel-Haenszel (CMH) test controlling for the randomization stratification factor (status of prior biologic use [yes or no] at randomization).

The CMH test analyses the association between treatments (brazikumab and placebo) and a binary outcome (clinical remission) while taking into account the stratification factor.

There are n_{ij} participants in each stratum, where i is the stratum and j is the treatment group. The number of participants achieving clinical remission is x_{ij} and the proportion of participants achieving clinical remission is $p_{ij} = x_{ij} / n_{ij}$.

For each stratum, the difference in proportion of participants achieving clinical remission is calculated as $d_i = p_{iB} - p_{iP}$, where B and P denote the different treatment groups (brazikumab and placebo, respectively).

Weights for each stratum are calculated as

$$w_i = \frac{n_{iB} \times n_{iP}}{n_{iB} + n_{iP}}$$

Then the weighted difference (WD) is calculated as

$$WD = \frac{\sum w_i d_i}{w_i}$$

The standard error (SE) of the weighted difference under the null hypothesis is given by

$$SE = \sqrt{\frac{\sum [w_i^2 Var(d_i)]}{(\sum w_i)^2}}$$

Where

$$Var(d_i) = \frac{p_i(1-p_i)n_i}{w_i(n_i - 1)}$$

And

$$p_i = \frac{x_i}{n_i} = \frac{x_{iB} + x_{iP}}{n_{iB} + n_{iP}}$$

To derive the confidence interval (CI) for the weighted difference in proportions, a correction will be applied to the variance to make it more accurate. For each strata,

$$Var(d_i) = \frac{p^*_{iB}(1-p^*_{iB})}{n_{iB}} + \frac{p^*_{iP}(1-p^*_{iP})}{n_{iP}}, where\ p^*_{ij} = \frac{x_{ij} + 2}{n_{ij} + 4}$$

The 95% CI is $WD \pm z_{0.975} * SE$.

The value of the test statistic is *WD/SE,* and the p-value from the two-sided test of no difference in treatment groups is calculated as

$$p = 2\ (1 - Prob(|^{WD}/_{SE}|))$$

Where $Prob(|^{WD}/_{SE}|)$is the distribution function of the standard normal distribution.

4.2.1.4.1 Presentation

The following analysis summary statistics will be presented for the primary efficacy endpoint and all endpoints on the FAS population by Induction treatment and overall:

- Responder counts for each treatment group
- Estimated treatment effect (i.e., difference in response rate for active versus placebo), and corresponding 95% CI and 2-sided p-value
- Number and percent of patients who are NRI due to:
    - Intercurrent event
        - Discontinues study intervention prematurely
        - Takes rescue medication
        - Uses prohibited medication
    - Missing data

  NRI will be assessed for intercurrent events, then for missing data. If a participant experiences more than one category of intercurrent event, that participant will be counted in each category.

A graph presenting clinical remission over time will be presented. In addition, a listing will be presented which includes, but is not limited to, whether the participant is a responder and which (if any) intercurrent events occurred.

4.2.1.5 Sensitivity Analyses of the Primary Endpoint

To assess the impact of Baseline BM on the primary efficacy endpoint, a logistic regression model will be performed controlling for the stratification factor (prior biologic use) and baseline serum IL-22 value as a covariate. The dependent variable for this sensitivity analysis is for the same primary estimand (clinical remission), and the Delta method will be used to estimate the rate differences and CIs (Guo et al, 2012).

A second sensitivity analysis will be performed with baseline serum IL-22 categorized into BM+ and BM- as identified in Section 4.2.4. If the model fails to converge with two factors, then each factor will be analysed individually as described in Section 4.2.1.7.

A third sensitivity analysis will use the previous definition of the primary endpoint, prior to implementation of Amendment 5:

Modified Mayo Score (mMS) at Week 10:

- Endoscopy subscore = 0 or 1, AND
- Rectal bleeding subscore = 0, AND
- Stool frequency subscore = 0

A fourth sensitivity analysis will use an alternate definition of the primary endpoint Modified Mayo Score (mMS) based on the updated definition of mMS from the FDA draft guidance (Guidance for Industry, 2022), at Week 10:

- Endoscopy subscore = 0 or 1, AND
- Rectal bleeding subscore = 0, AND
- Stool frequency subscore = 0 or 1

A fifth sensitivity analysis will be to run the primary endpoint using the post-Amendment 5 participants only.

These statistical comparisons are the same as that for the primary efficacy analysis described in Section 4.2.1.4. Both the odds ratio and rate difference will be included in the presentation of logistic regression results.

4.2.1.6 Supplementary Analyses of the Primary Endpoint

Treatment policy strategy

Per the treatment policy strategy, the actual value for clinical remission is used regardless of whether or not an intercurrent event occurs, i.e., the values of the variable of interest are used whether or not the following intercurrent events occur:

- Rescue therapy before Week 10
- Any prohibited treatments before Week 10

In this case, NRI will be used only for the following intercurrent event:

- Discontinue from study intervention before Week 10

The clinical question is the treatment effect on clinical remission regardless of which other therapies are to be used before clinical remission is experienced. Actual clinical remission values after the use of additional treatment will be relevant.

Alternate definition of clinical response

The primary analysis described in Section 4.2.1.4 will be repeated using the following alternate definition of clinical response at Week 10, where total MS is defined including the PGA as in Appendix 7.1: reduction in total MS ≥ 3 points from Baseline AND ≥ 30% from Baseline AND a decrease in the rectal bleeding score ≥ 1-point from Baseline or a score of 0 or 1.

4.2.1.7 Subgroup Analyses

The same analysis as described in Section 4.2.1.4 will be performed for the following subgroups:

- Prior biologic use (naïve vs prior use)
- Baseline CRP High (CRP+, ≥ 5 mg/L) vs. CRP Low (CRP-, <5 mg/L)

- Baseline FCP High (FCP+, ≥ 250 µg/g) vs. FCP Low (FCP-, < 250 µg/g)
- Region/Country (US, non-US)
- Race (White, non-White)
- Gender (Male, Female)
- Age group (<40, 40 to 65, >65)
- Baseline serum IL-22 (High/BM+, Low/BM-) and prior biologic use

These models will be specified using the "BY" statement to create each subgroup.

In case the model does not converge for any of these analyses, the prior biologic use strata will be dropped from the model. If the model does not converge in spite of this, no more changes will be made. Serum IL-22 subgroup classification will be determined in the interim (primary) analysis (see Section 4.2.4). For testing of "prior biologic use", the subgroup variable will not be included as a covariate. A forest plot will be presented.

4.2.2 Key Secondary Endpoint

N/A

4.2.3 Secondary Endpoints: Binary Endpoints

- Sustained clinical remission
- CS-free clinical remission at Week 54
- Clinical response at Week 10
- Endoscopic improvement at Week 10

4.2.3.1 Definition

For the following calculations, Baseline and percent Change from Baseline are defined in Section 3.3.1.2. Calculation of mMS components is described in Appendix 7.1 and clinical remission is defined in Section 4.2.1.1.

Sustained clinical remission: clinical remission at both Week 10 and Week 54.

CS-free clinical remission at Week 54: A participant is considered to have achieved CS-free clinical remission if that participant was taking CS at Baseline, was not taking corticosteroids for the 12 weeks before Week 54, and achieved clinical remission. Participants not taking CS at Baseline are excluded from this analysis.

Endoscopic improvement at Week 10: endoscopic subscore is ≤ 1 at Week 10.

4.2.3.2 Derivations

Sustained clinical remission at both Week 10 and Week 54: a successful response is achieved if the mMS score meets ALL of the following criteria at both Week 10 and Week 54:

- Endoscopy subscore = 0 or 1, AND
- Rectal bleeding subscore = 0, AND
- Stool frequency subscore = 0 or 1 AND at least a 1-point decrease from Baseline.

Clinical response at Week 10: a successful response is achieved if there is:

- a reduction in mMS ≥ 2 points from Baseline AND
- ≥ 30% from Baseline AND
- a decrease in the rectal bleeding score ≥ 1-point from Baseline or a score of 0 or 1 at Week 10.

4.2.3.3 Handling of Dropouts and Missing Data

Missing mMS PRO components and endoscopy score at Week 10 will be handled as described in Section 4.2.1.3.

If mMS PRO components are missing at Week 54, they will be handled similarly to the Week 10 missing data.

If the endoscopy score is missing at Week 54, it the endpoint will be imputed to non-responder.

4.2.3.4 Primary Analysis of Secondary Binary Endpoints

Analysis for each binary endpoint will be performed on the FAS population, by Induction treatment group, as for the Primary Endpoint as described in Section 4.2.1.4. Note that the FAS will be used with Induction treatment groups for Week 54 endpoints as well as Week 10 endpoints.

4.2.3.4.1 Presentation

Analysis results will be presented as described in Section 4.2.1.4.1.

4.2.3.5 Supplemental Analyses of the Secondary Binary Endpoints

Supplementary analyses will be performed on the binary secondary endpoints using a Treatment Policy strategy, similar to the supplementary analysis for the primary endpoint as described in Section 4.2.1.6 for the Week 10 endpoints.

An additional supplementary analysis will be performed on each Week 54 endpoint (sustained clinical remission and CS-free remission), in which the same analysis as in Section 4.2.1.4 will be performed on the MPP using the Induction/Maintenance treatment groups.

4.2.3.6 Subgroup Analyses

No subgroup analysis will be performed.

4.2.4 Secondary Endpoint: exposure-response model

An exposure-response model will link the primary endpoint to metrics of model-predicted individual brazikumab exposures. This will be described in the PK SAP.

4.2.5 Exploratory Endpoint: IL-22 cutoff determination

4.2.5.1 Definition

IL-22 is a cytokine produced by immune cells at sites of inflammation. The objective is the exploration of relationship of Baseline serum IL-22 concentration with efficacy of brazikumab at Week 10, and establishment of the serum IL-22 concentration clinical cutoff to stratify patients. The defined cutoff will be used to classify patients into BM+ and BM- strata.

