

STATISTICAL ANALYSIS PLAN

APD334-301

A Phase 3, Randomized, Double-Blind, Placebo-Controlled, 52-Week Study to Assess the Efficacy and Safety of Etrasimod in Subjects with Moderately to Severely Active Ulcerative Colitis

Author: [REDACTED]

Version Number and Date: V1.0 Final, 19 January 2022

Document: [REDACTED]

Author: [REDACTED]

Version Number: 1.0 Final

Version Date: 19 January 2022

Template No.: CS_TP_BS016 Revision 5

Reference: CS_WI_BS005

Effective Date: 01 April 2018

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V1.0 (dated 19 January 2022) for Clinical Study Protocol APD334-301 Amendment 4.0

	Name	Signature	Date
Author:	██████████		
Position:	Senior Biostatistician		
Company:	IQVIA		

Upon review of this document, the undersigned approves this version of the Statistical Analysis Plan, authorizing that the content is acceptable for the reporting of this study.

	Name	Signature	Date
Approved By:	██████████		
Position:	Associate Director, Biostatistics		
Company:	IQVIA		

Approved By:	██████████		
Position:	Vice President, Biostatistics & Data Management		
Company:	Arena Pharmaceuticals, Inc.		

Approved By:	██████████		
Position:	Senior Director, Biostatistics		
Company:	Arena Pharmaceuticals, Inc.		

MODIFICATION HISTORY

Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	19 January 2022	██████████	Not Applicable – First Version

TABLE OF CONTENTS

STATISTICAL ANALYSIS PLAN	1
STATISTICAL ANALYSIS PLAN SIGNATURE PAGE	1
MODIFICATION HISTORY	2
LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS	8
1. INTRODUCTION	12
2. STUDY OBJECTIVES	12
2.1. Primary Objective	12
2.2. Secondary Objectives	12
2.3. Safety Objectives	12
2.4. Other Objectives	12
2.5. Estimands	12
3. STUDY DESIGN	15
3.1. General Description	15
3.2. Schedules of Assessments	16
3.3. Changes to Analysis from Protocol	16
4. PLANNED ANALYSES	16
4.1. Interim Analysis	16
4.2. Final Analysis	17
5. ANALYSIS SETS	17
5.1. Screened Set	17
5.2. Randomized Set	17
5.3. Full Analysis Set	17
5.4. Per Protocol Sets	17
5.5. Modified Full Analysis Set	18
5.6. Safety Set	18
5.7. Pharmacokinetic Set	18
6. GENERAL CONSIDERATIONS	19
6.1. Reference Start Date and Study Day	19
6.2. Baseline	19
6.3. Retests, Unscheduled Visits, and Early Termination Data	19

6.4.	Windowing Conventions	19
6.5.	Statistical Tests	21
6.6.	Common Calculations	21
6.7.	General Study Information	22
7.	DETERMINATION OF SAMPLE SIZE	22
8.	STATISTICAL CONSIDERATIONS	23
8.1.	Multicenter Studies	23
8.2.	Adjustments for Covariates and Factors to be Included in Analyses	23
8.3.	Missing Data	24
8.4.	Multiple Comparisons and Multiplicity	24
8.5.	Subgroup Analyses	25
9.	OUTPUT PRESENTATIONS	26
10.	DISPOSITION AND PROTOCOL DEVIATIONS	27
11.	DEMOGRAPHIC AND BASELINE CHARACTERISTICS	28
11.1.	Derivations	30
12.	MEDICAL HISTORY	30
13.	MEDICATIONS	30
14.	STUDY TREATMENT EXPOSURE	30
14.1.	Derivations	31
15.	STUDY TREATMENT COMPLIANCE	31
15.1.	Derivations	31
16.	EFFICACY OUTCOMES	32
16.1.	Primary Efficacy	33
16.1.1.	Primary Efficacy Variables and Derivations	33
16.1.2.	Missing Data Methods for Primary Efficacy Variables	33
16.1.3.	Primary Analysis of Primary Efficacy Variables	34
16.1.4.	Sensitivity Analysis of Primary Efficacy Variables	35
16.1.4.1.	Multiple Imputation Under MAR	35
16.1.4.2.	Tipping Point Analysis	36
16.1.4.3.	Multiple Imputation with Copy Reference	37
16.1.4.4.	Multiple Imputation with NRI to Handle Missing Data Due to COVID-19 Pandemic	38

18.1.1.	All TEAEs	45
18.1.1.1.	Severity	45
18.1.1.2.	Relationship to Study Treatment	46
18.1.2.	TEAEs Leading to Discontinuation of Study Treatment.....	46
18.1.3.	Serious and Non-Serious TEAEs.....	46
18.1.4.	TEAEs Leading to Death.....	47
18.1.5.	TEAEs of Special Interest	47
18.1.6.	Overall Summary of Adverse Events	48
18.2.	Deaths	49
18.3.	Laboratory Evaluations.....	49
18.3.1.	Safety Laboratory Evaluations	49
18.3.2.	Pregnancy Tests.....	51
18.3.3.	Other Screening Laboratory Assessments	51
18.4.	ECG Evaluations	51
18.4.1.	ECG Markedly Abnormal Criteria	52
18.5.	Vital Signs	53
18.5.1.	Vital Signs Specific Derivations.....	54
18.5.2.	Vital Signs Markedly Abnormal Criteria	54
18.6.	Physical Examination	54
18.7.	Other Safety Assessments.....	54
18.7.1.	Pulmonary Function Tests	54
18.7.1.1.	PFT Markedly Abnormal and Potentially Important Criteria.....	55
18.7.2.	Ophthalmoscopy and Optical Coherence Tomography.....	55
18.7.3.	Tuberculosis Screening and Chest X-Ray	55
19.	EFFICACY-RELATED BIOMARKERS	56
20.	PHARMACOKINETICS	56
21.	DATA NOT SUMMARIZED OR PRESENTED	58
	REFERENCES	60
	APPENDICES	61

LIST OF APPENDICES

APPENDIX 1: PROGRAMMING CONVENTIONS FOR OUTPUTS61

APPENDIX 2: PARTIAL DATE CONVENTIONS63

APPENDIX 3: [REDACTED]63

APPENDIX 4: IBDQ SCORING RULES71

APPENDIX 5: SF-36 SCORING RULES72

APPENDIX 6: WPAI-UC SCORING RULES73

APPENDIX 7: CLASSIFICATION OF RESCUE THERAPY74

LIST OF TABLES

Table 1: List of Primary and Key Secondary Estimands13

Table 2: Visit Windows for Efficacy and Safety Analyses20

Table 3: Planned Analyses by Endpoint39

Table 4: Markedly Abnormal Criteria for Vital Signs54

LIST OF FIGURES

Figure 1: Study Design16

Figure 2: Gatekeeping Procedure Summary24

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation	Explanation
5-ASA	5-aminosalicylic acid
AE	adverse event
AESI	adverse events of special interest
ALC	absolute lymphocyte count
ALP	alkaline phosphatase
ALT	alanine aminotransferase
APD334	etrasimod
Arena	Arena Pharmaceuticals, Inc.
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AV	atrioventricular
β-hCG	beta-human chorionic gonadotropin
BLQ	below the limit of quantitation
BMI	body mass index
bpm	beats per minute
CI	confidence interval
CM	concomitant medication
CMH	Cochran-Mantel-Haenszel
COVID-19	coronavirus disease 2019
CR	copy reference
CRP	C-reactive protein
CS	clinically significant
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
C _{trough,ss}	average steady-state trough plasma concentration
CV	coefficient of variation
DBP	diastolic blood pressure
DLCO	diffusing capacity of the lungs for carbon monoxide
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram

Abbreviation	Explanation
eCRF	electronic case report form
eDISH	evaluation of drug-induced serious hepatotoxicity
EIM	extraintestinal manifestation
ES	endoscopic score
FAS	Full Analysis Set
FCP	fecal calprotectin
FCS	fully conditional specification
FDA	Food and Drug Administration
FEF	forced expiratory flow
FEV ₁	forced expiratory volume in 1 second
FVC	forced vital capacity
GGT	gamma-glutamyl transferase
HR	heart rate
HRQoL	health-related quality of life
hs-CRP	high-sensitivity C-reactive protein
IBDQ	Inflammatory Bowel Disease Questionnaire
ICF	informed consent form
JAK	Janus kinase
LLQ	lower limit of quantitation
LOESS	locally estimated scatterplot smoothing
LS	least squares
MAR	missing at random
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	modified Full Analysis Set
MMRM	mixed-effect model with repeated measures
MMS	modified Mayo score
MNAR	missing not at random
NCS	not clinically significant
NHI	Nancy Histological Index
NRI	nonresponder imputation

Abbreviation	Explanation
NRS	numeric rating scale
OCT	optical coherence tomography
OLE	open-label extension
PFT	pulmonary function test
PGA	Physician's Global Assessment
PK	pharmacokinetics
PRES	posterior reversible encephalopathy syndrome
PT	preferred term
RB	rectal bleeding
RHI	Robarts Histopathology Index
RNA	ribonucleic acid
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SDTM	Study Data Tabulation Model
SE	standard error
SF	stool frequency
SF-36	Medical Outcomes Study 36-Item Short Form Health Survey
SI	International System of Units
SOC	system organ class
ss	steady state
TB	tuberculosis
TEAE	treatment-emergent adverse event
TLC	total lung capacity
TME	targeted medical event
TMS	total Mayo score
TNF α	tumor necrosis factor alpha
UC	ulcerative colitis
ULN	upper limit of normal
ULQ	upper limit of quantitation

Abbreviation	Explanation
WHO	World Health Organization
WPAI-UC	Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis

1. INTRODUCTION

This Statistical Analysis Plan (SAP) describes the statistical rationale, methods, rules, and conventions to be used in the presentation and analysis of efficacy, safety, pharmacokinetic (PK), and efficacy-related biomarker data for Clinical Study Protocol APD334-301. It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed (ICH 1998). This SAP is based on the Clinical Study Protocol Amendment 4.0, dated 22 December 2020.

Analyses of the genetics (Clinical Study Protocol APD334-301 Section 9.9) [REDACTED] data will be specified in a separate biomarker analysis plan.

2. STUDY OBJECTIVES

2.1. Primary Objective

The primary objective is to assess the efficacy of etrasimod on clinical remission in subjects with moderately to severely active ulcerative colitis (UC) after 12 and 52 weeks of treatment.

2.2. Secondary Objectives

The secondary objective is to assess the efficacy of etrasimod on clinical response, symptomatic response and remission, endoscopic changes, corticosteroid-free remission, and mucosal healing in subjects with moderately to severely active UC at timepoints up to 52 weeks of treatment.

2.3. Safety Objectives

The safety objective is to assess the long-term safety of etrasimod after daily doses of 2 mg for up to 52 weeks in subjects with moderately to severely active UC.

2.4. Other Objectives

Other objectives include evaluation of etrasimod PK and the effect of etrasimod on health-related subject-reported outcomes and biomarkers.

2.5. Estimands

The primary and key secondary efficacy estimands to support regulatory decisions are described in Table 1. The analyses will be performed on the Full Analysis Set (FAS) as defined in Section 5.3. Other secondary efficacy variables (specified in Section 16.3.1) will have the population, intercurrent event handling strategy and population-level summary measure similar to the primary estimand. Supplementary analyses for the primary and the key secondary estimands will be performed on the modified Full Analysis Set (mFAS) and the Per Protocol populations, respectively (ICH 2019). Intercurrent events include: 1) initiate a rescue medication for UC, 2) have an increase in dose over Baseline levels in their existing UC medication, or 3) have a rescue non-drug treatment (eg, colectomy, ileostomy, or sigmoidectomy) before the

efficacy assessment. Refer to [APPENDIX 7](#) for the definition of rescue therapies (medication and medical procedures).

Table 1: List of Primary and Key Secondary Estimands

Estimand	Definition	Attribute			
		Population	Variable/Endpoint	Intercurrent Event Handling Strategy	Population-Level Summary Measure
Primary Estimand 1	Efficacy of etrasimod on clinical remission at Week 12	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with SF = 0 (or = 1 with a ≥ 1 -point decrease from Baseline), RB = 0, and ES ≤ 1 (excluding friability) at Week 12	Subjects with any of the 3 intercurrent events ^a before the efficacy assessment will be treated as nonresponders.	Difference between etrasimod and placebo in the endpoint at Week 12
Primary Estimand 2	Efficacy of etrasimod on clinical remission at Week 52	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with SF = 0 (or = 1 with a ≥ 1 -point decrease from Baseline), RB = 0, and ES ≤ 1 (excluding friability) at Week 52	Subjects with any of the 3 intercurrent events ^a before the efficacy assessment will be treated as nonresponders.	Difference between etrasimod and placebo in the endpoint at Week 52
Key Secondary Estimand 1	Efficacy of etrasimod on endoscopic improvement at Week 12	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with ES of ≤ 1 (excluding friability) at Week 12	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in the endpoint at Week 12
Key Secondary Estimand 2	Efficacy of etrasimod on endoscopic improvement at Week 52	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with ES of ≤ 1 (excluding friability) at Week 52	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in the endpoint at Week 52
Key Secondary Estimand 3	Efficacy of etrasimod on symptomatic remission at Week 12	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with SF = 0 (or = 1 with a ≥ 1 -point decrease from Baseline) and RB = 0 at Week 12.	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in the endpoint at Week 12

Estimand	Definition	Attribute			
		Population	Variable/Endpoint	Intercurrent Event Handling Strategy	Population-Level Summary Measure
Key Secondary Estimand 4	Efficacy of etrasimod on symptomatic remission at Week 52	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with SF = 0 (or = 1 with a ≥ 1 -point decrease from Baseline) and RB = 0 at Week 52.	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in proportion of responders at Week 52
Key Secondary Estimand 5	Efficacy of etrasimod on corticosteroid-free clinical remission at Week 52	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with clinical remission at Week 52 and who had not been receiving corticosteroids for ≥ 12 weeks prior to Week 52	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in proportion of responders at Week 52
Key Secondary Estimand 6	Efficacy of etrasimod on sustained clinical remission	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with sustained clinical remission (clinical remission at both Weeks 12 and 52)	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in proportion of responders
Key Secondary Estimand 7	Efficacy of etrasimod on mucosal healing at Week 12	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with ES of ≤ 1 (excluding friability) with histologic remission measured by a Geboes Index score < 2.0 at Week 12	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in proportion of responders at Week 12
Key Secondary Estimand 8	Efficacy of etrasimod on mucosal healing at Week 52	Randomized subjects who receive at least 1 dose of study treatment and with Baseline MMS 5 to 9	The proportion of subjects with ES of ≤ 1 (excluding friability) with histologic remission measured by a Geboes Index score < 2.0 at Week 52	Same as the intercurrent event handling strategy for the primary estimand	Difference between etrasimod and placebo in proportion of responders at Week 52

^a Intercurrent events include: 1) initiate a rescue medication for UC, 2) have an increase in dose over Baseline levels in their existing UC medication, or 3) have a rescue non-drug treatment (eg, colectomy, ileostomy, or sigmoidectomy) before the efficacy assessment. Refer to [Appendix 7](#) for the definition of rescue therapies (medication and medical procedures). ES, endoscopic score; RB, rectal bleeding; SF, stool frequency; UC, ulcerative colitis.