4.2.5.2 Derivations

Biomarker status (BM+ or BM-) will be defined using IL-22 and clinical remission data at Week 10. The IL-22 concentration cutoff will be selected as the value that maximizes the difference in clinical remission rates (defined in Section 4.2.1.1) between brazikumab and placebo when comparing the BM+ and BM- subpopulations for the pooled doses of brazikumab vs placebo. This cutoff will be algorithmically derived using normalized Z-scores across the population of all possible cutoffs using the differential effects method (See Appendix 7.6.

The BM cut-off will be determined using the differential effect method based on Week 10 results and pooled treatment groups. This cut-off will be used for the study analyses and reporting described by this SAP. A final cutoff to be used for phase 3 will be selected outside of the study reporting, considering totality of evidence including other data and non-statistical aspects.

4.2.5.3 Handling of Dropouts and Missing Data

The IL-22 cutoff will be derived from clinical remission at Week 10 (Section 4.2.1.2). Handling of missing mMS component data is described in Section 4.2.1.3.

The Baseline IL-22 value will be used to determine the cutoff value. If IL-22 is missing at Baseline, that observation will not be used in the analysis.

4.2.5.4 Primary Analysis of Exploratory Endpoint

The differential effect method (see Appendix 7.6) will be applied for clinical remission at Week 10 for each of the brazikumab doses and for the pooled brazikumab doses versus placebo for the Induction treatment group. All analyses will use the FAS. Prevalence of the BM+ for each scenario will be provided. The cutoff that will be used for all BM subgroup analyses, as reported in the CSR, will be that derived at Week 10 using pooled brazikumab doses.

4.2.5.4.1 Presentation

Graphs describing the relationship between baseline IL-22 (on a continuous scale) vs clinical remission (as a z-score) will be produced.

In addition, descriptive summaries with the proportions of responders for clinical remission based on Week 10 at the selected cutoff value of IL-22 for each of the brazikumab doses by Induction treatment group and for the pooled brazikumab doses versus placebo will be produced.

4.2.6 Exploratory Endpoint: Binary Endpoints

- Clinical remission at Week 10 in participants by BM status
- Sustained clinical remission in participants by BM status
- CS-free clinical remission in participants who took corticosteroids at Baseline, by BM status
- Clinical response at Weeks 10 participants by BM status
- Endoscopic improvement at Week 54
- Endoscopic improvement at Weeks 10 in participants by BM status
- Clinical remission at Week 54 overall and in participants by BM status
- Symptomatic remission at Week 10
- Symptomatic remission at Week 54
- Endoscopic remission at Week 54

4.2.6.1 Definition

mMS and the estimand framework are defined in Section 4.2.1.1. Baseline is defined in Section 3.3.1.2. The BM cutoff value will be determined as described in Section 4.2.4.

Clinical remission: Clinical remission is derived in Section 4.2.1.2. It will be calculated using the FAS for Week 10 by BM status and Induction treatment group.

Sustained clinical remission: Sustained clinical remission is derived in Section 4.2.3.1. It will be calculated using the FAS, by BM status and Induction treatment group.

CS-free clinical remission: CS-free clinical remission is derived in Section 4.2.3.2. It will be calculated from the subpopulation with CS use at Baseline using the FAS, by BM status and Induction treatment group.

4.2.6.2 Derivation

Endoscopic improvement at Week 54: Endoscopic improvement is derived in Section 4.2.3.2. It is calculated using the FAS for Week 54 by Induction treatment group.

Symptomatic remission: rectal bleeding = 0 and stool frequency = 0 or 1, and at least a 1-point decrease from baseline. It will be calculated for the FAS at Week 10 and at Week 54.

Endoscopic remission: a centrally read endoscopy subscore of 0 that will be calculated at Week 54 for the FAS.

4.2.6.3 Handling of Dropouts and Missing Data

The exploratory binary endpoints will be handled similarly to the secondary binary endpoints as described in Section 4.2.3.3.

4.2.6.4 Primary Analysis of Exploratory Endpoints

The binary exploratory endpoints will be analysed for the FAS using CMH for each time point as defined in Section 4.2.1.4.

4.2.6.4.1 Presentation

The number and percentage of patients will be summarized by visit and by Induction treatment group. CMH results will be presented as described in Section 4.2.1.4.1, for the appropriate time point.

The proportion of subjects achieving clinical remission at each time point will be presented graphically by BM status. Additionally, percentage of CS-users will be graphed over time by treatment group.

4.2.6.5 Supplementary analysis

Each endpoint calculated at Week 54 will also be analysed for the MPP using CMH and Induction/Maintenance treatment groups and presented as defined in Section 4.2.6.4.1. The presentation for the MPP will be overall only.

In addition, an alternate definition of clinical response at Week 10 will be analysed. mMS including PGA is defined in Appendix 7.1, and the alternate measure is defined as: reduction

of ≥ 3 points in mMS from Baseline AND ≥ 30% reduction from Baseline AND decrease in rectal bleeding score of ≥ 1 from Baseline OR a score of 0 or 1.

4.2.7 Exploratory Endpoint: Change from Baseline in mMS

4.2.7.1 Definition

The endpoint is Change from Baseline in the mMS total score (with a range of 0-9) for each scheduled mMS calculation (Week 10 and Week 54).

mMS total score is defined in Section 4.2.1.2.

4.2.7.2 Derivation

For each Week, Change from Baseline will be calculated as defined in Section 3.3.1.2.

4.2.7.3 Handling of Dropouts and Missing Data

Handling of missing mMS components is described in Section 4.2.1.3. If an mMS assessment is missing for a particular Week, then the Change from Baseline will be missing for that Week.

4.2.7.4 Primary Analysis of Exploratory Endpoints

Using the FAS and Induction treatment groups, ANCOVA will be used to model Change from Baseline in the mMS and its component subscores at Week 10 and Week 54, with the stratification variable (prior biologic use) as a covariate.

The model is parameterized as:

$$Y_{ij} = \mu_i + \alpha_i + \beta\left(X_{ij} - \bar{X}_{..}\right) + \varepsilon_{ij}$$

Where $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$,

the $Y_{ij} \sim N(\mu_i + \beta\left(X_{ij} - \bar{X}_{..}\right), \sigma^2)$ are independent and $\sum \alpha_i = 0$.

4.2.7.4.1 Presentation

For the overall mMS and the component subscores (rectal bleeding, stool frequency, and endoscopy score), least square (LS) means, standard error (COVTEST option), and 95% CIs will be presented by Induction treatment group, as well as the LS mean difference comparisons between brazikumab and placebo, and the standard error and 95% CI of the difference, overall and by BM.

4.2.8 Exploratory Endpoint: Change from Baseline in CS use

4.2.8.1 Definition

The endpoint is the proportion of participants with CS use (Y/N) for each visit Week (Weeks 2, 4, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, and 54). CS use is defined as any corticosteroid use during the 7 days prior to each Visit.

4.2.8.2 Derivation

For each Induction treatment group and using the FAS, CS use will be the proportion of participants taking any form of CS.

4.2.8.3 Handling of Dropouts and Missing Data

If there is no record of a Visit for a given Week, then CS usage will be set to missing. If CS usage is missing, then that participant will not be included in the numerator or the denominator of percentage using corticosteroids for the Week.

4.2.8.4 Primary Analysis of Exploratory Endpoints

Proportion of participants with CS use will be calculated, along with change from Baseline and percentage change from Baseline.

4.2.8.4.1 Presentation

The proportion of participants using CS at each Visit, change from Baseline, and percentage change from Baseline will be presented using descriptive statistics for each treatment group. Additionally, a graphic of percentage of CS-users over time with means and CI will be presented by Induction treatment group.

4.2.9 Exploratory Endpoint: Mucosal Inflammation through Week 54

4.2.9.1 Definition

Geboes Histological Index (GHI) is a measure of the degree of mucosal inflammation as follows: 0 = structural change only; 1 = chronic inflammation; 2a = lamina propria eosinophils; 2b = lamina propria neutrophils; 3 = neutrophils in epithelium; 4 = crypt destruction; and 5 = erosions or ulcers. The Geboe's score ranges from 0 to 5.4. The Robarts Histopathology Index (RHI) is calculated based on information from the GHI and contains 4 histological descriptors: severity of chronic inflammatory infiltrate, the number of lamina propria neutrophils, the number of neutrophils in the epithelium, and the severity of erosions or ulceration. Table 4-5 lists the GHI scores and their translation to RHI scores.

Table 4-5 GHI grading and conversion to RHI (Geboes, et al, 2000).

Grade	*Subgrades*	*Conversion to RHI*
Grade 0: structural changes	0.0 No abnormality 0.1 Mild abnormality	N/A

	0.2 Mild or moderate diffuse or multifocal abnormalities 0.3 Severe diffuse or multifocal abnormalities	
Grade 1: chronic inflammatory infiltrate	1.0 No increase 1.1 Mild but unequivocal increase 1.2 Moderate increase 1.3 Marked increase	0 1 2 3
Grade 2A: eosinophils	2A.0 No increase 2A.1 Mild but unequivocal increase 2A.2 Moderate increase 2A.3 Marked increase	N/A
Grade 2B: neutrophils	2B.0 None 2B.1 Mild but unequivocal increase 2B.2 Moderate increase 2B.3 Marked increase	0 1 2 3
Grade 3: neutrophiles in epithelium	3.0 None 3.1 < 5% crypts involved 3.2 < 50% crypts involved 3.3 > 50% crypts involved	0 1 2 3
Grade 4: crypt destruction	4.0 None 4.1 Probable- local excess of neutrophils in part of the crypt 4.2 Probable – marked attenuation 4.3 Unequivocal crypt destruction	
Grade 5: erosion or ulceration	5.0 No erosion, ulceration, or granulation tissue 5.1 Recovering epithelium + adjacent inflammation 5.2 Probable erosion, focally stripped 5.3 Unequivocal erosion 5.4 Ulcer or granulation tissue	0 1 1 2 3

The total RHI is calculated as follows in Table 4-6:

Table 4-6 RHI calculation

Descriptor	***Responses***	***Scoring metric***
Severity of chronic inflammatory infiltrate	0= none 1= mild but unequivocal increase 2= moderate increase 3= marked increase	× 1
Lamina propria neutrophils	0= none 1= mild but unequivocal increase 2= moderate increase 3= marked increase	× 2
Neutrophils in epithelium	0= none	× 3

	1= < 5% crypts involved 2= <50% crypts involved 3= >50% crypts involved	
Erosion or ulceration	0= no erosion, ulceration of granulation tissue 1= recovering epithelium + adjacent inflammation 1= probable erosion – focally stripped 2= unequivocal erosion 3= ulcer or granulation tissue	× 5
	Total	Σ

4.2.9.2 Derivation

The biopsies that are obtained during the endoscopy at Screening, Week 10, and Week 54 will be scored according to the GHI by the central pathologist as 0, 1, 2a, 2b, 3, 4, or 5.