3. STUDY DESIGN

3.1. General Description

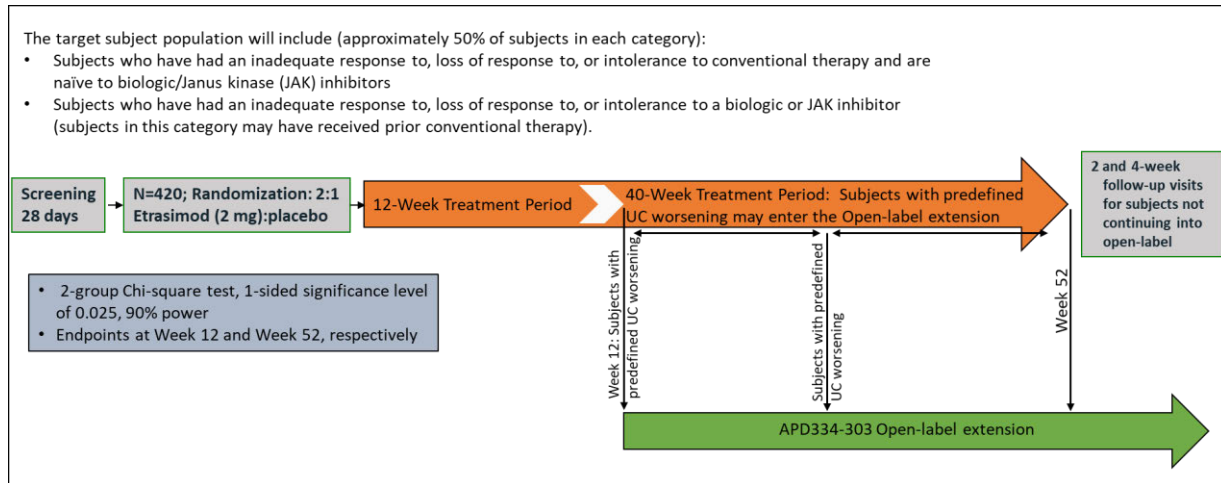
This is a multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of etrasimod 2 mg in subjects with moderately to severely active UC. The study consists of a 28-Day Screening Period, a 12-Week Treatment Period, a 40-Week Treatment Period, and a 2-Week and 4-Week Follow-Up Period. The study design diagram can be found in [Figure 1](#). Eligible subjects will be randomized (2:1 ratio) to receive either etrasimod (2 mg once daily) or matching placebo (once daily) in a double-blind fashion for up to 52 weeks of treatment. Randomization will be stratified by (a) naïve to biologic or Janus kinase (JAK) inhibitor therapy at study entry (Yes or No), (b) Baseline corticosteroid use (Yes or No), and (c) Baseline disease activity (modified Mayo score [MMS]: 4 to 6 or 7 to 9).

Approximately 420 subjects are planned to be enrolled. The target subject population will include (approximately 50% of subjects in each of the following categories):

1. Subjects who have had an inadequate response, loss of response, or intolerance to conventional therapy and are naïve to biologic or JAK inhibitor therapy
2. Subjects who have had an inadequate response, loss of response, or intolerance to a biologic or JAK inhibitor (subjects in this category may have received prior conventional therapy)

At the end of the 12-Week Treatment Period, subjects will undergo Week 12 efficacy and safety assessments and be evaluated for clinical response/remission as well as UC disease worsening. Subjects whose UC condition in the opinion of the Investigator has not improved or has worsened, compared with Baseline (Week 0/Day 1), as defined in Clinical Study Protocol APD334-301 Section 5.1.1 and who meet other eligibility criteria will have the option to enter an open-label extension (OLE) study (Study APD334-303). Subjects must complete Week 12 to be eligible for the OLE. Subjects who do not meet disease worsening criteria, including those demonstrating clinical response/remission at Week 12 will continue into the 40-Week Treatment Period and continue their double-blind treatment. Subjects who either experience disease worsening in the 40-Week Treatment Period or complete all study procedures at Week 52 will have the option to enroll into the OLE study if they meet all eligibility criteria. Subjects who do not participate in the OLE study will have 2-Week and 4-Week Follow-Up visits after their last treatment administration/Early Termination (ET) visit. The study design is outlined in [Figure 1](#).

Figure 1: Study Design



JAK, Janus kinase; UC, ulcerative colitis

3.2. Schedules of Assessments

Schedules of Assessments can be found in Clinical Study Protocol APD334-301 Appendix 1.

3.3. Changes to Analysis from Protocol

The primary analysis set for the efficacy endpoints is updated to the Full Analysis Set with MMS 5 to 9. Subgroup definition related to MMS is updated accordingly. The entire FAS will be used in supplementary analyses.

Definition of histologic improvement based on Geboes in Section 16 is updated from Geboes Index score < 3.1 to Geboes Index score ≤ 3.1.

The criterion of “Exclusion of other causes for loss of response unrelated to underlying UC (eg, infection)” is removed from the definition of “loss of response” in Section 16 since a more conservative approach is used.

4. PLANNED ANALYSES

The following analyses will be performed for this study:

- Final Analysis

4.1. Interim Analysis

There is no planned or unplanned interim analysis for this study.

4.2. Final Analysis

All final, planned analyses identified in this SAP will be performed following Sponsor (Arena Pharmaceuticals, Inc., hereafter referred to as Arena) authorization of the SAP, database lock, analysis sets, and unblinding of treatment. After Arena has authorized breaking of the study blind, the final analysis will be performed.

Any posthoc, exploratory analyses completed to support planned study analyses, which were not identified in this SAP, will be documented and reported in the clinical study report (CSR). Any results from these unplanned analyses will also be clearly identified in the text of the CSR ([ICH 1995](#)).

5. ANALYSIS SETS

Agreement and authorization by Arena of subjects included/excluded from each analysis set will be conducted prior to the unblinding of the study.

5.1. Screened Set

The Screened Set will consist of all subjects who sign informed consent to participate in the study.

5.2. Randomized Set

The Randomized Set will consist of all subjects who are randomized to study treatment.

5.3. Full Analysis Set

The FAS will consist of all randomized subjects who receive at least 1 dose of study treatment. Subjects will be summarized by the treatment to which they were randomized, regardless of treatment actually received.

5.4. Per Protocol Sets

The Per Protocol Set will be derived for Week 12 and Week 52, separately, and will consist of all subjects in the FAS who adhere to the protocol and complete Week 12 and Week 52, respectively. These sets will be used in sensitivity analyses of the primary and key secondary endpoints to evaluate the influences of important protocol deviations on the primary results.

Subjects may be excluded from the Week 12 Per Protocol Set if they meet any of the following criteria:

- Overall study treatment noncompliance in the first 12 weeks (< 80 or $> 120\%$)
- Receive incorrect study treatment for > 7 days in total in the first 12 weeks
- Use rescue medication (new or dose increase from the highest dose of background treatment for UC received during Screening) or undergo rescue medical procedure that may affect efficacy endpoint at Week 12

- Miss Week 12 endoscopic score (ES) in prespecified analysis window (per Section 6.4) for reasons other than UC flare, or discontinued due to disease worsening or lack of efficacy

Subjects may be excluded from the Week 52 Per Protocol Set if they meet any of the following criteria:

- Overall study treatment noncompliance (< 80 or > 120%)
- Receive incorrect study treatment for > 7 days in total
- Use rescue medication (new or dose increase from the highest dose received during Screening) or rescue medical procedure that may affect efficacy endpoint at Week 52
- Miss Week 52 ES in prespecified analysis window (per Section 6.4) for reasons other than UC flare, or discontinued due to disease worsening or lack of efficacy

Refer to [Appendix 7](#) for the definition of rescue therapies (medication and medical procedures). All protocol deviations will be reviewed in addition to the programmable deviations (eg, study treatment noncompliance) to determine if any deviation is significant enough to exclude a subject from each of the Per Protocol Sets. All exclusion flags will be finalized before study unblinding. Subjects excluded from each Per Protocol Set will be listed along with the reason category above. Subjects will be summarized by treatment to which they were randomized, regardless of treatment actually received.

5.5. Modified Full Analysis Set

The mFAS will consist of all randomized subjects who receive at least 1 dose of study treatment and have a Baseline and at least 1 postrandomization measurement. Subjects will be summarized by treatment to which they were randomized, regardless of treatment actually received. Note that the mFAS can vary between endpoints since some subjects may have the data needed for inclusion in the mFAS for some endpoints, but not other endpoints. Also note, since the Mayo component scores on the disease worsening CRF are manually entered by sites from another CRF, these scores will not be used in the derivation of mFAS for measurement of either MMS or any of its component scores.

5.6. Safety Set

The Safety Set will include all randomized subjects who receive at least 1 dose of study treatment. Subjects will be analyzed according to treatment received, regardless of randomization. The Safety Set will be used for all safety analyses.

5.7. Pharmacokinetic Set

The Pharmacokinetic Set will include all subjects in the Safety Set with at least 1 quantifiable postdose etrasimod concentration which is not impacted by protocol violations or events with potential to affect the etrasimod concentration.

6. GENERAL CONSIDERATIONS

6.1. Reference Start Date and Study Day

Study Day will be calculated from the reference start date and will be used to show start/stop day of assessments and events. Reference start date is defined as the date of first dose (Day 1) and will appear in every listing where an assessment date or event date appears.

- If the date of the event is on or after the reference start date, then:

$$\text{Study Day} = (\text{date of event} - \text{reference start date}) + 1$$

- If the date of the event is prior to the reference start date, then:

$$\text{Study Day} = (\text{date of event} - \text{reference start date})$$

In the situation where the event date is partial or missing, Study Day and any corresponding durations will appear missing in the listings.

6.2. Baseline

Unless otherwise specified, Baseline is defined as the last nonmissing measurement taken prior to the first dose (including unscheduled assessments). If measurements include time (except for health-related quality of life [HRQoL] instruments), the date/time will be used to define Baseline. Otherwise, only dates will be compared. For HRQoL instruments, only the date will be used to derive Baseline. In the case where the last nonmissing measurement and the date of first dose coincide and time is not collected, if the measurement was planned in the protocol to be done prior to the date of first dose, that measurement will be considered in defining Baseline.

6.3. Retests, Unscheduled Visits, and Early Termination Data

For by-visit analyses and summaries, efficacy, safety, HRQoL, and biomarker data (including scheduled, retests, unscheduled, and early termination) will be assigned to visits after the application of the windowing conventions described in Section 6.4. All measurements will be considered in summaries of abnormalities or worst-case values post-Baseline.

The visit windowing will be applied before missing data are imputed.

Listings will include scheduled, unscheduled, retest, and early termination data.

6.4. Windowing Conventions

All scheduled study visits are defined relative to Study Day 1, the date of first dose. Scheduled visit windows are defined in Clinical Study Protocol APD334-301 Appendix 1. A windowing convention will be used to determine the analysis visit value for a given measurement and will be applicable for all by-visit summaries and analyses for efficacy and safety data. Refer to [Table 2](#) for specific visit windows.

Table 2: Visit Windows for Efficacy and Safety Analyses

For all efficacy (except endoscopy/MMS/histology), efficacy-related biomarkers, safety labs and vital signs	
Scheduled Study Visit (Protocol Scheduled Day)	Analysis Visit Window (Study Day)
Baseline	≤ 1
Week 0 (Day 1)	1
Week 0 (Day 2) ^a	2
Week 2 (Day 15 ± 3)	2 to 22, or 3 to 22 ^a
Week 4 (Day 29 ± 3)	23 to 43
Week 8 (Day 57 ± 3)	44 to 71
Week 12 (Day 85 ± 7)	72 to 99
Week 16 (Day 113 ± 7)	100 to 127
Week 20 (Day 141 ± 7)	128 to 155
Week 24 (Day 169 ± 7)	156 to 197
Week 32 (Day 225 ± 7)	198 to 253
Week 40 (Day 281 ± 7)	254 to 309
Week 48 (Day 337 ± 7)	310 to 351
Week 52 (Day 365 ± 14)	> 351
For endoscopy/MMS/histology/all composite endpoints that include an endoscopy component and/or RB/SF component, Physician’s Global Assessment, health-related quality of life, ECGs, OCT, and PFT	
Scheduled Study Visit (Protocol Scheduled Day)	Analysis Visit Window (Study Day)
Baseline	≤ 1
Week 0 (Day 1)	1
Week 0 (Day 2) ^a	2
Week 12 (Day 85 ± 7) ^{b,c}	66 to 113 (66 to 141 for assessments impacted by COVID-19 pandemic) ^d
Week 32 (Day 225 ± 7) [PFT only]	198 to 253 (198 to 281 for assessments impacted by COVID-19 pandemic) ^d
Week 52 (Day 365 ± 14) ^{b,d}	337 to 393 (337 to 421 for assessments impacted by COVID-19 pandemic) ^e

^a Applicable only for subjects who require extended monitoring on Day 2 for vital signs and ECGs.

^b Applicable to all Week 12 and Week 52 efficacy endpoints based on any component of Mayo clinic score, such as clinical remission and symptomatic remission. Protocol visit window for PFT and OCT at Week 12 is Day 85 ± 7 (FDA 2020).

^c Due to overlapping windows at Week 12 and Week 8/16, a hierarchical algorithm will be used for the mapping. Week 12 will always be mapped first, then assess records for the Week 8/16 mapping. If a scheduled Week 8 and 12, or Week 12 and 16 visits are both present in the same visit window, they will be mapped to separate visits (ie, Week 8 and Week 12 or Week 12 and Week 16).

^d Due to overlapping windows in the RB/SF components at Week 52 and Week 48, a hierarchical algorithm will be used for the mapping. Week 52 will always be mapped first, then assess records for the Week 48 mapping. If a scheduled Week 48 and 52 visits are both present in the same visit window, they will be mapped to separate visits (ie, Week 48 and Week 52).

^e Assessments impacted by the COVID 19 pandemic are reported in the Date of Visit eCRF.

COVID-19, coronavirus disease 2019; ECG, electrocardiogram; eCRF, electronic case report form; MMS, modified Mayo Score; OCT, optical coherence tomography; PFT, pulmonary function test; RB, rectal bleeding; SF, stool frequency.

For cardiac remonitoring upon treatment reinitiation after Day 2, as manifested by more than 1 timed measurement in vital signs or electrocardiogram (ECG) parameters on the same collection date of an unscheduled visit, measurements will be mapped to analysis visit of Cardiac Remonitoring 1, Cardiac Remonitoring 2, etc. in the respective analysis dataset. For analysis, each timed measurement will be programmatically assigned to the nearest hourly timepoint (eg, Predose, 1-hour Postdose) based on their relationship to the dosing time on the same day. If the dosing date/time on the same day is missing, the timepoint will remain missing. Once mapped, these timed measurements from cardiac remonitoring visits will not be considered for any other analysis visit.

Windowing will be applied prior to any missing data calculations. The last nonmissing measurement taken prior to Day 1 (including unscheduled assessments) will be labeled as “Baseline”. Unless stated otherwise, data from all visits including scheduled, unscheduled, and ET visits will be eligible for allocation to an analysis visit. The 2-Week and 4-Week Follow-Up visits will not be included in the visit windows and will be summarized separately without any window applied.

If one or more results for a variable are assigned to the same analysis visit, the result with the date closest to the protocol scheduled day will be used in the analysis, except for component subscores of MMS or TMS, in which case, the composite score and component subscores from the same date will be used in the analysis. If 2 measurements in the same analysis visit window are equidistant from the protocol scheduled study day, the earliest measurement will be used in the analysis. If multiple assessments are available on the same day, then the average of the assessment will be used in the analysis, except for laboratory and ECG data where the assessment at the earliest time of the same day will be used. If both central and local assessments of the same ECG or lab test are available on the same day, the central result will take precedence over the local result.

For the overlapping visit windows at Week 12 and Week 8/16, a hierarchical algorithm will be applied as follows: always map to the Week 12 study visit if the study day falls within the visit window of the Week 12 study visit, then assess the other records for the Week 8 or Week 16 study visit. The same algorithm will be used for Week 48 and Week 52, with Week 52 taking priority.

6.5. Statistical Tests

The default significance level will be 0.05; confidence intervals (CIs) will be 95%, and all tests will be 2-sided, unless otherwise specified in the description of the analyses.

6.6. Common Calculations

For quantitative measurements:

- Change from Baseline = Test Value at Visit X – Baseline Value
- Percent change from Baseline = (Test Value at Visit X – Baseline Value) / Baseline Value × 100

- Proportion at Visit X = Number of subjects satisfying criteria at Visit X / Total number of subjects at Visit X

6.7. General Study Information

A general table with summary of study information will be generated, including the date of first subject signed informed consent form (ICF), the last subject visit date, and the database lock date. All analyses will be conducted using SAS® (v9.4 or later, SAS Institute Inc., Cary, NC).

7. DETERMINATION OF SAMPLE SIZE

Based on a 2-group Fisher’s exact test, a 1-sided significance level of 0.025, and a 2:1 randomization ratio, 420 total subjects (280 etrasimod, 140 placebo) are required to achieve 93.4% power to detect a difference of 13.5% in the primary endpoint of clinical remission at Week 52 between the etrasimod treatment group (23.5%) and the placebo treatment group (10.0%). With this sample size, there will be 96% power to detect a difference of 12.5% in the other primary endpoint of clinical remission at Week 12, assuming a placebo rate at 6.0%. The lower bound of overall power for both primary endpoints (as coprimary endpoints), is at least 90%; since the 2 primary endpoints are expected to be at least moderately, positively correlated, the actual overall power to reject both of their null hypotheses is likely to be greater than 90%. Workbook calculations from EAST® 6 statistical software are included below.

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Fisher's Exact

Test Parameters	
Design ID	Week 52 Clin Rem
Test Type	2-Sided
Specified α	0.05
Power	0.934
Sample Size (n)	420
Model Parameters	
Proportion Treatment (π_t)	0.235
Proportion Control (π_c)	0.1
$\delta = \pi_t - \pi_c$	
Under H0	0
Under H1	0.135
Allocation Ratio (n_t/n_c)	2

Sample Size Information

Sample Size (n)	420
Treatment (n _t)	280
Control (n _c)	140

Design: Discrete Endpoint: Two-Sample Test - Parallel Design - Fisher's Exact

Test Parameters	
Design ID	Week 12 Clin Rem
Test Type	2-Sided
Specified α	0.05
Power	0.96
Sample Size (n)	420
Model Parameters	
Proportion Treatment (π_t)	0.185
Proportion Control (π_c)	0.06
$\delta = \pi_t - \pi_c$	
Under H0	0
Under H1	0.125
Allocation Ratio (n_t/n_c)	2

⊖ **Sample Size Information**

Sample Size (n)	420
Treatment (n _t)	280
Control (n _c)	140

8. STATISTICAL CONSIDERATIONS**8.1. Multicenter Studies**

This study will be conducted by multiple investigators at multiple centers internationally. Randomization to treatment groups is stratified by:

- Naïve to biologic or JAK inhibitor therapy at study entry (Yes or No)
- Baseline corticosteroid use (Yes or No)
- Baseline disease activity (MMS: 4 to 6 or 7 to 9)

Data from all sites will be pooled and statistical analyses will not be adjusted for investigational site, country, or geographic region.

8.2. Adjustments for Covariates and Factors to be Included in Analyses

The following covariates and factors, and/or endpoint Baseline measure are used in the analyses. For details of their inclusion in the models, refer to the specific analysis section.

- Naïve to biologic or JAK inhibitor therapy at study entry (Yes or No)
- Baseline corticosteroid use (Yes or No)
- Baseline disease activity (MMS: 4 to 6 or 7 to 9)

If a subject was assigned the wrong stratum at randomization, their stratum based on the randomization system will be used for the statistical analyses. Analyses may be repeated using actual stratum if mis-randomization occurs in more than 10% of randomized subjects.

8.3. Missing Data

Missing adverse event (AE) relationship to study treatment and AE seriousness will be imputed as described in Section 18.1.1.2 and Section 18.1.3, respectively. Partial or missing AE start dates, concomitant medication (CM) start dates, and hospitalization dates will also be imputed as described in Appendix 2. No other missing safety data will be imputed.

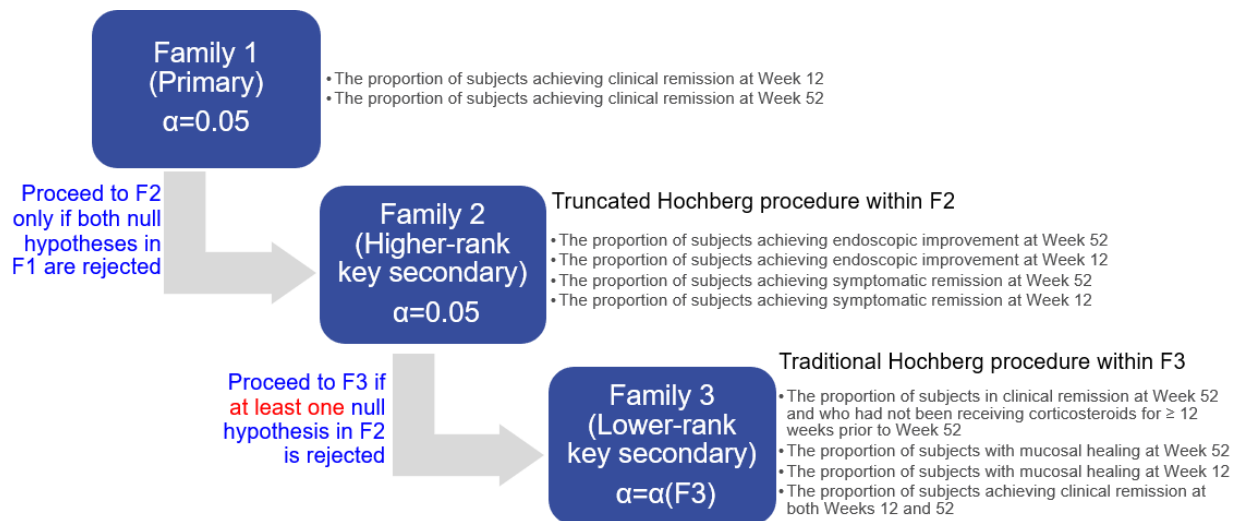
Missing efficacy data will be handled as described in Section 16.1.2, Section 16.2.2, and Section 16.3.2. In the primary analysis of the primary endpoint and main analyses of all binary responder-type endpoints, all subjects with missing data, regardless of reason for missingness, will be considered as nonresponders. In the main analysis of continuous or score endpoints, such as biomarker measures, urgency Numeric Rating Scale (NRS), abdominal pain NRS, and HRQoL measures, subjects with missing data will be handled using observed cases only, multiple imputation, or a mixed-effect model with repeated measures.

8.4. Multiple Comparisons and Multiplicity

There are multiple null hypotheses for the comparison of etrasimod and placebo in the primary and key secondary endpoints. The family-wise type I error rate will be controlled at a fixed α level at 0.05 (2-sided) using the following testing procedure. First, the whole α will be spent on testing family 1 (F1) consisting of the coprimary endpoints. This study will be considered as an overall success only if both of the primary null hypotheses are rejected, each at the α level. This study will be considered as a partial success if only 1 of the 2 primary null hypotheses are rejected at $\alpha/2$ if the other has p-value $> \alpha$ (FDA 2017).

Figure 2 illustrates the parallel gatekeeping procedure, which will be followed to control the family-wise error rate.

Figure 2: Gatekeeping Procedure Summary



F1, family 1; F2, family 2, F3, family 3

Traditional Hochberg procedure:

Consider testing the family of hypotheses $H_{0i}; i = 1 \dots k$. Let p_i , for $i = 1 \dots k$, denote the sample p-values of tests for $H_{0i}; i = 1 \dots k$, computed without multiplicity adjustment. Let $[1] \dots [k]$ denote the random indices such that $p_{[1]} \leq \dots \leq p_{[k]}$ ($p_{[1]}$ is the smallest p-value, $p_{[k]}$ is the largest p-value). Hochberg procedure is implemented as follows:

Step 1: if $p_{[k]} < \alpha$, reject $H_{0[i]}, i = 1, \dots, k$ and stop; otherwise proceed to next step.

Step 2: if $p_{[k-1]} < \alpha/2$, reject $H_{0[i]}, i = 1, \dots, k-1$, and stop; otherwise proceed to next step.

...

Step k : if $p_{[1]} < \alpha/k$, reject $H_{0[1]}$, and stop; otherwise no null hypothesis is rejected.

Truncated Hochberg procedure (with truncation fraction f where $0 < f < 1$):

Consider testing the family of hypotheses $H_{0i}; i = 1 \dots k$. Let p_i , for $i = 1 \dots k$, denote the sample p-values of tests for $H_{0i}; i = 1 \dots k$, computed without multiplicity adjustment. Let $[1] \dots [k]$ denote the random indices such that $p_{[1]} \leq \dots \leq p_{[k]}$ ($p_{[1]}$ is the smallest p-value, $p_{[k]}$ is the largest p-value). Truncated Hochberg procedure is implemented as follows:

Step 1: if $p_{[k]} < f\alpha + (1-f)\alpha/k$, reject $H_{0[i]}, i = 1, \dots, k$ and stop; otherwise proceed to next step.

Step 2: if $p_{[k-1]} < f\alpha/2 + (1-f)\alpha/k$, reject $H_{0[i]}, i = 1, \dots, k-1$, and stop; otherwise proceed to next step.

...

Step k : if $p_{[1]} < f\alpha/k + (1-f)\alpha/k = \alpha/k$, reject $H_{0[1]}$, and stop; otherwise no null hypothesis is rejected.

A demonstration of alpha level for F3:

A truncation fraction $f = 0.8$ will be used in the truncated Hochberg procedure in testing F2, which results in the following alpha level to be passed to F3 for testing.