Each component of the RHI will also be graded between 0 and 3 based on information from the GHI. Using the scoring table above, the RHI can be calculated as: 1 × chronic inflammatory infiltrate level + 2 × lamina propria neutrophils + 3 × neutrophils in epithelium + 5 × erosion or ulceration, yielding a score that ranges from 0 (no disease activity) to 33 (severe disease activity) (Mosli, et al, 2015).

4.2.9.3 Handling of Dropouts and Missing Data

The GHI and RHI are based on pathology scoring from the Screening, Week 10 and Week 54 endoscopies. There will be no imputation. If the Week 10 or Week 54 endoscopy is missing, the endoscopy score will be set to missing.

Missing data will be handled in the mixed model. Missing at Random (MAR) is assumed. Mixed models for repeated measures (MMRM) does not employ formal imputation; rather, it uses all available data (including participants with partial data) to estimate the mean treatment effect in an unbiased manner.

4.2.9.4 Primary Analysis of Exploratory Endpoints

GHI will be described descriptively as a categorical variable (1, 2A, 2B, 3, 4, 5) by Induction treatment group.

RHI will be analysed as a continuous variable using MMRM, as described in Section 4.2.7.4, using the Induction treatment groups.

Using the FAS and Induction treatment groups, MMRM will be used to generate the least squares (LS) mean and 95% CI for each Week's endpoint, assuming unstructured covariance matrix. If use of the covariance matrix does not converge, repeat using compound symmetry (then AR(1), Toeplitz, Heterogeneous Toeplitz, and Heterogeneous Compound Symmetric,

in that order). The MMRM will only be run if there are at least 10 data points per treatment group.

For subjects i=1, …, I and repeated observations j=1,…,J, the MMRM model can be parameterized as

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + Z_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i$$

Where

$\mathbf{Y}_i$ = a J_i dimensional vector of outcome measurements for the i^{th} subject,

$\boldsymbol{\beta}$ = a p dimensional vector containing the fixed effects (baseline, treatment effect, time)

X_i, Z_i = ($J_i \times p$) and ($J_i \times q$) dimensional design matrices of known covariates

$\mathbf{b}_i$ = q dimensional vector containing the random effects ($\mathbf{b}_i \sim N(0, D)$)

$\boldsymbol{\varepsilon_i}$ = J_i dimensional vector of residual components ($\boldsymbol{\varepsilon_i} \sim N(0, \Sigma_i)$)

D = a general ($q \times q$) covariance matrix with the (i,j) element $d_{ij} = d_{ji}$

Σ_i = a ($J_i \times J_i$) covariance matrix which depends on i only through its dimension J_i

The MMRM model will include fixed effects for Induction treatment group, time point, and Induction treatment group-time point interaction. Covariates will include the baseline value of RHI, and the stratification variable (prior biologic use) as factors. The Kenward and Roger method will be used to calculate the denominator degrees of freedom for the test of fixed effects. All visits with a scheduled assessment through Week 54 will be included in the model.

4.2.9.4.1 Presentation

GHI will be presented with descriptive statistics overall and by BM status, by Induction treatment group, for each Grade. In addition, the number and percent in each subgrade will be presented for the MPP by Induction/Maintenance treatment groups.

The RHI Change from Baseline at each Visit will be presented using descriptive statistics for each score. Least square (LS) means, standard error (COVTEST option), and 95% CIs will be presented by Induction treatment group, as well as the LS mean difference comparisons between brazikumab and placebo, and the standard error and 95% CI of the difference will be presented.

GHI and RHI total and subscores will be presented in a listing.

4.2.9.5 Supplementary analysis

As a supplementary analysis, the MMRM described in Section 4.2.9.4 will be repeated for the MPP and Induction/Maintenance treatment groups.

4.2.9.6 Subgroup analysis

The analysis described in Section 4.2.9.4 will be performed by Baseline serum IL-22 (High/BM+, Low/BM-), with the cut-point as determined in Section 4.2.4.

4.2.10 Exploratory Endpoints: Change in Signs and Symptoms (BMD, Evening Diary, FACIT-F)

4.2.10.1 Definition

The exploratory endpoints are Change from Baseline through Week 54 in signs and symptoms derived from the Bowel Movement Diary, Evening Diary, and Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F). Frequency and type of stool (stool, blood, mucus), stool consistency (using BSFS) and urgency are collected from the BMD. Worst abdominal pain, worst fatigue, worst tiredness, worst weakness, worst lack of energy, worst joint pain, and frequency/severity of flatulence, are collected from the Evening eDiary (CSP Appendix G1 and G2). BMD and Evening Diary scores are evaluated at Weeks 2, 4, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, and 54. For visits where no endoscopy is performed, averages will be generated within the 7 days prior to each Visit. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep). The FACIT-F is assessed on-site during Weeks 0, 6, 10, 18, 30, 42, and 54.

The FACIT-F is a 13-item instrument that measures fatigue in chronic illness patients. See CSP Appendix G 3.3.

BMD items

Average stool frequency is derived from the number of entries in the BMD, including any BM entries entered in the Evening Diary.

Stool type includes the responses Stool, Blood, and Mucus on question E02 (Table 7-2). For instance, for a given day, it is possible that stool =7, blood=1, and mucus=2 and these will be averaged separately.

Stool consistency will be derived from question E04 (Table 7-2).

Urgency will be derived from question E03 (Table 7-2).

Evening diary

The Evening Diary contains 7 numeric rating scale (NRS) items with responses from 0 (none) to 10 (as bad as I can imagine), in which the participant rates the symptom at its worst in the past 24 hours:

- Abdominal pain
- Fatigue
- Tiredness
- Weakness
- Lack of energy
- Joint pain
- Flatulence

In addition, frequency of flatulence is measured on a 5-point scale ranging from 0 (never) to 4 (all of the time). (Spiegel et al, 2009).

FACIT-F

The FACIT-F Scale (Version 4) is a 13-item instrument that measures fatigue and its impact on daily functions over a recall period of 7 days. Five of the items assess the experience of fatigue, and 8 items assess the impact of fatigue. Items are scored on a 5-point verbal rating scale, yielding a score ranging from 0 to 52, with lower scores indicating greater fatigue (see Appendix 7.4 and CSP Appendix G 3.3).

4.2.10.2 Derivation

Baseline and Change from Baseline are defined in Section 3.3.1.2.

Average stool frequency is the number of entries in the BMD in the 3 most recent (consecutive or non-consecutive) days out of the 7 days prior to the Visit, divided by 3. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep). The endpoint is Change from Baseline.

For each stool type Stool, Blood, and Mucus, the percentage for a day is calculated by dividing the number of responses (Stool=[1], Blood=[2], Mucus=[3]) by the total number of stools for the day, multiplied by 100.

Percentage of stools with consistency 6 or 7 for each day will be number of responses to question E04 (Table 7-2) that equal 6 or 7, divided by the number of stools that day multiplied by 100. The average will be taken from the 3 most recent (consecutive or non-consecutive) days out of the 7 days prior to the Visit. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep). The endpoint is Change from Baseline.

Urgency for the day is calculated as the number of stools in which the answer to "Did you experience urgency (a sudden, almost irresistible need) to pass this bowel movement?" = [1]. The percentage for a day is calculated by dividing the number of urgent stools by the total number of stools for the day, multiplied by 100. The average will be taken from the 3 most recent (consecutive or non-consecutive) days out of the 7 days prior to the Visit. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep). The endpoint is Change from Baseline.

Evening diary NRS responses (abdominal pain, fatigue, tiredness, weakness, lack of energy, joint pain, and flatulence) range from 0 to 10. Calculate the average score for each Visit by summing the scores for the 3 most recent (consecutive or non-consecutive) days of the last 7 days prior to the respective Visits, and dividing by 3. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep).

Frequency of flatulence is captured in the Evening Diary and measured on a 5-point scale. Calculate the average score across the 3 most recent (consecutive or non-consecutive) days of the last 7 days prior to each Visit. For visits where endoscopy is performed, averages will be generated within the 7 days of endoscopy, exclusive of days of endoscopy and bowel prep (i.e., during the 5 days prior to bowel prep).

FACIT-F is completed on site, recalling the past 7 days. It is scored by reversing all items except AN5 and AN7, adding the available item scores, multiplying by 13, and divided by the number of items answered. (FACIT-F.org). The endpoint is Change from Baseline.

4.2.10.3 Handling of Dropouts and Missing Data

Evening Diary items

If there are not 3 (consecutive or non-consecutive) days of data available for the week preceding a Visit (or in the 5 days prior to bowel prep for visits where endoscopy is performed), the window will be expanded as described for mMS PRO items in Section 4.2.1.3. If 3 (consecutive or non-consecutive) days of data are still not available, the value will be set to missing and their data will not be included in the summary table.