- All 4 null hypotheses in F2 are rejected:
 $\alpha(F3) = \alpha(F2) = 0.05$
- 3 of 4 null hypotheses in F2 are rejected:
 $\alpha(F3) = \alpha(F2) - [f\alpha(F2) + (1-f)\alpha(F2)/4] = 0.0075$
- 2 of 4 null hypotheses in F2 are rejected:
 $\alpha(F3) = \alpha(F2) - 2*[f\alpha(F2)/2 + (1-f)\alpha(F2)/4] = 0.005$
- 1 of 4 null hypotheses in F2 is rejected:
 $\alpha(F3) = \alpha(F2) - 3*[f\alpha(F2)/3 + (1-f)\alpha(F2)/4] = 0.0025$

8.5. Subgroup Analyses

The following subgroups will be assessed for the primary and key secondary endpoints:

- Sex (Female or Male)

- Age (\leq or $>$ Median, $<$ or ≥ 65)
- Race (White or Non-White)
- Region (North America, Western Europe, Eastern Europe, Other)
- Baseline oral corticosteroid usage (Yes or No)
- Naïve to biologic or JAK inhibitor therapy at study entry (Yes or No)
- Baseline disease activity (MMS: 4 to 6, 7 to 9)
- Actual Baseline MMS (5 to 7, 8 to 9)
- Baseline fecal calprotectin (\leq or $>$ Median)
- Baseline high-sensitivity C-reactive protein (\leq or $>$ Median)
- Baseline total Mayo score (TMS) (\leq or $>$ 8)
- Duration of UC (\leq or $>$ Median)
- Extent of disease (Proctosigmoiditis/Left-sided colitis, Pancolitis, Proctitis as reported on the eCRF)
- Proctitis (Yes or No, based on central read)
- Prior UC treatment of oral 5-aminosalicylic acid (5-ASA) only (Yes or No)
- Prior UC treatment failure of oral 5-ASA only (Yes or No)
- Number of prior biologic or JAK inhibitor therapies (1 or $>$ 1)
- Prior UC treatment failure of anti-tumor necrosis factor alpha (anti-TNF α) (Yes or No)

Additional subgroups may be assessed, if deemed necessary. The medians will be derived based on the FAS for all subgroups cut at the median. It should be noted that the study was not designed to detect treatment differences with high statistical power within subgroups. If any subgroup includes $<$ 5% of all subjects, no inferential statistics will be generated.

The actual stratum at Baseline will be used for all subgroup analyses.

Forest plots of the overall analysis set and all subgroups will be presented for each of the primary and key secondary endpoints. Inference from the model described in Section 16.1.3 will be used and differences in remission percentages along with 95% CIs will be displayed.

9. OUTPUT PRESENTATIONS

[Appendix 1](#) contains the conventions for presentation of data in outputs.

10. DISPOSITION AND PROTOCOL DEVIATIONS

All subjects who provide informed consent will be accounted for in this study. Inclusion criteria not met and exclusion criteria met will be listed.

Among the randomized subjects, the number and percent of subjects who completed/discontinued treatment, reasons off treatment, the number and percent of subjects who completed/discontinued the study, and reasons off study will be summarized. This summary will also be provided by region and country. The number and percent of subjects in each analysis set will be summarized for all randomized subjects. An additional summary of the number of subjects screened, the number of screen-failed subjects, and reason for screen failure will be presented for all screened subjects. A listing of subjects whose blind was broken will be provided. The number of subjects whose visit was impacted by the COVID-19 pandemic per CRF will also be summarized by nominal visit.

During site monitoring, protocol deviations will be graded as Critical, Major, or Minor.

According to ICH E3 and ICH E3(R1), important protocol deviations are a subset of protocol deviations that might significantly affect the completeness, accuracy, and/or reliability of the study data or that might significantly affect a subject's rights, safety, or well-being (ICH 1995, ICH 2012). For example, important protocol deviations might include enrolling subjects in violation of key eligibility criteria designed to ensure a specific subject population or failing to collect data necessary to interpret primary endpoints, as this may compromise the scientific value of the study.

During the review of all reported deviations, important deviations related to study inclusion or exclusion criteria, conduct of the study, patient management, or patient assessment will be identified. Where relevant the importance of a potentially important protocol deviations will be assessed in the context of the study's estimands to evaluate potential impact.

All important protocol deviations will be summarized for the FAS in the following categories in descending frequency in the etrasimod group.

All protocol deviations will be listed, including whether a deviation was impacted by the COVID-19 pandemic (Yes or No). Protocol deviation categories include but are not limited to the following:

- Informed Consent
- Eligibility and Entry Criteria
- Concomitant Medication
- Laboratory Assessment
- Study Procedures
- Serious Adverse Event
- Randomization
- Visit Schedule

- Investigational Product Compliance
- Efficacy
- Administrative
- Source Document
- Regulatory or Ethics Approvals
- Other

Important protocol deviations will also be summarized by whether they were impacted by COVID-19 pandemic (Yes or No). Additionally, a summary of missing endoscopy regardless of reason and missing endoscopy due to visit impacted by the COVID-19 pandemic will be presented by visit.

11. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

The following demographic data and Baseline characteristics will be summarized for the FAS.

- Age on consent (years)
- Sex
- Race
- Ethnicity
- Woman of childbearing potential (Yes or No)
- Height (cm)
- Weight (kg)
- Body mass index (BMI) (kg/m²)
- Alcohol consumption (Yes or No)
- Caffeine consumption (Yes or No)
- Tobacco use (Yes or No)

This summary will be repeated for age, sex, race, ethnicity, or woman of childbearing potential for subjects who fail screening.

The following Baseline characteristics related to UC will be summarized:

- Extent of disease (Proctosigmoiditis/Left-sided colitis, Pancolitis, Proctitis reported on the electronic case report form [eCRF], and Proctitis reported from central reader review)
- Baseline MMS
- Baseline rectal bleeding (RB) subscore

-
- Baseline stool frequency (SF) subscore
 - Baseline endoscopic subscore
 - Baseline Physician's Global Assessment (PGA)
 - Baseline TMS
 - Duration of UC (years)
 - Any acute exacerbations within past 12 months (Yes or No), including the number of acute exacerbations among those with any acute exacerbation = Yes
 - Colonoscopy within past 12 months (Yes or No)
 - Surgery for UC (Yes or No), including the number of surgeries among those with surgery for UC = Yes
 - Hospitalizations for UC (Yes or No), including the number of hospitalizations
 - Naïve to biologic or JAK inhibitor therapy (Yes or No) – Reported (used for stratification at randomization)
 - Naïve to biologic or JAK inhibitor therapy (Yes or No) – Actual (medications reported on the eCRF)
 - Baseline corticosteroid use (Yes or No) – Reported (used for stratification at randomization)
 - Baseline corticosteroid use (Yes or No) – Actual (medications reported on the eCRF)
 - Baseline MMS group (4 to 6, 7 to 9) – Reported (used for stratification at randomization)
 - Baseline MMS group (4 to 6, 7 to 9, and 5 to 9) – Actual (scores reported on the eCRF)
 - Naïve to biologic or JAK inhibitor therapy – Difference between the Reported and the Actual
 - Baseline corticosteroid use (Yes or No) – Difference between the Reported and the Actual
 - Baseline MMS group (4 to 6, 7 to 9) – Difference between the Reported and the Actual
 - Prior failure of oral 5-ASA only (Yes or No)
 - Prior failure of anti-TNF α (Yes or No)
 - Prior failure of anti-TNF α or vedolizumab (Yes or No)

Prior treatment for UC will be summarized, including category of treatment, reason for discontinuation, and estimated duration (weeks) of corticosteroid use over the last 12 months.

11.1. Derivations

- Duration of UC (year) = (Informed consent date – Date of diagnosis + 1) / 365.25
- Weight (kg) = Weight (lb) × 0.4536
- Height (cm) = Height (in) × 2.54
- Height (m) = Height (in) × 0.0254 = Height (cm) × 0.01
- BMI (kg/m²) = Weight (kg) / Height (m)²

12. MEDICAL HISTORY

Medical history will be collected on the medical history eCRF and coded using Medical Dictionary for Regulatory Activities (MedDRA, v24.1). The version used to code medical history will be displayed in the outputs. All medical history will be summarized for the Safety Set by system organ class (SOC) and preferred term (PT).

13. MEDICATIONS

Medications will be captured on the Concomitant Medications eCRF and coded using the WHO Drug dictionary (WHODDE01SEP2021). Refer to [Appendix 2](#) for handling of partial dates for medications. In the case where it is not possible to define a medication as prior or concomitant, the medication will be classified by the worst case (ie, concomitant).

- ‘Prior’ medications are medications which started and stopped prior to the first dose of study treatment
- ‘Concomitant’ medications are medications which started prior to, on or after the first dose of study treatment AND ended on or after the date of first dose of study treatment or were ongoing at the end of the study

Concomitant medications will be summarized by Anatomical Therapeutic Chemical (ATC) Level 2 and Preferred Drug name for the Safety Set. Concomitant medications used for UC will be flagged as such on the Concomitant Medications eCRF and summarized by ATC Level 2 and Preferred Drug name for the Safety Set.

14. STUDY TREATMENT EXPOSURE

The date of first and last study treatment administration will be taken from the Day 1 Onsite Dosing Administration eCRF and the End of Study eCRF, respectively. Interruptions, compliance, and dose changes are not taken into account for duration of exposure. Exposure to study treatment in weeks will be summarized for the Safety Set.

Dose interruptions will be recorded on the Dosing Administration eCRF, and overdose will be recorded on the Overdose eCRF. The frequency and percentage of subjects who had at least

1 dose interruption, who had at least 1 dose interruption of > 7 days, who had at least 1 dose interruption of > 14 days, who had 1 overdose, and who had > 1 overdose will be summarized.

14.1. Derivations

Duration of exposure (weeks) = (date of last study treatment administration – date of first study treatment administration + 1) / 7. For subjects with missing Date of Last Dose on the End of Study eCRF, their date of last study treatment administration will be imputed by the last date of all dosing administration start/stop dates recorded.

15. STUDY TREATMENT COMPLIANCE

The total number of tablets expected, total number of tablets taken, total number of tablets missed, overall compliance with study treatment, and frequency and percentage of subjects with overall compliance of < 80% or > 120% will be summarized for the Safety Set for both the first 12 weeks and the entire study.

15.1. Derivations

Compliance to study treatment is based on the Drug Accountability eCRF and will be calculated as the total number of tablets taken (total dispensed – total returned) divided by the number of tablets expected during the Treatment Period, expressed as a percentage, refer to calculations below.

The total number of tablets expected is defined as the number of tablets that a subject is expected to have taken between their first and last study treatment administration and is numerically identical to the subject's overall study treatment exposure, since the medication is to be taken once daily. On any site visit day, the medication is to be held and taken at the site, after all predose assessments have been completed. For example, if a subject took their last dose of study treatment on Day 200 and returned to the site on Day 207 to return the study treatment bottle, then the total number of tablets expected would be 200, not 206.

- Overall compliance to study treatment will be calculated as follows; total number of tablets returned subtracted from total number of tablets dispensed, divided by the duration of treatment, multiplied by 100

For all bottles not returned, it will be assumed that all dispensed tablets were taken. For each subject, if a high percentage (> 25%) of bottles were not returned by a subject, additional analyses may be done where bottles not returned are excluded from the overall compliance calculation for the subject. In such analysis, the date of last dose or the date of last bottle return, whichever is earlier, will be used as the “date of last dose” in the calculation above.

Both scheduled and unscheduled study treatment dispensations will be used in the compliance calculation. Overall compliance calculations will be performed for both the 12-Week Treatment Period and the entire study and will be used in determining inclusion/exclusion of subjects in the respective Per Protocol Set. For subjects discontinued prior to Week 12, their overall compliance for the 12-Week Treatment Period and the entire study will be identical. For subjects who

remained in the study beyond Week 12, the 12-Week total tablets expected will be defined as Date of Week 12 visit – Date of first dose, since their last dose in the 12-Week Treatment Period should have been taken on the day before Week 12 visit.

16. EFFICACY OUTCOMES

Unless otherwise stated, the primary efficacy analyses will be based on subject stratum used at the time of randomization for the FAS with Baseline MMS 5 to 9. Primary and key secondary efficacy analyses will be repeated for all subjects in FAS with Baseline MMS 4 to 9 as a supplementary analysis. The following definitions will be used to assess efficacy outcomes:

- Clinical response: A ≥ 2 -point and $\geq 30\%$ decrease from Baseline in MMS, and a ≥ 1 -point decrease from Baseline in RB subscore or an absolute RB subscore ≤ 1
- Clinical remission: SF subscore = 0 (or = 1 with a ≥ 1 -point decrease from Baseline), RB subscore = 0, and ES ≤ 1 (excluding friability)
- Endoscopic improvement: ES of ≤ 1 (excluding friability)
- Endoscopic normalization: ES = 0
- Mucosal healing: ES of ≤ 1 (excluding friability) with histologic remission measured by a Geboes Index score < 2.0
- Symptomatic remission: SF subscore = 0 (or = 1 with a ≥ 1 -point decrease from Baseline) and RB subscore = 0
- Complete symptomatic remission: SF subscore = 0 and RB subscore of 0
- Symptomatic response: Decrease from Baseline $\geq 30\%$ in composite RB and SF subscores
- Noninvasive clinical response: A $\geq 30\%$ decrease from Baseline in composite RB and SF, and a ≥ 1 -point decrease from Baseline in RB subscore or an absolute RB subscore ≤ 1
- Histologic improvement: Geboes Index score ≤ 3.1
- Histologic remission: Geboes Index score < 2.0
- Histologic improvement using Roberts Histopathology Index (RHI): $\geq 50\%$ reduction from Baseline in RHI or RHI ≤ 3
- Histologic remission using RHI: RHI ≤ 3 , with scores of 0 (zero) for both Geboes Grade 2B (lamina propria neutrophils) and Grade 3 (neutrophils in epithelium)
- Histologic improvement using Nancy Histologic Index (NHI): ≥ 1 point reduction from Baseline NHI
- Histologic remission using NHI: ≤ 1

- Clinical remission using total Mayo Clinic score: total Mayo Clinic score of ≤ 2 points with no individual subscore of > 1 point
- Clinical response using total Mayo Clinic score: $A \geq 3$ -point and $\geq 30\%$ decrease from Baseline in total Mayo Clinic score, and a ≥ 1 -point decrease from Baseline in RB subscore or an absolute RB subscore ≤ 1
- Corticosteroid-free clinical remission: Clinical remission at Week 52 and who had not been receiving corticosteroids for ≥ 12 weeks prior to Week 52
- Sustained clinical remission: Clinical remission at both Week 12 and Week 52
- Loss of response:
 - Achieved clinical response at Week 12
 - $A \geq 2$ -point increase from Week 12 in the combined SF + RB scores and combined SF + RB score of ≥ 4 , on 2 consecutive visits (≥ 7 days apart), and
 - Confirmed by centrally read ES ≥ 2

16.1. Primary Efficacy

16.1.1. Primary Efficacy Variables and Derivations

The primary efficacy endpoints will compare etrasimod to placebo for:

- The proportion of subjects achieving clinical remission at Week 12
- The proportion of subjects achieving clinical remission at Week 52

Clinical remission is defined in Section 16. The SF, RB, and ES subscores will be based on the Mayo Clinic score eCRF. Subjects who achieve clinical remission will be referred to as responders. Subjects who do not achieve clinical remission will be referred to as nonresponders.