NRS items

In the case of missing days in the diary, data from the completed days will be averaged and analysed. If there are not 3 (consecutive or non-consecutive) days of data available for the week preceding a Visit (or in the 5 days prior to bowel prep for visits where endoscopy is performed), the window will be expanded as described for mMS PRO items in Section 4.2.1.3. If 3 (consecutive or non-consecutive) days of data are still not available, the NRS score will be set to missing.

FACIT-F Fatigue

If there are missing items on the FACIT-F, and there are at least 50% of items completed, the subscore can be imputed (Webster, Cella, and Yost 2003). Multiply the sum of the responses by the number of items (13), then divide by the number of items answered:

score = [(sum of item scores) * (13)] / (number of items answered)

If more than 50% of the items are missing, then then FACIT-F score will be set to missing. Missing data will be handled in the mixed model.

4.2.10.4 Primary Analysis of Exploratory Endpoints

The average score over the 3 most recent (consecutive or non-consecutive) days from the 7 days prior to each Visit, or in the 5 days prior to bowel prep for visits where endoscopy is performed, will be calculated for each endpoint and the Change from Baseline calculated for each Visit for each score. Each measure will be analysed separately using a mixed model as described in Section 4.2.9.4.

4.2.10.4.1 Presentation

The Change from Baseline at each Visit will be presented using descriptive statistics for each score. Mixed model results will be presented as in Section 4.2.9.4.1.

4.2.11 Exploratory Endpoints: Health Related Quality of Life (HRQoL) (IBDQ, SF-36, EQ-5D-5L)

4.2.11.1 Definition

Inflammatory Bowel Disease Questionnaire (IBDQ)

The IBDQ is presented in Appendix I 3.4 of the CSP, and measures Health Related Quality of Life (HRQoL) in patients with inflammatory bowel disease (IBD). It will be assessed during the site visit at Weeks 0, 6, 10,18, 30, 42, 54, and ET or ED. It covers how the participant has been feeling in the past two weeks (Guyatt et al, 1989). Four dimensions are covered: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items). Items are scored on a 7-point Likert scale (from 1 [worse] to 7 [better]). The total score is obtained by summing the score from each item, yielding a global score in the range 32 to 224 (with higher scores indicating better quality of life).

EQ-5D-5L

The EQ-5D-5L is presented in Appendix G 3.5 of the CSP and will be assessed at Baseline and during site visits at Weeks 0, 10, 30, and 54. The EQ-5D-5L includes a VAS that allows respondents to rate their own current health on a 101-point scale ranging from “best imaginable” (100) to “worst imaginable” (0) health.

SF-36v2

The SF-36v2 (SF36) is presented in Appendix I 3.6 of the CSP and will be assessed at Baseline and during site visits at Weeks 0, 10, 30, and 54. It is a standardized instrument used to measure self-reports of health status and functional well-being. It consists of 8 domains: physical functioning (10 questions, #3a to 3j), role physical (4 questions, #4a to 4b), bodily pain (2 questions, #7 and 8), general health (5 questions, #1,11a to 11d), vitality (4 questions, #9a, 9e, 9g, 9i), social functioning (2 questions, #6, #10), role emotional (3 questions, #5a to 5c), and mental health (5 questions, #9b to 9d, 9f, 9h). The subscores from the 8 dimensions can be directly calculated. Two summary indices can be scored: the Physical Component Score (PCS) and the Mental Component Score (MCS). The lower the score, the more disability (Brazier, Roberts, and Deverill, 2002).

4.2.11.2 Derivation

IBDQ

The total score is obtained by summing each item and will range from 32 to 224. Change from Baseline is calculated as described in Section 3.3.1.2. A positive Change from Baseline value indicates increased quality of life compared to Baseline.

The mapping of items to the 4 subscores is found in Table 4-7 along with the maximum score per dimension, calculated by summing the response to each item in the subscore.

Table 4-7 IBDQ subscores

Subscore	*Item numbers*	*Max points*
Bowel symptoms	1,5,9,13,17,20,22,24,46,29	70
Systemic symptoms	2,6,10,14,18	35
Emotional function	3,7,11,15,19,21,13,15,17,30,31,32	84
Social function	4,8,12,16,28	35

Each of the 4 IBDQ domain scores is calculated by summing the responses from the items in that domain, as defined in Table 4-7. The average per item scores can be calculated by dividing the each subscore by the number of items in the subscore.

EQ-5D-5L

Change from Baseline is calculated for the 101-point scale. A negative Change from Baseline value indicates a worse health state.

SF-36v2

The SF36 will be scored using the algorithm provided by Optum with the instrument license (Health Outcomes™ Scoring Software 4.5 or higher). Change from Baseline is calculated for the Physical and Mental component subscores (PCS and MCS).

4.2.11.3 Handling of Dropouts and Missing Data

IBDQ, EQ-5D-5L, and SF36 will be administered at the site. If a participant discontinues early or misses a Visit, the scores for that Visit will be set to missing and the mixed model will use the non-missing data to estimate the means.

If more than half of the items in a subscore are missing for the IBDQ, the total score will be set to missing as well as that subscore. Otherwise, the mean of the remaining items in the subscore can be imputed for that subscore (Rubin et al, 2004).

Per the EQ-5D-5L user guide (EQ-5D-5L User Guide, 2019), missing values should be coded as '9', or '999' for the VAS. These should be set to missing prior to analysis.

The SF36 is scored by averaging items in the same subscale. If a question is missing values will be imputed according to the algorithm provided by Optum.

4.2.11.4 Primary Analysis of Exploratory Endpoints

Using the FAS, Change from Baseline at each Week will be calculated for each Induction treatment group from the score for each Week. Change from Baseline for each endpoint (IBDQ, IBDQ subscores, EQ-5D-5L VAS, SF36 PCS, and SF36 MCS) will be analysed using a mixed model with the Induction treatment group and stratification variable (prior biologic use) as factors and baseline value as a covariate, as described in Section 4.2.9.4.

The 5 domains of the EQ-5D-5L will not be analysed.

4.2.11.4.1 Presentation

Absolute values and change from Baseline in IBDQ, EQ-5D-5L VAS, SF36 PCS, and SF36 MCS will be presented using descriptive statistics by visit and treatment for the FAS.

For each mixed model, results will be presented as described in Section 4.2.9.4.1, as well as descriptive statistics for each score.

4.2.11.5 Subgroup analysis

The analysis described in Section 4.2.9.4 will be performed by Baseline serum IL-22 (High/BM+, Low/BM-), with the cut-point as determined in Section 4.2.4.

4.3 Pharmacodynamic Endpoints

4.3.1 Definition

Pharmacodynamic (PD) variables collected for this study include IL-22 as measured by the R&D Systems assay, serum C-reactive protein (CRP), and fecal calprotectin (FCP).

4.3.2 Handling of Dropouts and Missing Data

There will be no imputation of missing data. If any IL-22, CRP, or FCP are missing for a Visit, then Change from Baseline and percentage Change from Baseline will be set to missing and not be used in the analysis.

4.3.3 Primary Analysis of Exploratory Endpoints

Analysis of IL-22 is described in Section 4.2.5.
Change from Baseline for each PD parameter will be analysed with an MMRM model with Induction treatment group and timepoint as factors and Baseline value and the stratification variable (prior biologic use) as covariates, as described in Section 4.2.9.4.

4.3.3.1 Presentation

Individual PD parameters will be listed with Change from Baseline and percentage Change from Baseline by Induction treatment group. Absolute values, Change from Baseline, and percentage Change from Baseline will be summarized by Induction treatment group and BM status using descriptive statistics.
For each mixed model, results will be presented as described in Section 4.2.9.4.1.

Mean absolute values, mean Change from Baseline, and mean percentage Change from Baseline will also be presented graphically by Induction treatment group.

4.3.4 Subgroup analysis

The analysis described in Section 4.2.9.4 will be performed by Baseline serum IL-22 (High/BM+, Low/BM-) for the three PD parameters, with the cut-point as determined in Section 4.2.4.

4.4 Pharmacokinetics

Serum samples for determination of brazikumab concentrations and the presence of anti-drug antibodies (ADA) will be collected at Baseline prior to first study intervention administration, at multiple time points during the Double-Blind Treatment Period, and at the end of the follow-up period. All analysis of PK variables will be based on the PK analysis set defined in Section 3.2.
Individual brazikumab serum concentrations for each nominal sampling time will be listed and summary descriptive statistics will be presented and plotted versus time. Individual

brazikumab serum concentrations will be graphically illustrated and summarized by ADA status.

For each time point, the geometric mean, geometric standard deviation, coefficient of variation (CV), arithmetic mean, SD, median, minimum, and maximum will be reported as summary statistics.

Predose PK blood samples will be collected at Weeks 0, 2, 10 14, and 30. Data will be assigned to analysis windows as defined in Section 3.3.2.

Linear and log-linear geometric mean (± SE) concentration profile plots against time will be produced for each treatment arm, with nominal sampling time on the x-axis.

Further model-based evaluation of PK will be detailed in a separate PK analysis plan and reported outside the CSR.

4.5 Immunogenicity

Serum samples to measure the presence of anti-drug antibodies (ADAs) against brazikumab (binding antibodies and neutralizing antibodies) will be collected at Week 0, 6, 10, 30, and 54. ADA results from each sample will be reported as either positive or negative. If the sample is positive, the ADA titer will be reported as well. In addition, the presence of nAb will be tested for all ADA-positive samples.

4.5.1 Derivations

The number and percentage of ADA-evaluable participants in the following ADA categories in each treatment group and overall will be determined. The number of ADA-evaluable participants in the treatment group will be used as the denominator for percentage calculation.

- ADA positive at any visit (at baseline and/or post-baseline). The percentage of these participants in a population is known as ADA prevalence.
- Treatment-emergent ADA positive (sum of treatment-induced ADA positive and treatment-boosted ADA positive). The percentage of these participants in a population is known as ADA incidence.
- Treatment-induced ADA positive (ADA negative at baseline and ADA positive at >=1 post-baseline assessment).
- Treatment-boosted ADA positive (baseline ADA titer that was boosted by ≥4-fold following OLE study intervention administration).
- Treatment-emergent ADA negative (ADA positive but not fulfilling the definition of TE-ADA positive).
- ADA positive at baseline and post-baseline.
- ADA positive at baseline and not detected post-baseline.

- Treatment-emergent ADA (TE-ADA) persistently positive, defined as treatment-emergent ADA positive participants having at least 2 post-baseline ADA positive measurements at least 16 weeks (112 days) between the first and last positive measurement, or an ADA positive result at the last available assessment.
- Treatment-emergent ADA (TE-ADA) transiently positive is defined as treatment-emergent ADA positive participants not fulfilling the conditions for TE-ADA persistently positive.
- Neutralizing Antibody (nAb) positive is defined as nAb positive at any time baseline and/or post-baseline (nAb prevalence).
- Treatment-induced nAb positive (nAb incidence) is defined as nAb negative at baseline (or ADA negative at baseline) and nAb positive at >=1 post-baseline assessment.
- Treatment-emergent ADA positive with maximum titre > median of maximum titre (calculated based on the maximum titre for each ADA positive subject within each treatment group).

4.5.2 Primary analysis of ADA

A summary of the number and percentage of participants who developed positive ADA in different treatment groups and overall will be presented based on the ADA evaluable analysis set. In addition, descriptive statistics of the maximum titre will be presented by ADA category. ADA results and titers for each ADA category (Section 4.5.1) will be summarized by visit, treatment group and overall.
Additional immunogenicity outcomes may be described in separate analysis plans and will be reported separately.

4.6 Safety Analyses

Safety data will be summarized descriptively using the Safety Population, which includes participants who enrolled prior to and after the initiation of Amendment 5. Pre-Amendment 5 participants randomized to the vedolizumab group will be listed only. Participants in the pre-Amendment 5 high dose brazikumab group (2100 mg) will be included in the total (grouped) brazikumab summaries but will not be summarized separately. The other Pre-Amendment 5 participants will be grouped with their corresponding post-Amendment 5 dose group and included in safety summary tables.
All safety data will be summarized by Treatment Period. Induction Period tables will be presented by Induction treatment group. Maintenance Period tables will be presented by Induction/Maintenance treatment group.
The safety parameters will include AEs, clinical laboratory, vital signs, physical examinations, and ECG parameters. Continuous variables will be summarized by the number

of patients and mean, SD, median, minimum, and maximum values. Categorical variables will be summarized by number and percentage of patients.

4.6.1 Exposure

4.6.1.1 Definitions and Derivations

Exposure to the study intervention will be calculated as the number of days from the first study intervention to 28 days after the last study intervention for the study for the Safety Population. Any amount of exposure to brazikumab in the placebo group will be counted as a brazikumab exposure. Any gaps in dosing will be ignored when calculating the total.

Total dose will be calculated as the IV doses given during the Induction Period (from the form Exposure- Intravenous Infusion Bag) during Week 0, 2, and 6, plus the SC doses given during the Maintenance Period through Week 50 (from the form Exposure- Subcutaneous).

4.6.1.2 Presentation

Exposure to the study intervention (in days) for the Safety Population during the Induction and Maintenance Treatment Periods will be summarized by Induction and Induction/ Maintenance treatment groups, respectively. The total dose will be summarized by Induction/Maintenance treatment group for the Double-blind Treatment Period. Participants who did not enter the Maintenance Treatment Period and are therefore without an Induction/Maintenance treatment category, will be summarized by Induction treatment group only.

4.6.2 Adverse Events

Adverse events will be coded by system organ class and preferred term using the MedDRA, version 24.0 or newer.

4.6.2.1 Definitions and Derivations

Adverse event data will be categorized according to their onset date based on the last dose of investigational product as follows:

- An AE occurring during the Treatment Period is defined as an AE with a date of onset ≥ the first dose of investigational product and ≤ 28 days after the last dose of investigational product
- An AE occurring during the Study Period is defined as an AE that occurs during treatment including follow-up, with a date of onset ≥ the first dose of investigational product and ≤ 18 weeks after the last dose of investigational product

Per case report form instructions, a new AE record will be created for any AE that worsens.

Treatment related

An AE will be considered as treatment related if the investigator considers that there is a reasonable possibility that the event may have been caused by the investigational product.

Serious

An AE will be considered a serious AE (SAE) if it is an AE that additionally meets any SAE criteria.

Missing Severity Assessment

Adverse events with missing severity are assumed to be severe. The imputed values for severity assessment will be used for the incidence summary; the values will be shown as missing in the data listings.

Missing Causal Relationship to Study Intervention

There will be no imputation of missing causal relationship for any non-serious Aes. For SAEs, if the causal relationship to the study intervention is missing for an SAE that started on or after the date of the first dose of the study intervention, a causality of 'yes' will be assigned.

The imputed values for causal relationship during the Treatment Period will be used for the incidence summary; the values will the shown as missing in the data listings.

Missing Date Information for Adverse Events

See Appendix 7.5 for missing date imputation rules.

4.6.2.2 Presentation

For the Safety population, summary tables will be presented separately for the Induction Period using Induction treatment groups, and for the Maintenance Period using the Induction/Maintenance treatment groups. The term "treatment group" will refer to the above standard for the respective summaries. Each table will also include brazikumab total.

For the Safety analysis set, the following events will be summarized by treatment group:

- Overall summary of AE, including
    - Any AE
    - Any SAE
    - Any SAE with outcome death
    - Any AE leading to discontinuation of IP
    - Any possibly related AE
    - Any possibly related SAE
- AE by SOC and PT
- AE by maximum reported intensity on preferred term level
- Possibly related AE by SOC and PT

- AE per 100 subject years by SOC and PT
- Adverse events of special interest (AESI)
- SAE with outcome death by SOC and PT
- SAE by SOC and PT
- AE leading to discontinuation of investigational product by SOC and PT

The AESI summary will include infusion related reactions, injection-site reactions, malignancies, hypersensitivity (including anaphylaxis), and infections.

The incidence of common (> 5% of patients in either 720 mg IV, 1440 mg IV, 120 mg SC or 240 mg SC treatment groups) non-serious AEs will be summarized by system organ class, preferred term, and treatment group.

Listings of AEs, deaths, SAEs and AEs leading to discontinuation of the investigational product by participant will be presented, to include key subject information, including Subject identifier, sex, age, event term as reported, AE (PT), time from first dose to AE, treatment period, last dose prior to event, time from first dose to <death or discontinuation>, whether study intervention was received for the AE, action taken with respect to study intervention, and AE relationship to study intervention.

If more than 1 AE is coded to the same preferred term for the same participant, the participant will be counted only once for that preferred term using the most severe severity for the summaries by severity. Similarly, the participant will be counted using the closest relationship for summaries by causality.

4.6.3 Clinical Laboratory, Blood Sample

4.6.3.1 Definitions and Derivations

The following are protocol-required laboratory parameters as listed in Table 4-8 through Table 4-9. All laboratory assessments will be done at Screening. Clinical chemistry and hematology will be carried out at Weeks 0, 4, 10, 30, 54, unscheduled visits and ED or ET. CRP will be assessed at Weeks 0, 6, 10, 30, 54, unscheduled visits and ED or ET.

Table 4-8 Protocol-Required Safety Laboratory Assessments

Laboratory Assessments	***Parameters***			
Hematology	Platelet count	RBC indices: MCV MCH % Reticulocytes		WBC count with differential: Neutrophils Lymphocytes Monocytes Eosinophils Basophils
	RBC count			
	Hemoglobin			
	Hematocrit aPTT/PT/INR (reflexively as needed)			
Clinical Chemistry	BUN C-reactive protein	Potassium	AST	Total and direct bilirubin
	Creatinine eGFR	Sodium	ALT	Total protein
	Glucose (nonfasting)	Calcium	Alkaline phosphatase	Bicarbonate
	Magnesium	Chloride	Albumin	Phosphate
	Uric acid			
Screening serology	• Serum hCG			
Fecal tests	• *Clostridium difficile* • Fecal calprotectin • Fecal lactoferrin			
Routine urinalysis	• Specific gravity • pH, glucose, protein, blood, ketones, bilirubin, urobilinogen, nitrite, leukocyte esterase by dipstick • Microscopic examination (reflexively as needed)			

ALT = alanine aminotransferase; aPTT = activated partial thromboplastin time; AST = aspartate aminotransferase; BUN = blood urea nitrogen; eGFR = estimated glomerular filtration rate;hCG = human chorionic gonadotropin HIV = human immunodeficiency virus; IgM = immunoglobulin M; INR = international normalized ratio; MCH = mean corpuscular hemoglobin; MCV = mean corpuscular volume; PT = prothrombin time; RBC = red blood cell; WBC = white blood cell.