Multiplicity adjustments for the coprimary endpoints are discussed in Section 8.4.

16.1.2. Missing Data Methods for Primary Efficacy Variables

Subjects who 1) discontinue the study for worsening of disease, lack of efficacy, or adverse event related to UC, 2) initiate a rescue medication for UC, 3) have an increase in dose over Baseline levels in their existing UC medication, or 4) have rescue medical procedure (eg, colectomy, ileostomy, or sigmoidectomy) during the study as confirmed after blinded review by clinical and medical team members before study unblinding will be considered to have a known (ie, nonmissing) outcome of nonresponse in the analysis of all efficacy endpoints at any subsequent timepoints, including the primary endpoint. The rescue medications will be identified by the Arena clinical team during blinded data review. Refer to [Appendix 7](#) for the definition of rescue therapies (medication and medical procedures). In scenarios 2 and 3 above, if a subject has an efficacy measurement collected after the initiation or dose increase from Baseline of UC medication, the observed data will be censored at the time of initiation or dose increase and they will be considered as having a nonresponse. Subjects in all 4 scenarios above may still be included in the respective Per Protocol Set, provided they do not violate other criteria for Per

Protocol Set. For example, they will be excluded from the respective Per Protocol Set if they initiate a prohibited medication before the Week 12 or Week 52 efficacy assessment that can affect efficacy of the study treatment and the indication is unrelated to UC.

Analysis visits will be mapped as per Section 6.4 before any missing data imputation method is applied.

Subjects with a missing efficacy outcome will be included in the primary analyses using the following missing data methods:

- Primary method: Single imputation as nonresponder
- Sensitivity analyses: Multiple imputation under missing at random (MAR), tipping point analysis, multiple imputation with Copy Reference (CR) under missing not at random (MNAR), and a hybrid imputation with multiple imputation to handle missing endoscopy data due to the COVID-19 pandemic impact and nonresponder imputation (NRI) for missing data not due to the COVID-19 pandemic impact. More details are provided in Section 16.1.4
- Supplementary analyses: mFAS analysis with data as observed, Per Protocol Set analysis (Week 12 Per Protocol Set for Week 12 endpoints and Week 52 Per Protocol Set for Week 52 endpoints)

16.1.3. Primary Analysis of Primary Efficacy Variables

The following hypotheses (H_{01} and H_{02}) for the coprimary endpoints will be tested:

H_{01} : The proportion of subjects achieving clinical remission at Week 12 is the same between etrasimod and placebo.

H_{11} : The proportion of subjects achieving clinical remission at Week 12 is different between etrasimod and placebo.

H_{02} : The proportion of subjects achieving clinical remission at Week 52 is the same between etrasimod and placebo.

H_{12} : The proportion of subjects achieving clinical remission at Week 52 is different between etrasimod and placebo.

Both hypotheses will be tested at the 2-sided α level, 0.05.

The primary efficacy analysis will be performed for the FAS with Baseline MMS 5 to 9. The Cochran-Mantel-Haenszel (CMH) test will be used, stratified by naïve to biologic or JAK inhibitor therapy at study entry (Yes or No), Baseline corticosteroid use (Yes or No), and Baseline disease activity (MMS: 4 to 6 or 7 to 9). Reported randomization stratum will be used in the model. Results will be expressed as the number and percentage of subjects in remission, difference in remission percentages and 95% CI and p-value, odds ratio and 95% CI. If there are no responders in both treatment groups in any stratum, the difference in remission percentages and associated inferences can still be made. However, the odds ratio and its CI may be calculated from the unstratified CMH.

16.1.4. Sensitivity Analysis of Primary Efficacy Variables

Four sensitivity analyses will be implemented to explore different types of missing data approaches: Multiple imputation under MAR, tipping point analysis, multiple imputation with CR under MNAR, and multiple imputation under MAR/NRI imputation hybrid. These sensitivity analyses will all use the FAS with Baseline MMS 5 to 9. There are 2 different types of missing data patterns: non-monotone and monotone. A subject with missing data at any post-Baseline visit, and a non-missing data point after that visit is said to have intermittent missing data (non-monotone). A subject with missing data at any post-Baseline visit and at all subsequent visits is said to have monotone missing data.

16.1.4.1. Multiple Imputation Under MAR

Any missing component scores of MMS at the planned assessments will be imputed using multiple imputation under MAR. MAR assumes the missing value is independent of unobserved outcomes given observed data (eg, subjects with missing RB subscores can be modeled based on subjects with observed RB subscores) (Rubin 1987).

The following steps will be implemented:

Step 1

Regardless of the arbitrary missing data pattern (ie, non-monotone or monotone), a fully conditional specification (FCS) method with predictive mean matching for continuous variables will be used to impute the missing data at all timepoints. The FCS method allows for separate conditional distributions for each imputed variable. The predictive mean matching approach creates a regression model using parameters sampled from the posterior distribution and then a predicted value for each missing value is computed. The missing value is replaced by randomly selecting an observation from a set of 'k' values that are the closest predicted values to the missing predicted value. Missing data imputation will be performed using the SAS PROC MI procedure. The number of imputations will be 40 and is based on a projected overall dropout rate of 50% and relative efficiency of ~ 99%. A separate imputation model will be used for each treatment group. The RB subscore, SF subscore, ES, PGA, stratification variables, a binary variable indicating if a subject is a prior UC treatment failure of oral 5-ASA only (ie, subject had an inadequate response, loss of response, or intolerance to previous treatment with oral 5-ASA and did not have an inadequate response, loss of response, or intolerance to any other previous UC medication), Geboes index score, and biomarker measurements (absolute lymphocyte count [ALC], fecal calprotectin [FCP], hs-CRP, log-transformed) at each planned visit will be included in the imputation models. Refer to [Appendix 3](#) for the order of variables and timepoints in the model.

Step 2

For each of the imputed datasets at Week 12 and Week 52, subjects will be classified as responder or nonresponder based on the criteria specified in Section 16. For each dataset, the proportion of responders will be calculated. For endpoints requiring MMS, MMS will be recomputed by summing the scores of ES, SF, and RB.

Step 3

The CMH test, stratified by naïve to biologic or JAK inhibitor therapy at study entry (Yes or No), Baseline corticosteroid use (Yes or No), and Baseline disease activity (MMS: 4 to 6 or 7 to 9) will be run for each of the 40 datasets to obtain 40 estimators of interest at Week 12 and Week 52.

Step 4

Use SAS PROC MIANALYZE to produce an overall pooled estimate (mean of 40 estimates) with its associated standard error (SE), CI and pooled p-value.

However, there could be certain adjustments due to unexpected data issues after unblinding treatment. All post-unblinding modifications to the multiple imputation model or approaches to address missing data will be described in the CSR.

16.1.4.2. Tipping Point Analysis

A tipping point approach based on multiple imputation of values through Week 12 and Week 52 will be used with a specified adjustment (referred to as delta adjustment or shift) applied to values imputed under a MNAR-based imputation model. To find a tipping point, a series of imputations will be performed with increasing values of delta.

The goal is to evaluate the plausibility of the assumed expected values for missing outcomes on each treatment group under which the conclusions change, ie, under which values of delta there are no longer evidence of significant treatment difference.

The following steps will be implemented:

Step 1

Same as Step 1 described in Section [16.1.4.1](#).

Step 2

Impute missing RB, SF, and ES subscores, then apply a random shift under the MNAR assumption to the imputed subscores and rounded to the nearest integer at Week 12 and Week 52 in both treatment groups, so that the expected value of the additive shift in RB, SF, and ES subscores combined is delta. The same variables used in Step 1 will be used in the MNAR model. The tipping point analysis sample code is specified in [Appendix 3](#). After the shift, if any assessments have values less than 0, the value will be set to 0. Similarly, if any assessments have values greater than 3, the value will be set to 3. MMS will then be rederived for each of the 40 imputed datasets, using the shifted values. Next, classify subjects as a responder or nonresponder at Week 12 and Week 52 based on the criteria specified in Clinical Study Protocol APD334-301 Section 10.5. For each dataset, the proportion of responders and nonresponders will be calculated.

Step 3

Same as Step 3 described in Section 16.1.4.1.

Step 4

Same as Step 4 described in Section 16.1.4.1.

Note: Imputations with delta adjustment will be performed with progressively increasing delta values until a tipping point is reached based on the significance of the CMH test. A tipping point will correspond to the smallest value of delta for which the primary hypotheses are no longer rejected. The shift parameter (delta) for the etrasimod treatment group will take increasing values from 0 with increments of 1 (maximum MMS of 9) at the MMS level, representing adjustments towards worse outcomes (ie, after dropout, on average, subjects on treatment will have MMS worsen). For the placebo arm, delta adjustments will decrease with increments of 1, representing adjustments in the direction of better outcomes. The adjustments in each treatment group will be continued until the CMH test for the primary hypotheses are no longer significant at the 2-sided α level, 0.05. The corresponding shift parameters in both treatment groups at the tipping point will be reported along with the proportions and p-values from each CMH test.

Results will be plotted as a heat map on a two-dimensional plot with axes corresponding to the delta values used in each treatment group respectively, and different colors will be used to represent the magnitude of the 2-sided p-values corresponding to analysis with each combination of delta values. The clinical interpretation about the plausibility of the assumptions underlying the tipping point will be provided in the CSR.

16.1.4.3. Multiple Imputation with Copy Reference

Missing data will also be explored assuming an MNAR approach. MNAR assumes the missing values depend on unobserved outcomes even after accounting for the observed data. Therefore, subjects cannot be modelled based on subjects with observed data, and more assumptions are needed. Copy reference is a type of MNAR approach where each missing value for subjects in the treatment group will be imputed using observed data in the placebo group, and missing values for subjects in the placebo group will be imputed under the MAR approach.

The following steps will be implemented:

Step 1**Step 1 (a)**

Intermittent missing data (before discontinuation from study) at each scheduled visit will be imputed, separately for each treatment group, under the MAR assumption using a multivariate normal imputation model and the Markov Chain Monte Carlo (MCMC) method with multiple chains. The number of imputations will be 40. The resulting data will have a monotone missing data pattern. The imputed data will be used together with the observed data to impute postdiscontinuation missing values.

Step 1 (b)

Only after Step 1a is completed, the remaining monotone missing data after discontinuation will be imputed using the FCS predictive mean matching method based on data from the placebo group only. The RB, SF, ES, and PGA subscores, stratification variables, a binary variable indicating if a subject is a treatment failure, and biomarker measurements (ALC, FCP, hs-CRP, log-transformed) at each planned visit will be included in the imputation model.

Step 2

Same as Step 2 described in Section 16.1.4.1.

Step 3

Same as Step 3 described in Section 16.1.4.1.

Step 4

Same as Step 4 described in Section 16.1.4.1.

16.1.4.4. Multiple Imputation with NRI to Handle Missing Data Due to COVID-19 Pandemic

The COVID-19 pandemic introduces unexpected and unknown impact on the clinical study due to unforeseen intercurrent events. As a sensitivity analysis, missing endoscopy data will be handled using a hybrid approach combining NRI and multiple imputation, where subjects with missing endoscopy data due to COVID-19 pandemic will be imputed using multiple imputation method under a MAR assumption, as described in Section 16.1.4.1. Missing data for any other reason will be considered a nonresponse.

16.1.5. Supplementary Analysis of Primary Efficacy Variables

The primary model as described in Section 16.1.3 will be repeated using the mFAS (with data as observed) and Per Protocol Set as supplementary analyses in subjects with Baseline MMS 5 to 9.

Analysis with the primary model as described in Section 16.1.3 will be repeated for both primary endpoints using the FAS with Baseline MMS 4 to 9. The multiple imputations, tipping point analysis, mFAS and Per Protocol Set analyses will also be repeated with Baseline MMS 4 to 9.

16.2. Key Secondary Efficacy

Key secondary efficacy endpoints are defined in Section 16. In general, all the primary and supplementary analyses planned for the primary endpoints in Section 16.1.3 and Section 16.1.4.4 will be repeated for each key secondary efficacy endpoint (Table 3). Additionally, the sensitivity analyses planned for the primary endpoints in Section 16.1.4 will be repeated for key secondary endpoints of endoscopic improvements and symptomatic remission (Table 3). Multiplicity adjustments for the family of the key secondary endpoints are discussed in Section 8.4.