Table 4-9 Normal ranges for laboratory values (Central Laboratory Services Manual)

Laboratory Parameter	***LLN - Male***	***ULN- Male***	***LLN – Female***	***ULN- Female***	***SI Unit***
Hematology					
Hemoglobin	Age 12-59: 127 Age >59: 125	Age 12-59: 181 Age >59:170	Age 12-59: 116 Age >59: 115	Age 12-59: 164 Age >59: 158	g/L
Hematocrit	Age 12-59: 0.39 Age >59: 0.37	Age 12-59: 0.54 Age >59: 0.51	0.34	0.48	ratio
RBC	Age 12-59: 4.5 Age >59: 4.0	Age 12-59: 6.4 Age >59: 5.8	Age 12-59: 4.1 Age >59: 3.9	Age 12-59: 5.6 Age >59: 5.5	TI/L
Eosinophils	0.00	0.57	0.00	0.57	GI/L

Laboratory Parameter	***LLN - Male***	***ULN- Male***	***LLN – Female***	***ULN- Female***	***SI Unit***
Neutrophils	40.5	75.0	40.5	75.0	%
Basophils	0.00	0.20	0.00	0.20	GI/L
Monocytes	0.12	0.92	0.12	0.92	GI/L
Lymphocytes	Age 18-59: 0.91 Age >59: 0.80	Age 18-59: 4.28 Age >59:3.00	Age 18-59: 0.91 Age >59: 0.80	Age 18-59: 4.28 Age >59: 3.00	GI/L
Absolute neutrophil count	1.96	7.23	1.96	7.23	GI/L
Platelet count	Age 12-60: 140 Age > 60: 130	Age 12-60: 400 Age > 60: 394	Age 12-60: 140 Age > 60: 130	Age 12-60: 400 Age > 60: 394	10^9/L
Albumin	Age 18-69: 33 Age >69: 33	Age 18-69: 49 Age >69: 46	Age 18-69: 33 Age >69: 33	Age 18-69: 49 Age >69: 46	g/L
Alkaline phosphatase	Age 18-19: 55 Age >19: 40	Age 18-19: 149 Age > 19: 129	Age 18-19: 45 Age > 19: 35	Age 18-19: 87 Age > 19: 104	U/L
Alanine aminotransferase (ALT)	Age 18-69: 6 Age > 69: 6	Age 18-69: 43 Age > 69: 35	Age 18-69: 6 Age > 69: 6	Age 18-69: 34 Age > 69: 32	U/L
Aspartate aminotransferase (AST)	11	36	9	34	U/L
Gamma-glutamyl transferase (GGT)	Age 18-59: 10 Age >59: 10	Age 18-59: 61 Age >59: 50	Age 18-59: 4 Age >59: 5	Age 18-59: 49 Age >59: 50	U/L
Blood urea nitrogen or Urea	1.4	Age 18-70: 8.6 Age >70: 10.4	1.4	Age 18-70: 8.6 Age >70: 10.4	mmol/dL
Calcium	2.07	2.64	2.07	2.64	mmol/dL
Chloride	94	112	94	112	mmol/L
Creatinine	40	Age 18-50: 110 Age 50-70: 119 Age >70: 137	31	Age 18-70: 101 Age >70: 110	µmol/L
Glucose, nonfasting	3.9	5.6	3.9	5.6	mmol/L
Magnesium	Age 18-20: 0.70 Age 20-60: 0.66 Age >60: 0.66	Age 18-20: 0.90 Age 20-60: 1.07 Age >60: 0.98	Age 18-20: 0.70 Age 20-60: 0.66 Age >60: 0.66	Age 18-20: 0.90 Age 20-60: 1.07 Age >60: 0.98	mmol/L
Potassium	3.5	5.2	3.5	5.2	mmol/L
Sodium	Age 18-59: 132 Age >59: 135	Age 18-59: 147 Age >59: 145	Age 18-59: 132 Age >59: 135	Age 18-59: 147 Age >59: 145	mmol/L
Total bilirubin	3	21	3	21	µmol/L

Laboratory Parameter	*LLN - Male*	*ULN- Male*	*LLN – Female*	*ULN- Female*	*SI Unit*
Total protein	Age 18-59: 61 Age >59: 60	Age 18-59: 84 Age >59: 80	Age 18-59: 61 Age >59: 60	Age 18-59: 61 Age >59: 60	g/L
Uric acid or Urate	Age 18-50: 125 Age > 50: 149	Age 18-50: 488 Age >50: 494	Age 18-50: 125 Age >50: 149	Age 18-50: 428 Age >50: 446	mg/dL

Blood urea nitrogen or Urea are the same parameters, and Uric acid or Urate are the same parameters.
LLN = lower limit of normal; SI = *Le Système International d'Unités* (International System of Units; ULN = upper limit of normal laboratory reference range.

Table 4-10 Criteria for Potentially Clinically Significant Laboratory Values

Laboratory Parameter	*SI Unit*	*Conversion Factor*[a]	*Conventional Unit*	*PCS Criterion*[b] *Low Value*	*PCS Criterion*[b] *High Value*
Hematology					
Hemoglobin	g/L	0.1	g/dL	< 0.9 × LLN	
Hematocrit	Volume fraction	100	%	< 0.9 × LLN	
Eosinophils	%	1	%	—	> 10
Neutrophils	%	1	%	< 30	> 90
Basophils	%	1	%	—	> 6
Monocytes	%	1	%	—	> 20
Lymphocytes	%	1	%	< 10	> 60
Absolute neutrophil count	$\times 10^9$/L	1	1000/µL	< 1.0	—
Platelet count	$\times 10^9$/L	1	1000/µL	≤ 75	≥ 700
White blood cell count with differential	$\times 10^9$/L	1	1000/µL	≤ 2.5	≥ 15
Chemistry					
Albumin	g/L	0.1	g/dL	< 0.9 × LLN	> 1.1 × ULN
Alkaline phosphatase	U/L	1	U/L	—	≥ 3 × ULN
Alanine aminotransferase (ALT)	U/L	1	U/L	—	≥ 3 × ULN
Aspartate aminotransferase (AST)	U/L	1	U/L	—	≥ 3 × ULN
Gamma-glutamyl transferase (GGT)	U/L	1	U/L	—	≥ 3 × ULN
Blood urea nitrogen or Urea	mmol/L	2.8011	mg/dL	—	> 1.2 × ULN
Calcium	mmol/L	4.008	mg/dL	< 0.9 × LLN	> 1.1 × ULN
Chloride	mmol/L	1	mg/dL	< 0.9 × LLN	> 1.1 × ULN
Creatinine	µmol/L	0.0113	mg/dL	—	> 1.3 × ULN
Glucose, nonfasting	mmol/L	18.018	mg/dL	< 0.8 × LLN	> 1.2 × ULN
Magnesium	mmol/L	2	mEq/L	< 0.9 × LLN	> 1.1 × ULN
Potassium	mmol/L	1	mEq/L	< 0.9 × LLN	> 1.1 × ULN

Laboratory Parameter	*SI Unit*	*Conversion Factor*[a]	*Conventional Unit*	*PCS Criterion*[b] *Low Value*	*PCS Criterion*[b] *High Value*
Sodium	mmol/L	1	mEq/L	< 0.9 × LLN	> 1.1 × ULN
Total bilirubin	µmol/L	0.0585	mg/dL	—	> 1.5 × ULN
Total protein	g/L	0.1	g/dL	< 0.9 × LLN	> 1.1 × ULN
Uric acid or Urate	µmol/L	0.0168	mg/dL	—	> 1.1 × ULN

[a] Conversion factor from SI units to conventional (traditional) units.
[b] Criteria refer to SI units.
Blood urea nitrogen or Urea are the same parameters, and Uric acid or Urate are the same parameters.
LLN = lower limit of normal; PCS = potentially clinically significant; SI = *Le Système International d'Unités* (International System of Units); ULN = upper limit of normal laboratory reference range.

Clinical laboratory test values will be considered potentially clinically significant (PCS) if they meet either the lower-limit or higher-limit PCS criteria listed in Table 4-10 (using the same periods as defined for AEs in Section 4.6.2.1).

4.6.3.2 Presentations

Baseline, postbaseline, and Change from Baseline for Hematology and clinical chemistry parameters will be summarized separately for Induction and Maintenance Periods by treatment group for the Safety population, as described in Section 4.6.2.2.

Descriptive statistics for clinical laboratory values (in SI units) at baseline, postbaseline, and Change from Baseline values at each postbaseline timepoint will be listed and summarized by treatment group. If the reported value of a clinical laboratory parameter cannot be used in a statistical summary table because, for example, a character string is reported for a parameter of the numeric type, a coded value must be appropriately determined for use in the statistical analyses. The actual values, however, as reported in the database will be presented in the data listings.

The number and percentage of patients who have PCS postbaseline clinical laboratory values will be listed and tabulated by treatment group for the Induction and Maintenance Periods. The percentages will be calculated relative to the number of patients with available baseline values and at least 1 postbaseline assessment for the Treatment Period. The numerator will be the total number of patients with available baseline values and at least 1 PCS postbaseline value for the Treatment Period. A supportive tabular display of key subject information of participants with PCS postbaseline values will be provided, including the PID number, baseline and all postbaseline (including non-PCS) values.

Chemistry lab results will be similarly listed and summarized with descriptive statistics.

4.6.4 Clinical Laboratory, Urinalysis

4.6.4.1 Definitions and Derivations

Urinalysis will be carried out at Screening and at Weeks 0, 4, 10, 30, 54, unscheduled visit and ED or ET. Routine urinalysis includes specific gravity, glucose, protein, ketones,

bilirubin, urobilinogen, nitrite, and leukocyte esterase, all of which are considered a high value if 2+. In addition, pH normal range is ≥ 5 to ≤ 8. Microscopic examination is performed as needed.

4.6.4.2 Presentations

Listing of urinalysis results will be provided for the Safety analysis set. A shift table from Baseline to the end of the Treatment Period will be presented by treatment group. Treatment emergent urinalysis abnormalities will be summarized by treatment group.

4.6.5 Other Laboratory Evaluations

4.6.5.1 Definitions and Derivations

Table 4-11 presents the other laboratory testing.

Table 4-11 Other laboratory testing

- QuantiFERON-TB Gold In-Tube
- FSH (as needed in women of non-childbearing potential only)
- Serum hCG pregnancy test (as needed for women of childbearing potential); urine hCG should be performed locally at each visit, except Week 4
- Serology (HIV antibody, hepatitis B surface antigen, hepatitis B core total antibody, hepatitis B core IgM antibody, and hepatitis C virus antibody with reflex testing when required)
- All study-required laboratory assessments will be performed by a central laboratory unless indicated (eg, local urine hCG testing)

FSH = follicle stimulating hormone; hCG = human chorionic gonadotropic; SoA = schedule of activities; HIV = human immunodeficiency virus

4.6.5.2 Presentations

Positive pregnancy and other screening tests will be listed.