Table 3: Planned Analyses by Endpoint

Endpoint	Primary Analysis of Stratified CMH	Sensitivity Analyses of MI, Tipping Point Analysis, MI with CR, and NRI + MI	Supplementary Analyses in mFAS and Per Protocol Set
Primary endpoints	X ^a	X	X
Key secondary endpoints:			
Endoscopic improvement	X	X	X
Symptomatic remission	X	X (MI, Tipping Point Analysis, and MI with CR only) ^b	X
Corticosteroid-free clinical remission	X		X
Mucosal healing	X	X (MI, MI-CR, and NRI + MI only)	X
Sustained clinical remission	X		X

^a Inference from the primary analysis method on the primary endpoints will be the gatekeeper for testing in the families of key secondary endpoints as per Section 8.4.

^b The hybrid missing data handling approach of NRI + MI will not be applied to the symptomatic remission endpoint since COVID 19 pandemic is not expected to impact subject eDiary entries.

Analyses in the table will be repeated for subjects in the specified analysis population with Baseline MMS 4 to 9. CMH, Cochran-Mantel-Haenszel; CR, copy reference; mFAS, modified full analysis set; MI, multiple imputation; NRI, nonresponder imputation.

16.2.1. Key Secondary Efficacy Variables and Derivations

16.2.1.1. Endoscopic Improvement

For each of the 2 timepoints (Week 12 and Week 52):

The null hypothesis is that the proportion of subjects achieving endoscopic improvement is the same between etrasimod and placebo.

The alternative hypothesis is that the proportion of subjects achieving endoscopic improvement is different between etrasimod and placebo.

16.2.1.2. Symptomatic Remission

For each of the 2 timepoints (Week 12 and Week 52):

The null hypothesis is that the proportion of subjects achieving symptomatic remission is the same between etrasimod and placebo.

The alternative hypothesis is that the proportion of subjects achieving symptomatic remission is different between etrasimod and placebo.

16.2.1.3. Corticosteroid-Free Clinical Remission and Sustained Clinical Remission

For corticosteroid-free clinical remission:

The null hypothesis is that the proportion of subjects achieving corticosteroid-free clinical remission at Week 52 is the same between etrasimod and placebo.

The alternative hypothesis is that the proportion of subjects achieving corticosteroid-free clinical remission at Week 52 is different between etrasimod and placebo.

For sustained clinical remission:

The null hypothesis is that the proportion of subjects achieving sustained clinical remission (ie, at both Week 12 and Week 52) is the same between etrasimod and placebo.

The alternative hypothesis is that the proportion of subjects achieving sustained clinical remission (ie, at both Week 12 and Week 52) is different between etrasimod and placebo.

16.2.1.4. Mucosal Healing

For each of the 2 timepoints (Week 12 and Week 52):

The null hypothesis is that the proportion of subjects with mucosal healing is the same between etrasimod and placebo.

The alternative hypothesis is that the proportion of subjects with mucosal healing is different between etrasimod and placebo.

16.2.2. Missing Data Methods for Key Secondary Efficacy Variables

The same missing data methods used for the primary endpoints described in Section 16.1.2 will be used for the key secondary endpoints.

16.2.3. Primary Analysis of Key Secondary Efficacy Variables

The primary analysis of these key secondary efficacy endpoints will be performed for the FAS with Baseline MMS 5 to 9. For all key secondary endpoints, the model described in Section 16.1.3 will be used.

16.2.4. Sensitivity Analysis of Key Secondary Efficacy Variables

For the key secondary endpoints of endoscopic improvement, symptomatic remission, and mucosal healing, the sensitivity analyses described in Section 16.1.4 will be used.

16.2.5. Supplementary Analysis of Key Secondary Efficacy Variables

For all key secondary endpoints, the primary model as described in Section 16.1.3 will be repeated using the mFAS (with data as observed) and Per Protocol Set as supplementary analyses with Baseline MMS 5 to 9.

For the key secondary endpoints, analyses listed in Table 3 will be repeated in subjects with Baseline MMS 4 to 9.

16.3. Other Secondary Efficacy

Other secondary efficacy endpoints are defined in Section 16.

16.3.1. Other Secondary Efficacy Variables

The other secondary endpoints are:

- The proportion of subjects who had not received corticosteroids for ≥ 4 weeks and achieved clinical remission at Week 52 among subjects receiving corticosteroids at baseline
- The proportion of subjects achieving clinical response at Week 52
- The proportion of subjects achieving clinical response at Week 12
- The proportion of subjects achieving clinical response at both Weeks 12 and 52
- The proportion of subjects with mucosal healing at both Weeks 12 and 52
- The proportion of subjects achieving endoscopic normalization at Week 52
- The proportion of subjects achieving endoscopic normalization at Week 12
- The proportion of subjects achieving endoscopic normalization at both Weeks 12 and 52
- The proportion of subjects achieving symptomatic remission at Weeks 2, 4, 8, 16, 20, 24, 32, 40, and 48
- The proportion of subjects achieving complete symptomatic remission at each study visit (Weeks 2, 4, 8, 12, 16, 20, 24, 32, 40, 48, and 52)
- The proportion of subjects achieving noninvasive clinical response at each study visit (Weeks 2, 4, 8, 12, 16, 20, 24, 32, 40, 48, and 52)
- The proportion of subjects achieving symptomatic response at each study visit (Weeks 2, 4, 8, 12, 16, 20, 24, 32, 40, 48, and 52)
- The proportion of subjects achieving clinical remission at Week 52 among subjects in clinical response at Week 12

16.3.2. Missing Data Methods for Other Secondary Efficacy Variables

All subjects with missing data, regardless of reason for missingness, will be considered as nonresponders. The intercurrent events that are considered in the primary efficacy analysis will be handled similarly for the other secondary efficacy endpoints, as specified in Section 16.1.2.

16.3.3. Analysis of Other Secondary Efficacy Variables

The analysis of other secondary efficacy endpoints will be performed for the FAS with Baseline MMS 5 to 9. The primary analysis method planned for the primary endpoints in Section 16.1.3 will be repeated for each other secondary efficacy endpoint. No multiplicity adjustment will be made.



will be analyzed using the stratified CMH model described in Section 16.1.3 for the FAS with Baseline MMS 5 to 9.

17.1. Variables and Derivations

17.1.1. Inflammatory Bowel Disease Questionnaire

The Inflammatory Bowel Disease Questionnaire (IBDQ) is a 32-item self-administered questionnaire which has 4 dimensions: bowel symptoms (10 items), systemic symptoms (5 items), emotional health (12 items), and social function (5 items). Responses are graded on a 7-point Likert scale where 7 denotes “not a problem at all” and 1 denotes “a very severe problem”. Scores range from 32 to 224, a higher score indicates better quality of life.

The IBDQ total score and 4 domain subscores are derived on the eCRF directly. Details about scoring rules can be found in [Appendix 4](#).

17.1.2. 36-Item Short Form Health Survey, Version 2

The 36-Item Short Form Health Survey (SF-36) is a 36-item, subject-reported survey of subject health. The SF-36 consists of 36 questions measuring 8 health domains: Physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, general health perceptions, mental health, social function, and vitality. The subject’s responses are solicited using Likert scales that vary in length, with 3 to 6 response options per item. The SF-36 will be scored using 2 overall summary scores: Physical component summary and mental component summary scores. A higher score indicates better health status.

The Physical Component Summary Score (norm-based), the Mental Component Summary Score (norm-based), 8 domain subscores (0 to 100 based), and the SF-6D health utility index score are derived by Optum and integrated into the SF-36 eCRF. Details about scoring rules and derivations of domain scores can be found in [Appendix 5](#).

17.1.3. Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis

The Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis (WPAI-UC) consists of 6 questions asking about the effect of UC on the subject’s ability to work and perform regular activities.

The percent work time missed due to problem (absenteeism), percent impairment while working due to problem (presenteeism), percent overall work impairment due to problem, and percent activity impairment due to problem are derived on the eCRF directly. Details about derivations can be found in [Appendix 6](#).

17.1.4. Urgency Numeric Rating Scale

The urgency NRS is a single item that measures the severity for the urgency (sudden or immediate need) to have a bowel movement in the past 24 hours using an 11-point NRS ranging from 0 (no urgency) to 10 (worst possible urgency).

17.1.5. Abdominal Pain NRS

The abdominal pain NRS is a single item that measures the “worst abdominal pain in the past 24 hours” using an 11-point NRS ranging from 0 (no pain) to 10 (pain as bad as can imagine).

17.1.6. UC-Related Hospitalizations and Surgeries

The UC-related hospitalizations are captured on the Adverse Events eCRF. The UC-related surgeries, including colectomy, are captured on the Non-Drug Treatment eCRF. Blinded review of AE preferred terms will be performed to determine UC-related hospitalizations.

17.2. Missing Data Methods

No missing data will be imputed for HRQoL endpoints.

18. SAFETY OUTCOMES

All outputs for safety outcomes will be based on the Safety Set. There will be no statistical comparisons between the treatment groups for safety data, unless otherwise specified.

18.1. Adverse Events

AEs will be coded using MedDRA (v24.1). The version used to code AEs will be displayed in the analyses. Treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first dose of study treatment.

Refer to [Appendix 2](#) for handling of partial dates for AEs for the purpose of assigning treatment-emergent flags. In the case where it is not possible to define an AE as treatment-emergent or not, the AE will be classified as treatment-emergent.

18.1.1. All TEAEs

All TEAEs will be summarized by SOC and PT. This summary table and all other TEAE summaries by SOC and PT will be presented by descending frequency in the etrasimod group. All AEs, regardless of treatment-emergent status, will be included in an AE listing. Additionally, a listing of other AE details as collected on the CRF will be presented.

Exposure-adjusted incidence rate (EAIR) of TEAEs will be summarized by SOC and PT. Exposure is defined as the sum of either time (year) from first dose to the onset of first such event for those who experienced this AE, or time (year) from first dose to last participation for those who did not experience this AE. The EAIR is calculated as the number of subjects with the AE divided by the total exposure in subject-years.

18.1.1.1. Severity

Severity is classified as Grade 1: Mild, Grade 2: Moderate, Grade 3: Severe, Grade 4: Life-threatening, Grade 5: Death related to AE, using the Common Terminology Criteria for Adverse Events, v5.0 (CTCAE v5.0).

All TEAEs will be summarized by SOC, PT, and maximum severity, with SOC and PT presented by descending frequency. If a subject reports a TEAE more than once within a SOC/PT, the AE with the worst-case severity will be used in this summary.

18.1.1.2. Relationship to Study Treatment

Relationship is classified as “not related”, “unlikely related”, “probably related”, or “related” by the Investigator.

All related TEAEs will be summarized by SOC and PT. A “related TEAE” for the purpose of this summary is defined as a TEAE with relationship to study treatment of “probably related” or “related”.

All TEAEs will be summarized by SOC, PT, and highest relationship (as reported on the eCRF, not grouped), with SOC and PT presented by descending frequency. If a subject reports a TEAE more than once within a SOC/PT, the AE with the worst-case relationship to study treatment will be used in this summary. TEAEs with a missing relationship to study treatment will be regarded as “Related” to study treatment for summary tabulation purpose only.

18.1.2. TEAEs Leading to Discontinuation of Study Treatment

TEAEs leading to discontinuation of study treatment will be identified by action taken being recorded as “Drug withdrawal” on the Adverse Events eCRF.

All TEAEs leading to discontinuation of study treatment will be summarized by SOC and PT. A listing of all TEAEs leading to discontinuation of study treatment will also be presented.

Time to study treatment discontinuation due to liver-related AE will be summarized for the Safety Set for subjects who had such an event reported during the study. Descriptive statistics, Kaplan-Meier estimate of 25th, 50th, and 75th percentile time to event, 95% CIs, and p-value from the log rank test will be displayed. If a subject did not have the event, they will be censored at the date of last dose. If the date of last dose is missing, the date of last study visit up to Week 52 will be used. If no event is ever observed, then summary will not be generated.

Liver-related AEs are defined by either SOC of hepatobiliary disorders or PT of alanine aminotransferase abnormal, alanine aminotransferase increased, aspartate aminotransferase abnormal, aspartate aminotransferase increased, hepatic enzyme abnormal, hepatic enzyme increased, liver function test abnormal, liver function test increased, transaminases abnormal, or transaminases increased.

18.1.3. Serious and Non-Serious TEAEs

Serious adverse events (SAEs) are those events recorded as “Serious” on the Adverse Events eCRF.

All serious TEAEs will be summarized by SOC and PT. If the seriousness is missing, the AE will be considered as “Serious” for summary tabulation purpose only. A serious TEAE listing will also be presented. All nonserious TEAEs will be summarized by SOC and PT.

18.1.4. TEAEs Leading to Death

TEAEs leading to Death are those events which are recorded with an outcome as “Fatal” on the Adverse Events eCRF. A listing of TEAEs leading to death will be presented.

Time to death due to liver-related AE will be summarized for the Safety Set for subjects who had such an event reported during the study. Descriptive statistics, Kaplan-Meier estimate of 25th, 50th, and 75th percentile time to event, 95% CIs, and p-value from the log-rank test will be displayed. If a subject did not have the event, they will be censored at the date of last dose. If the date of last dose is missing, the date of last study visit up to Week 52 will be used. If no event is observed, then the summary will not be generated.

18.1.5. TEAEs of Special Interest

Categories of Targeted Medical Events (TMEs) and a list of preferred terms associated with these TME categories were developed based on the mechanism of action of etrasimod, prior experience with other agents acting via a similar mechanism, and disease-specific clinical judgment. In addition, where standard testing has been implemented to screen for potential AESI (eg, electrocardiograms, spirometry, serum transaminases, etc.), the relevant data will be reviewed to identify potential cases of AESI that investigators may not have identified and to provide quantitative data for AESIs. The proposed candidate terms will be reviewed to identify which events reflect AESI.

TEAEs of special interest will be summarized by category, subcategory, and PT by descending frequency by treatment group. Categories and subcategories of TEAEs of special interest are the following:

- Cardiovascular Events
 - Bradycardia
 - AV conduction delay
 - Hypertension
- Macular Edema
- Pulmonary Disorders
 - Airflow obstruction (forced expiratory volume in 1 second [FEV₁], forced vital capacity [FVC])
 - Decrease gas exchange (diffusing capacity of the lungs for carbon monoxide [DLCO])
- Infections
 - Severe infections
 - Opportunistic infections (Narrow)
 - Herpes simplex and herpes zoster

- Liver Injury
 - Liver transaminases elevation
 - Bilirubin elevation
- Posterior Reversible Encephalopathy Syndrome (PRES)
- Malignancies

For each category, subcategory, and PT, model-based analysis will be performed to compare treatment effect between etrasimod and placebo. An extended Cox regression model will be used to handle recurring AEs (Cao 2011). The stratification variables, naïve to biologic or JAK inhibitor therapy status at study entry (Yes or No), Baseline corticosteroid use (Yes or No), and Baseline disease activity (MMS: 4 to 6 or 7 to 9), will be included in the model. A subject with K events contributes (K+1) observations to the input data set; here K might be 0. The k-th observation of the subject identifies the time interval from the start date of the (k-1)-th event, or time 0 (if k = 1), to the start date of the kth event, k = 1, ..., K. The (K+1)-th observation represents the time interval from the K-th event to the date of censorship, ie, the date of last dose or date of last study visit up to Week 52, whichever is later. Hazard ratios, 95% CIs, and p-values for treatment effect comparison will be presented.

18.1.6. Overall Summary of Adverse Events

In addition to the summaries above, an overview of TEAEs will be summarized (not broken down by SOC or PT) by number and frequency of subjects and by number of AEs:

- Any TEAE
- Any related TEAE*
- Any serious TEAEs
- Any related serious TEAEs*
- TEAEs leading to death
 - Liver-related TEAEs leading to death
- TEAEs leading to study treatment discontinuation
 - Related TEAEs leading to study treatment discontinuation*
 - Liver-related TEAEs leading to study treatment discontinuation
- TEAEs leading to study treatment interruption
 - Related TEAEs leading to study treatment interruption*
- TEAEs by maximum severity
- Related TEAEs by maximum severity*
- TEAEs by relationship to study treatment

* “Related TEAEs” refers to TEAEs related or probably related to study treatment or are missing relationship.

Comparative analyses will be performed for:

- Deaths
- Serious TEAEs
- TEAEs leading to study treatment discontinuation
- TEAEs of special interest
- Serious infections
- Opportunistic infections (Narrow)
- Malignancies

Risk differences between etrasimod and placebo along with 95% CIs using the Wilson’s score method will be presented.

18.2. Deaths

Information collected about death (eg, date of death, primary cause of death) will be presented in a data listing, as described in Section [18.1.4](#).

18.3. Laboratory Evaluations

No local laboratory assessments will be used in any summaries except for lipid panel and thyroid panel tests. No local laboratory assessments will be used to derive maximum/minimum/worst value. Local laboratory assessments will be listed.

18.3.1. Safety Laboratory Evaluations

Hematology, serum chemistry, coagulation, and urinalysis are analyzed and reported by central laboratory and sometimes by local laboratory. Results out of reference range are flagged by the performing laboratory (eg, low, high). A full list of laboratory assessments to be included in the outputs is included in Clinical Study Protocol APD334-301 Table 6.

In general, presentations will use SI units. Quantitative laboratory measurements reported as “< X” or “> X”, where X may be the lower limit of quantitation (LLQ) or the upper limit of quantitation (ULQ), respectively, will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, ie, as “< X” or “> X” in the listings. For urinalysis, only pH and specific gravity are considered as quantitative tests. If local and central laboratory assessments are available from the same day, central laboratory assessments will be used in the summary as per Section [6.4](#).

The following summaries will be provided for laboratory data:

- Value and change from Baseline by visit (for hematology, serum chemistry, quantitative urinalysis [pH and specific gravity], and coagulation)

- Incidence of abnormal values according to laboratory reference ranges by visit
- Incidence of lymphocytes $< 0.2 \times 10^9/L$, $0.5 \times 10^9/L$, or neutrophils $< 0.5 \times 10^9/L$, $1 \times 10^9/L$ at end of treatment and anytime in the study
- Shift from end of treatment to each follow-up visit in incidence of lymphocytes in normal range and in incidence of lymphocytes of at least 80% percent of Baseline, by visit (2-Week Follow-Up, 4-Week Follow Up, and Last Follow-Up)
- Shift from Baseline to end of treatment according to laboratory reference range (for quantitative measurements and categorical measurements)
- Evaluation of drug-induced serious hepatotoxicity (eDISH) plots for the following laboratory assessments (using values after the first administration of study treatment):
 - Maximum aspartate aminotransferase (AST) versus same-day total bilirubin
 - Maximum alanine aminotransferase (ALT) versus same-day total bilirubin
 - Maximum gamma-glutamyl transferase (GGT) versus same-day total bilirubin
 - Maximum alkaline phosphatase (ALP) versus same-day total bilirubin

The eDISH plots above will be repeated with adjustment for elevated Baseline.

- Incidence of hepatic enzyme elevations by visit
 - $> 1 \times, 2 \times, 3 \times, 5 \times, 8 \times, 10 \times, 20 \times$ upper limit of normal (ULN) elevation in ALT
 - $> 1 \times, 2 \times, 3 \times, 5 \times, 8 \times, 10 \times, 20 \times$ ULN elevation in AST
 - $> 3 \times, 5 \times, 10 \times, 20 \times$ ULN elevation in either ALT or AST
 - $> 1 \times, 1.5 \times, 2 \times, 3 \times$ ULN elevation in total bilirubin
 - $> 1 \times, 1.5 \times, 2 \times, 3 \times, 5 \times, 8 \times$ ULN elevation in ALP
 - $> 1 \times, 2 \times, 3 \times, 5 \times, 8 \times$ ULN elevation in GGT
 - $> 3 \times$ ULN elevation in either ALT or AST and $> 1.5 \times$ ULN elevation in total bilirubin
 - $> 3 \times$ ULN elevation in either ALT or AST and $> 2 \times$ ULN elevation in total bilirubin
 - $> 3 \times$ ULN elevation in either ALT or AST and $> 1.5 \times$ ULN elevation in ALP
 - $> 3 \times$ ULN elevation in either ALT or AST in temporal association with treatment-emergent nausea, vomiting, anorexia, abdominal pain, or fatigue identified by PT, where temporal association is defined as ± 14 days of onset date from the time of elevation.

If both central and local assessments of total bilirubin are available on the same day, the central result will take precedence over the local result in the eDISH plot. If the maximum AST, ALT, GGT, or ALP assessment occurs at 2 different dates, the assessment with the higher

accompanying Bilirubin value will be used. Two versions of the eDISH plots will be presented, with one showing values as multiples of ULN, and the other one showing values as multiples of ULN, or subject's baseline, whichever is higher.

Only laboratory tests completed by 50 or more subjects in the Safety Set will be presented in the summaries. All laboratory tests will be listed.

Subject's laboratory assessments at all timepoints will be listed in chronological order. Values outside of the laboratory reference range will be flagged. Values obtained from local laboratory will be flagged. A listing of lymphocytes and neutrophils over time in subjects who ever had lymphocytes $< 0.5 \times 10^9/L$ or neutrophils $< 1 \times 10^9/L$ will also be provided.

18.3.2. Pregnancy Tests

Urine beta-human chorionic gonadotropin (β -hCG) and/or serum β -hCG pregnancy tests are performed throughout the study in female subjects of childbearing potential. All pregnancy test results (positive or negative) will be listed in chronological order.

18.3.3. Other Screening Laboratory Assessments

Screening laboratory assessments for virology, drug screen/toxicology, QuantiFERON Tuberculosis (TB) Gold, stool pathogens, and *Clostridioides difficile* (formerly known as *Clostridium difficile*) will be listed. Analyses of genetics [REDACTED] data based on samples collected at Screening in subjects who provided consent will be described in a separate plan.

18.4. ECG Evaluations

ECGs are recorded on a 12-lead ECG machine and read centrally. The following ECG parameters are reported for this study:

- HR (bpm)
- PR interval (ms)
- RR interval (ms)
- QRS interval (ms)
- QT interval (ms)
- QTcF interval (ms)
- Overall interpretation of ECG (Investigator's judgment):
 - Normal
 - Abnormal, Not Clinically Significant (Abnormal NCS)
 - Abnormal, Clinically Significant (Abnormal CS)

- Overall interpretation of ECG (central reader):
 - Normal
 - Abnormal NCS
 - Abnormal CS
- AV conduction abnormalities
 - First-degree AV block
 - Second-degree AV block type 1
 - Second-degree AV block type 2
 - Third-degree AV block

The following summaries will be provided for ECG data:

- Value and change from Baseline by visit (for quantitative measurements)
- Incidence of markedly abnormal values (defined in Section 18.4.1) and AV blocks by visit
- Shift in normal/abnormal NCS/abnormal CS in the overall interpretation (by investigator) from Baseline to end of treatment and to the worst-case post-Baseline
- Shift in markedly abnormal categories from Baseline to post-Baseline by visit

Listings of ECG results, including first dose cardiac monitoring, and discharge criteria for first dose cardiac monitoring will be provided.

A listing of all ECG assessments over time in subjects meeting markedly abnormal criteria will also be provided. For each subject, only ECG parameters ever meeting markedly abnormal criteria will be included.

18.4.1. ECG Markedly Abnormal Criteria

Markedly abnormal quantitative ECG measurements will be identified in accordance with the following predefined markedly abnormal criteria:

- Absolute values in QT and QTcF:
 - ≥ 450 ms (male) or ≥ 470 ms (female) in QTcF
 - > 500 ms in QT
- Change from Baseline in QT and QTcF:
 - > 30 ms increase from Baseline
 - > 60 ms increase from Baseline

In shift tables, subjects will be classified according to each binary category for each parameter and the predefined markedly abnormal criterion (ie, Markedly abnormal versus Not markedly abnormal, with Markedly abnormal defined in the footnote).

18.5. Vital Signs

The following vital signs measurements are reported for this study:

- Systolic blood pressure (mm Hg)
- Diastolic blood pressure (mm Hg)
- Heart rate (bpm)
- Respiratory rate (resp/min)
- Temperature (°C)
- Weight (kg)
- Height (cm) (Screening only)

The following summaries will be provided for vital signs data:

- Value and change from Baseline by visit
- Value and change from predose on Day 1 (as reported on the eCRF) by timepoint
 - For heart rate only, value and change from predose to minimum postdose heart rate on Day 1 will be included in the same table
- Incidence of markedly abnormal values (defined in [Table 4](#) by visit)
- Listing of subjects meeting markedly abnormal criteria
- Incidence of minimum heart rate on Day 1 by postdose timepoint (1, 2, 3, 4, and > 4 hours postdose, and Day 1 overall) and heart rate interval (≥ 65 , 60 to 64, 55 to 59, 50 to 54, 45 to 49, 40 to 44, < 40 bpm)
 - If minimum heart rate is attained at multiple postdose timepoints, only the earliest timepoint will be counted in this incidence summary
- Time to minimum heart rate on Day 1 by planned hourly timepoint (if minimum heart rate is attained at multiple postdose timepoints, only the earliest timepoint will be counted).
 - Actual time elapsed from first dose to minimum heart rate on Day 1 will also be summarized using descriptive statistics (n, mean, SD, median, minimum, and maximum) in the same table
- Line plots for mean value and mean change from predose on Day 1, up to 4 hours postdose, by parameter and timepoint (systolic blood pressure, diastolic blood pressure, and heart rate)
- Line plots for mean value and mean change from Baseline by parameter and visit (systolic blood pressure, diastolic blood pressure, heart rate, body temperature, and respiratory rate)

Listings of all vital signs, including first dose cardiac monitoring, and discharge criteria for first dose cardiac monitoring will be provided. A listing of vital signs assessments over time in subjects meeting a markedly abnormal criterion will also be provided. For each subject, only the parameters with at least one markedly abnormal criterion satisfied will be included.

For subjects with extended monitoring or remonitoring, a listing of systolic blood pressure, diastolic blood pressure, heart rate values, and change from predose in all parameters on Day 1, Day 2, and any remonitoring visit will be provided.

18.5.1. Vital Signs Specific Derivations

- Temperature (°C) = (5/9) (Temperature (°F) – 32)

18.5.2. Vital Signs Markedly Abnormal Criteria

Markedly abnormal quantitative vital signs measurements will be identified in accordance with the following predefined markedly abnormal criteria:

Table 4: Markedly Abnormal Criteria for Vital Signs

Variable	Unit	Low	High
SBP	mm Hg	≤ 90 mm Hg	> 150 mm Hg
DBP	mm Hg	≤ 50 mm Hg	> 90 mm Hg
Heart rate	bpm	< 40 bpm < 50 bpm < 50 bpm and decrease from predose (Baseline) of > 10 bpm at 4 hours on Day 1 or Day 2 or remonitoring visit	> 100 bpm

bpm, beats per minute; DBP, diastolic blood pressure; SBP, systolic blood pressure

18.6. Physical Examination

A listing of all physical examination assessments in subjects who had at least 1 abnormal physical examination finding will be provided.