4.6.6 Physical Examinations

4.6.6.1 Definitions and Derivations

A complete physical examination will include assessments of the respiratory, cardiovascular, gastrointestinal, and neurological systems as well as assessment of EIM. Physical examinations will be performed at Weeks 0, 2, 6, 10, 18, 30, 42, 54, and unscheduled as well as E/T or ED. A listing of the date of physical examinations will be provided.

Any new findings or aggravated existing abnormalities, judged as clinically significant by the investigator, will be reported as an AE.

4.6.7 Vital Signs

4.6.7.1 Definitions and Derivations

Vital signs will be collected at every Visit and include:

- Body temperature

- Pulse rate (sitting)
- Respiratory rate
- Blood pressure (sitting)
- Height (Screening Visit 1 only) recorded in centimeters
- Weight, recorded in kilograms

Vital sign values will be considered PCS if they meet both the observed value criterion and the Change from Baseline value criterion, if both criteria are available, or meet either the observed value criterion or the Change from Baseline value criterion detailed in Table 4-12.

Table 4-12 Criteria for Potentially Clinically Significant Vital Signs

Parameter	***Flag***	***Criteria[a]***	
		Observed Value	*Change from Baseline*
Sitting systolic blood pressure, mm Hg	High	≥ 140	Increase of ≥ 20
	Low	<90	Decrease of ≥ 20
Sitting diastolic blood pressure, mm Hg	High	≥ 90	Increase of ≥ 15
	Low	<60	Decrease of ≥ 15
Sitting pulse rate, bpm	High	≥ 100	Increase of ≥ 15
	Low	≤ 60	Decrease of ≥ 15
Weight, kg	High	—	Increase of ≥ 7%
	Low	—	Decrease of ≥ 7%

[a] A postbaseline value is considered potentially clinically significant if it meets both the observed-value and the change-from-baseline criteria.
bpm = beats per minute; hG = Mercury

4.6.7.2 Presentations

A listing and summary of descriptive statistics for vital signs (systolic and diastolic blood pressures, pulse rate, body temperature, respiratory rate, and weight) values at Baseline, postbaseline, and Changes from Baseline values at each postbaseline timepoint will be presented by treatment group (as described in Section 4.6.2.2) for the Safety population for the Induction and Maintenance Periods.

The number and percentage of patients who have PCS postbaseline vital sign values will be listed and tabulated by treatment for each assessment for each Treatment Period. The percentages will be calculated relative to the number of participants who have available baseline values and at least 1 postbaseline assessment for the Treatment Period. The numerator will be the total number of participants with at least 1 PCS postbaseline value for the Treatment Period. A supportive listing of Key Subject Information for participants with PCS postbaseline values will be provided.

4.6.8 Electrocardiogram

4.6.8.1 Definitions and Derivations

Single 12-lead ECGs will be obtained during Screening and Weeks 0, 10, 30, 54, and E/T visits using an ECG machine that automatically calculates the heart rate and measures PR, QRS, QT, and QTc intervals. The upper limit for each measure is shown in Table 4-13.

Table 4-13 Criteria for Potentially Clinically Significant Electrocardiograms

Parameter	*Unit*	*Low Value*	*Low Decrease*	*High Value*	*High Increase*
ECG heart rate	bpm	≤ 60	≥ 15	≥ 100	≥15
RR interval	msec	< 500	NA	> 1500	NA
PR interval	msec	NA	NA	≥ 240	NA
QRS duration	msec	≤ 60	NA	≥ 140	NA
QT	msec	≤ 300	NA	> 450	> 60
QTcF	msec	≤ 300	NA	> 480	> 60
QTcB	msec	≤ 300	NA	> 480	> 60

QTc = QT interval corrected for heart rate.
QTcB = QT interval corrected for heart rate using the Bazett formula.
QTcF = QT interval corrected for heart rate using the Fridericia formula.

The outcome of the ECG evaluation is to be recorded as normal or abnormal, with any abnormalities recorded as either not clinically significant, or clinically significant.

4.6.8.2 Presentations

A listing and summary of descriptive statistics for ECG parameters (i.e., heart rate, PR interval, QRS interval, QT interval, and QTc interval) at Baseline, post baseline, and Change from Baseline values at Week 54 will be presented by treatment group for the Safety analysis set.

ECG parameter values are considered PCS if ECG values meet the value listed in Table 4-13. The number and percentage of participants with PCS post baseline values will be tabulated by treatment group. The percentages will be calculated relative to the number of participants with an available non-PCS Baseline value and at least one post baseline assessment.

The numerator will be the total number of participants with an available non-PCS Baseline value and at least one PCS post baseline ECG value.

A supportive listing of participants with PCS post baseline values will be provided and will include the participant number and the Baseline and post baseline values.

A listing of all AEs for participants with PCS ECG values will also be provided.

4.6.9 Other Safety Assessments

4.6.9.1 Definitions and Derivations

Potential Hy's law is defined as AST ≥ 3 × ULN or ALT ≥ 3 × ULN and total bilirubin ≥ 2 × ULN, where at least one time total Bilirubin ≥ 2 × ULN occurred after the first occurrence of ALT ≥ 3 × ULN or AST ≥ 3 × ULN after start of treatment with the investigational product. Hy's Law is defined as ALT or AST ≥ 3 × ULN together with total bilirubin ≥ 2 × ULN, where for no other reason, other than study intervention, can be found to explain the combination of increases (e.g., elevated alkaline phosphatase indicating cholestasis, viral hepatitis, or another drug).

More information can be found in the CSP, Appendix E.

4.6.9.2 Presentations

Participants with elevated liver tests based on measured laboratory values will be summarized for the Safety population. Participants who meet the potential Hy's Law criteria from the first dose of the study intervention to the end of the Induction Period, and in the Maintenance Period to within 18 weeks after the last dose of the study intervention will be listed and summarized by treatment group for the Safety population as described in Section 4.6.2.2.

5 INTERIM/PRIMARY ANALYSIS

Details on how trial integrity will be maintained is documented in a separate Study Integrity Plan.

Interim Analysis (IA), as per ICH E9, is an analysis prior to the time point of the formal completion of the trial. The primary analyses in study D5272C00001 will be performed when all participants have completed their Week 10 visit or discontinued early. Accordingly, the pre-specified primary analysis fulfils the ICH E9 definition of an interim analysis. These analyses are described in Section 4.2.

There will be 2 clinical database locks in the study, one at the time of the primary analyses, and one after the last participant has completed the second safety follow-up visit or discontinued early.

All data collected, including Week 54 will be analysed at the time of the primary analyses. All available data from participants who have completed Week 54, or would have completed Week 54 if they hadn't discontinued, will be analysed.

All analyses that will be performed beyond Week 10 at the time of interim (primary) analyses are exploratory and should be interpreted with caution.

6 REFERENCES

Best WR, Becktel JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. Gastroenterology 1976; 70(3):439-44.

Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002 Mar;21(2):271-92. doi: 10.1016/s0167-6296(01)00130-8. PMID: 11939242.

EuroQol Research Foundation. EQ-5D-5L User Guide, 2019. Available from: https://euroqol.org/publications/user-guides.

FACIT group. Available at: https://www.facit.org/measures/FACIT-F (Accessed 22 November 2022).

Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Lofberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut. 2000 Sep:47(3):404-409.

Guidance for Industry Ulcerative Colitis: Developing Drugs for Treatment (April 2022).

Guo Y, Wu V Li X, Xu X, and Cheng C. An Illustration of Rate Differences Estimation with SAS in Logistic Regression. 2012. PharmaSUG China 1st Conference.

Guyatt G, Mitchell A, Irvine EJ, et al. A New Measure of Health Status for Clinical Trials in Inflammatory Bowel Disease. Gastroenterology 1989;96:804-10.

Han SW, McCollE, Steen N, Barton JR, Welfare MR. The Inflammatory Bowel Disease Questionnaire: a valid and reliable measure in ulcerative colitis patients in the North East of England. Scand J Gastroenterol 1998;33:961-966.

Irvine EJ. Development and subsequent refinement of the inflammatory bowel disease questionnaire: a quality-of-life instrument for adult patients with inflammatory bowel disease. J Pediatr Gastroenterol Nutr. 1999 Apr;28(4):S23-7.

Loftus EV, Feagan BG, Colombel JF, Rubin DT, Wu EQ, Pollack PF, Chao J, Mulani P. Effects of adalimumab maintenance therapy on health-related quality of life of patients with Crohn's disease: patient-reported outcomes of the CHARM trial. Am J Gastroenterol 2008; 103:3132-3141.

Mosli MH, Feagan BG, Zou G, Sandborn WJ, D'Haens G, Khanna R et al. Development and vadlidation of a histological index for UC. Gut. 2015;0:1-9. doi:10.1136/gutjnl-2015-310393.

Rubin GP, Hungin APS, Chinn DJ, Dwarakanath D. Quality of life in patients with established inflammatory bowel disease: a UK general practice survey. Aliment Pharmacol Ther 2004; 19:529-535.

Sandborn, WJ, Vermeire S, Sands, BE, et al. Defining Remission Using Patient-reported Outcome Components of the Crohn's Disease Activity Index: Post-hoc Analyses of Vedolizumab Data from the GEMINI 2 Study, American Journal of Gastroenterology: October 2015 - Volume 110 - Issue - p S792-S793.

Sands BE, Ooi CJ. A survey of methodological variation in the Crohn's disease activity index. Inflamm Bowel Dis. 2005; 11:133-38.

Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid 548 therapy for mildly to moderately active ulcerative colitis. N Engl J Med. 1987;317:1625-1629.

Spiegel B, Bolus R, Harris LA, et al. Measuring irritable bowel syndrome patient-reported outcomes with an abdominal pain numeric rating scale. *Aliment Pharmacol Ther*. 2009;30(11-12):1159-1170. doi:10.1111/j.1365-2036.2009.04144.x.

Stokes ME, Davis CS, Kock GG. Categorical Data Analysis Using SAS. Third Edition, 2012. SAS Institute, Inc, Cary, North Carolina, USA.

Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes*. 2003;1:79. Published 2003 Dec 16. doi:10.1186/1477-7525-1-79.

7 APPENDIX

7.1 Modified Mayo Score

Patient Reported Outcomes Questionnaire: mMS was removed due to copyrights.

7.2 Bowel Movement Diary

Patient Reported Outcomes Questionnaire: Bowel Movement Diary was removed due to copyrights.

Patient Reported Outcomes Questionnaire: Bowel Movement Diary was removed due to copyrights.

7.3 Evening eDiary

Patient Reported Outcomes Questionnaire: Evening eDiary was removed due to copyrights.

Patient Reported Outcomes Questionnaire: Evening eDiary was removed due to copyrights.

7.4 FACIT-F Scale

Patient Reported Outcomes Questionnaire: FACIT-F Scale was removed due to copyrights.

Patient Reported Outcomes Questionnaire: FACIT-F Scale was removed due to copyrights.

7.5 Missing start and stop dates

For adverse events and prior or concomitant medications, including rescue medications, incomplete (i.e., partly missing) start dates and/or stop dates will be imputed. When the start date and the stop date are both incomplete for a participant, the start date will be imputed first.

7.5.1 Incomplete Start Date

The following rules will be applied to impute the missing numeric fields for an incomplete prior or concomitant medication start date. If the stop date is complete (or imputed) and the imputed start date is after the stop date, the start date will be imputed using the stop date.

If the start date is missing, then it will be set to the date of the first study intervention, unless the end date suggests that the start date could be prior to the date of the first study intervention. In that case, set the start date to January 1 of the same year as the end date.

Missing month and day
The start date will be set to January 1 of the year of occurrence.

If the start year is the same as the year of the first study intervention, then the start date will be set to the date of first study intervention.

Missing month only
If only the month is missing, the day will be treated as missing and both the month and the day will be replaced according to the above procedure.

Missing day only
The missing start day will be set to the first day of the month of occurrence, if the start month and year is after the month and year of the first study intervention.

If the month and year of the incomplete start date are the same as the month and year of the first dose of the study intervention, the day of the first dose of the study intervention will be assigned to the missing day.

If the month and year of the incomplete start date is before the month and year of the first study intervention, the date of informed consent will be assigned to the incomplete date.

7.5.2 Incomplete Stop Date

Because participants will receive their double-blind study intervention doses under the direct supervision of study centre personnel, the date of dose of study intervention will always be known. The following rules will be applied to impute missing date fields for an incomplete AE, prior, or concomitant medication stop date. If the imputed stop date is before the start date (imputed or non-imputed start date), the imputed stop date will be equal to the start date.

If the end date is missing and the event is not recorded as ongoing, then set the end date as the date of the first study intervention, if the start date is prior to the date of the first study intervention. If the start date is on or after the date of first study intervention, set the end date to the date of last visit.

Missing month and day

If the year of occurrence is the same as the last study intervention year, then the incomplete date will be set to the date of last study intervention. If the year of the incomplete stop date is the same as the year of the last dose of study treatment, the month and day of the last dose of study treatment will be assigned to the missing fields.

If the year of the incomplete stop date is before the year of the last dose of study treatment, *December 31* of the year of occurrence will be assigned to the missing fields.

The stop date will be set to December 31 of the year of occurrence. If the participant died in the same year, the stop date will be set to the date of death.

Missing month only

If only the month is missing, the day will be treated as missing and both the month and the day will be replaced as described in Section 7.5.1.

Missing day only

If the month and year of the incomplete stop date are after the month and year of the first study intervention, then the missing end day will be set to the last day of the month of the occurrence.

If the participant died, the end date will be set to the death date.

If the month and year of the incomplete stop date are the same as the month and year of the last dose of study intervention, the day of the last dose of study intervention will be assigned to the missing day.

If either the year of the incomplete stop date is before the year of the date of the last dose of study intervention or if both years are the same, but the month of the incomplete stop date is before the month of the date of the last dose of study intervention, the last day of the month will be assigned to the missing day.

If either the year of the incomplete stop date is after the year of the date of the last dose of study intervention or if both years are the same, but the month of the incomplete stop date is after the month of the date of the last dose of study intervention, the first day of the month will be assigned to the missing day.

7.6 Differential effect method

David Svensson and Fredrik Ohrn. A technical note on determining a biomarker split for differentiating treatment effect in Subpopulations. February 2021

1 Introduction

In this document we generically discuss an approach that could be useful in a RCT setting when a subgroup with enhanced treatment effect is desired to be found in a data-driven fashion. There is a vast literature on the general case where baseline data for multiple covarriates is available, but we here restrict ourselves to the special case of a single biomarker. This is not so uncommon in the practice, and the question then is if there is some simpler and transparent approach that doesn't rely on Machine Learning and furthermore has reasonable theoretical properties and is relatively non-parametric (in the sense of not assuming any particular functional form).

We here describe one such idea in mathematical terms. The setting we have in mind is rather general but assume for now the endpoint Y_i to be binary, let T_i denote treatment allocation $\in \{0,1\}$ and x_i the biomarker; here, $i = 1, ..., n$ denotes subjects in the trial. In the standard primary analysis for an overall treatment effect, we assume that some statistical model M is applied to the full population, rendering an estimate $\hat{\theta}$, with associated standard error $\hat{s}$. (Since our biomarker splitting approach is generic and doesn't rely on any particular statistical modelling for the primary analysis, we simply illustrate the key concepts with a standard test for a difference in proportions).

Assume, without loss of generality, that the range of the biomarker x in the trial is $R = (0,1)$ (hence, possibly scaled), and assume furthermore that any possible value $h \in R$ is considered as a candidate split. This would render two candidate subgroups $S_h^a = \{x : x < h\}$ and $S_h^b = \{x : x \geq h\}$. In each of these sub-populations, the model M can be applied, rendering an estimate and the corresponding standard error. Denote these with $\hat{\theta}_a^{(h)}$, $\hat{\theta}_b^{(h)}$, $\hat{s(h)}_a$ and $\hat{s(h)}_b$. To assess how well the split differentiate in terms of treatment effect hetereogeneity,

we suggest the simple metric

$$g(h) = \frac{\hat{\theta}_b^{(h)} - \hat{\theta}_a^{(h)}}{\hat{s}_\Delta}.$$

where $\hat{s}_\Delta^2$ is the variance of the nominator; since the data in $\{x : x < h\}$ and $\{x : x \geq h\}$ is mutually exclusive, independence holds and

$$\hat{s}_\Delta^2 = s(\hat{h})_a^2 + s(\hat{h})_b^2$$

The split h that maximizes the differential effects is now found by

$$argmax_h|\{g(h)\}.|$$

If instead a directional analysis is desired (under an assumption of monotonicity of effect with x), the absolute value can be removed and h is chosen accounting for the underlying directional assumption.

This approach maximizes the differential effects, while penalizing splits where one of the subgroups is small in which case the term $\hat{s}_\Delta^2$ would tend to increase, hence avoiding results with a large point estimate but a very large standard error.

So, this idea simply uses the difference between the effect sizes, but normalizes by the standard error of that difference. (Note that the variance of the difference is the sum of the variances of the two subgroup estimates, since the data is mutually exclusive; i.e., due to independence)

In practice, if the candidate split is too close to the edges, one of the subgroups will have very little data, and possibly only data from one arm is present. To protect numerically against such artefacts, we suggest restricting the analysis to a subset $[a, b]$, $0 < a < b < 1$, where a and b are such that for $h \in [a, b]$, there is sufficient data in both arms to fit the model M for any subgroups resulting from the cut h (to avoid numerical instabilities).

Moreover, in the bulk of the data range, where estimates and standard errors are defined, $g(h)$ will decrease if one of the estimates $\hat{\theta}_a^{(h)}$ and $\hat{\theta}_b^{(h)}$ is uncertain due one of the subgroups being small.

The advantages with this simple non-parametric approach is that it is straightforward to implement, it avoids splits in the extreme ends, and relies on a well known principle in the literature on causal effects and ITE estimation, namely looking for partitions of the data that maximizes the differential effect. This principle underlies many modern ML methods in this area, such as SIDES [1], GUIDE [2], and QUINT [3]. Similarly, discovering splits that maximizes the differential effects was explored in a spline setting in [4].

As a final remark, it is standard in ITE literature to optimize splits with respect to differential effects rather than to directly optimize the effect in one of the subgroup nodes; the latter approach (optimizing only wrt $S(h)_b$) risk to result in low prevalence subgroups with high uncertainty in the effect estimate (unless some manual arbitrary lower bounds are imposed in the search), whereas the former approach balance out this, see e.g., [1].

References

[1] Lipkovich I., Dmitrienko A., Denne J., Enas G. *Subgroup Identification based on Differential Effect Search (SIDES) a recursive partitioning method for establishing response to treatment in patient subpopulations.* Stat. In. Medicine, 2011.

[2] Loh W., He X., Man M. *A Regression Tree Approach to Identifying Subgroups with Differential Treatment Effects Statistics in Medicine.* January 2014.

[3] Dusseldorp E., Mechelen I. *Qualitative interaction trees: a tool to identify qualitative treatment subgroup interactions.* Statistics in Medicine, 2014, 33(2).

[4] Thomas M., Bornkamp B. *Comparing Approaches to Treatment Effect Estimation for Subgroups in Clinical Trials.* Statistics in Biopharmaceutical Research, 9:2, 2017.

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: d5272c00001-sap-ed-3		
Document Title:	Statistical Analysis Plan Edition 3	
Document ID:	Doc ID-004424330	
Version Label:	3.0 CURRENT LATEST APPROVED	
Server Date (dd-MMM-yyyy HH:mm 'UTC'Z)	**Signed by**	**Meaning of Signature**
19-Dec-2022 17:22 UTC	PPD	Content Approval
19-Dec-2022 14:38 UTC	PPD PPD	Author Approval

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