18.7. Other Safety Assessments

18.7.1. Pulmonary Function Tests

The following pulmonary function test (PFT) measurements (actual and % Predicted) will be reported for this study:

- FEV₁
- FVC
- Total Lung Capacity (TLC)
- FEV₁/FVC ratio
- Forced Expiratory Flow (FEF) 25-75

- DLCO (if available)

The following summaries will be provided for PFT data:

- Value and change from Baseline by visit
- Incidence of markedly abnormal values by visit (Section 18.7.1.1)

All PFT data will be listed. A listing in subjects who ever reported an abnormality in PFT will also be provided, including a flag for whether markedly abnormal criterion is also met.

18.7.1.1. PFT Markedly Abnormal and Potentially Important Criteria

Markedly abnormal quantitative PFT measurements will be identified in accordance with the following predefined markedly abnormal criteria:

- % Predicted FEV₁ < 50%
- % Predicted FVC < 50%
- % Predicted FEV₁/FVC ratio < 50%

Potentially important PFT measurements will also be identified using the criteria below:

- Decrease from Baseline > 20% in FEV₁, ie, percent change from Baseline < -20%
- Decrease from Baseline > 20% in FVC, ie, percent change from Baseline < -20%
- Decrease from Baseline > 20% in DLCO, ie, percent change from Baseline < -20%

18.7.2. Ophthalmoscopy and Optical Coherence Tomography

The following summaries will be provided for ophthalmoscopy and optical coherence tomography (OCT) data:

- Values and change from Baseline in central foveal thickness and intraocular pressure by visit
- Categorical result in ophthalmoscopy with OCT parameters by visit (the categories are listed on the eCRF)

Only PFT and OCT assessments completed by 50 or more subjects in the Safety Set will be presented in the summaries.

A listing of all ophthalmoscopy and OCT assessments will be provided. A listing of subjects who ever reported an abnormality in OCT will also be provided. Additionally, a listing of retinal photograph and eye pressure assessments will be provided for subjects who experienced an AE related to eye disorders.

18.7.3. Tuberculosis Screening and Chest X-Ray

Screening TB results and chest X-rays will be listed. Results of TB questionnaires will also be listed.

19. EFFICACY-RELATED BIOMARKERS

Value and change from Baseline in level of FCP, hs-CRP, and lymphocyte counts at protocol-specified visits will be summarized by visit and treatment for the FAS with Baseline MMS 5 to 9. Additionally, percent change from Baseline in lymphocyte counts will also be summarized. Change from Baseline in these biomarkers will also be analyzed using an MMRM, with factors for naïve to biologic or JAK inhibitor therapy at study entry (Yes or No), Baseline corticosteroid use (Yes or No), Baseline disease activity (MMS: 4 to 6 or 7 to 9), treatment, visit, treatment by visit interaction, a covariate of the corresponding Baseline value, and a random subject effect. Reported randomization stratum will be used in the model. Unstructured covariance will be used. If there is a convergence issue in fitting the unstructured covariance, a compound symmetry covariance structure will be used. The same model will be fit to the percent change from Baseline in lymphocytes. LS means, SEs, CIs, and p-values by visit will be reported.

Additionally, the following summaries will be provided for biomarker data:

- Line plots over time for LS mean and SE for the following laboratory assessments:
 - Change from Baseline in lymphocytes
 - Percent change from Baseline in lymphocytes
 - Change from Baseline in hs-CRP
 - Change from Baseline in FCP

The exploratory efficacy-related biomarker analyses for immunophenotyping, proteomics, RNA transcriptomics, and fecal microbiome will be described in a separate biomarker analysis plan.

20. PHARMACOKINETICS

Plasma concentrations of etrasimod, and if warranted M3 (AR503641) and M6 (AR504344) will be assessed from PK samples collected prior to dosing and 4 hours (\pm 15 minutes) postdose (after 12-lead ECG) on Week 0/Day 1. Additionally, PK samples will be collected prior to dosing (trough) at Weeks 2, 4, 8, 12, 16, 20, 24, 32, 40, 48, and 52, and for PK samples collected at the 2-Week and 4-Week Follow-Up visits for subjects not enrolled into any extension study. Concentrations below the limit of quantitation (BLQ) will be assigned a numerical value of zero for the calculation of descriptive statistics. For geometric mean and geometric % coefficient of variation [CV], the zero values will be excluded.

The pharmacokineticist will determine the strategy for dealing with data affected by protocol deviations or events which may impact the quality of etrasimod concentration data on a case-by-case basis with input from the Arena study physician and Arena clinical pharmacologist, as needed. Examples for protocol deviations or events include, but may not be limited to, vomiting within 4 hours after etrasimod dose administration on Week 0/Day 1 or on the day prior to trough sample collection, sample processing errors that lead to inaccurate bioanalytical results, and/or inaccurate dosing prior to PK sampling. In the case of an important protocol deviation or event, the affected PK data collected may be excluded from the summaries, calculation of the

average steady-state trough plasma concentration ($C_{\text{trough,ss}}$) based on Week 2 to Week 52 trough concentrations ($C_{\text{trough,ss,W2-W52}}$), exposure-response analysis, and/or population PK analysis, but will still be reported in the study result listings.

Unless otherwise specified, PK summaries will use the Pharmacokinetic Set. Exposure-response analysis will use the Safety Set.

Individual subject etrasimod plasma concentrations (and etrasimod metabolite concentrations, if applicable) will be presented in the data listings, including subject ID, treatment received, sex, age, weight, tobacco use, nominal timepoint, actual blood collection date/time, concentration, and time since last dose and also summarized using descriptive statistics (n, mean, SD, % CV, geometric mean, geometric % CV, median, minimum, and maximum) by nominal timepoint.

All subjects are expected to achieve steady-state plasma concentration by Week 2. For each subject, the average steady-state trough (predose) plasma concentration based on Week 2 to Week 52 trough concentrations ($C_{\text{trough,ss,W2-W52}}$) per subject will also be calculated and presented in a data listing and summarized using descriptive statistics.

Mean (\pm SD) etrasimod concentration versus nominal time (Weeks 0 to 52) will be plotted on linear scales. Additionally, a mean (\pm SD) concentration-time plot with etrasimod and metabolites overlaid on linear and semi-logarithmic scale will be generated.

Box plots of etrasimod, M3, and M6 concentrations at Day 1, 4-hour postdose timepoint (C_{4h}), steady-state C_{trough} (Weeks 2 through 52), and $C_{\text{trough,ss,W2-W52}}$ will be plotted versus nominal timepoint on ordinal scale.

Individual subject plasma concentrations versus actual time (Weeks 0 to 52) will be plotted since start of treatment on linear and semi-logarithmic scales. Individual subject plots with etrasimod and metabolite concentration data will be overlaid on the same plot.

Scatter plots of individual subject average steady-state $C_{\text{trough,ss,W2-W52}}$ versus Baseline body weight and versus Baseline age (age at consent) as continuous variables will be generated for the Pharmacokinetic Set, which will also include the Spearman's rank correlation coefficient, p-value, and locally estimated scatterplot smoothing (LOESS) trend line.

To explore potential etrasimod plasma exposure-response relationships, the following scatter plots will be generated for the Safety Set, including data from placebo subjects (plotted at 'zero' concentration). Any contributing data from the placebo subjects will be annotated with a different color or symbol. All scatter plots will include the Spearman's rank correlation coefficient, p-value, and LOESS trend line (based on subjects that received active treatment only).

- Individual subject absolute change from Baseline in MMS versus etrasimod steady-state C_{trough} at Week 12
- Individual subject absolute change from Baseline in MMS versus etrasimod steady-state C_{trough} at Week 52
- Individual subject lymphocytes versus etrasimod steady-state C_{trough} at Week 52 and grouped all visits (Weeks 2 through 52 pooled)

- Individual subject absolute changes from Baseline in lymphocytes versus etrasimod steady-state C_{trough} at Week 52 and grouped all visits (Weeks 2 through 52 pooled)
- Individual subject percent changes from Baseline in lymphocytes versus etrasimod steady-state C_{trough} at Week 52 and grouped all visits (Weeks 2 through 52 pooled)
- Individual subject heart rate versus etrasimod concentration at matching visit/timepoint for Week 0/Day 1; the 4-hour postdose vital results will be presented on the left panel and safety ECG on the right panel.
- Individual subject absolute changes from Day 1 Predose in heart rate versus etrasimod concentration at matching visit/timepoint for Week 0/Day 1; the 4-hour postdose vital results will be presented on the left panel and safety ECG on the right panel.

Furthermore, the following overlay plots will be generated for the Pharmacokinetic Set within the etrasimod group only:

- Mean (SD) absolute lymphocyte counts and mean (SD) etrasimod concentration versus nominal timepoint (Weeks 0 through 52)
- Mean (SD) absolute change from Baseline in lymphocytes and mean (SD) etrasimod concentration versus nominal timepoint (Weeks 0 through 52)
- Mean (SD) percent change from Baseline in lymphocytes and mean (SD) etrasimod concentration versus nominal timepoint (Weeks 0 through 52)
- Mean (SD) heart rate and mean (SD) etrasimod concentration versus nominal timepoint (Weeks 0 through 52)
- Mean (SD) absolute change from Baseline in heart rate and mean (SD) etrasimod concentration versus nominal timepoint (Weeks 0 through 52)

All summary figures for etrasimod plasma concentrations and for change from Baseline in MMS versus etrasimod steady-state C_{trough} will be repeated by:

- Sex (Male or Female)
- Weight (\leq Median or $>$ Median)
- Age group (< 18 , 18 to < 65 , or ≥ 65 years)
- Tobacco use (Yes or No)

The plasma concentrations over time will be used in a population PK analysis, which will be described in a separate plan.

21. DATA NOT SUMMARIZED OR PRESENTED

The other variables and/or domains not summarized are:

- Comments

These domains and/or variables will not be summarized but will be available in the clinical study database and Study Data Tabulation Model (SDTM) datasets.

REFERENCES

- Cao D, He X. Statistical analysis of adverse events in randomized clinical trials using SAS. PharmaSUG Conference Proceedings; 2011.
- FDA. Multiple endpoints in clinical trials. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry Web site. <https://www.fda.gov/media/102657/download>. Published 2017. Accessed February 19, 2021.
- FDA. Statistical considerations for clinical trials during the COVID-19 public health emergency. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry Web site. <https://www.fda.gov/media/139145/download>. Published 2020. Accessed February 19, 2021.
- ICH. Structure and content of clinical study reports, E3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. https://database.ich.org/sites/default/files/E3_Guideline.pdf. Published 1995. Accessed March 10, 2021.
- ICH. Statistical principles for clinical trials, E9. International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. https://database.ich.org/sites/default/files/E9_Guideline.pdf. Published 1998. Accessed March 19, 2021.
- ICH. Structure and content of clinical study reports questions & answers (R1). https://database.ich.org/sites/default/files/E3_Q%26As_R1_Q%26As.pdf. Published 2012. Accessed Dec 14, 2021.
- ICH. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials, E9(R1). International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf. Published 2019. Accessed March 19, 2021.
- Rubin DB. *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley and Sons; 1987. On file at Arena.

APPENDICES

APPENDIX 1: PROGRAMMING CONVENTIONS FOR OUTPUTS

OUTPUT CONVENTIONS

Outputs will be presented as shown in the Output shells.

DECIMALS, PERCENTAGES, AND P-VALUES

- If the original data has N decimal places, then the summary statistics should have the following decimal places:
 - Minimum and maximum: N
 - Mean (and LS Means), median: N + 1
 - SD or SE: N + 2
- Percentages will be reported to one decimal place. Where counts are zero, percentages will not appear in the output.
- P-values will be reported to three decimal places, except values < 1.000 but > 0.999 will be presented as ' > 0.999 ' (eg, 0.9998 is presented as > 0.999).
- Values < 0.001 will be presented as ' < 0.001 ' (eg, 0.0009 is presented as < 0.001). Rounding will be applied after the < 0.001 and > 0.999 rule.

CONVENTIONS RELATED TO PHARMACOKINETIC DATA

All etrasimod concentrations will be reported and analyzed with the same precision as the source data provided by the bioanalytical laboratory regardless of how many significant figures or decimals the data carry. The derived PK data (etrasimod concentration) will be rounded to carry 3 significant digits and considered the source data for the calculation of descriptive statistics and the statistical analysis source data.

For the reporting of descriptive statistics, the mean, geometric mean, median, SD, % CV, and % geometric CV will be presented with 4 significant digits. The minimum and maximum will be presented with 3 significant digits.

DATES AND TIMES

Depending on data available, dates and times will take the form DDMMYYYY or DDMMYYYY:hh:mm.

SPELLING FORMAT

English US.

PRESENTATION OF TREATMENT GROUPS

For outputs, treatment groups will be represented as follows and in that order:

Treatment Group	For Tables, Graphs, and Listings
Etrasimod	Etrasimod
Placebo	Placebo
Screen Failure	Screen Failure
Not Treated ^a	Not Treated

^a To be used for subjects in safety listings who are randomized but do not receive study treatment.

PRESENTATION OF VISITS AND STUDY PERIOD

For outputs, visits and study periods will be represented as follows and in that order:

Visit Name	Study Period
Screening	Screening
Baseline, Day 1, Week 2, Week 4, Week 8, Week 12, Week 16, Week 20, Week 24, Week 32, Week 40, Week 48, and Week 52	Treatment
End of Study/Early Termination	Dependent on Analysis Visit assigned, described in Section 6.4
2-Week Follow-Up, 4-Week Follow-Up	Follow-Up

LISTINGS

All listings will be ordered by the following (unless otherwise indicated in the template):

- Randomized treatment group (or treatment received if it is a safety output), first by Etrasimod, then Placebo, then Screen Failure and then No Treatment (only in safety listings if there are any randomized subjects who did not receive study treatment)
- Subject number (which is expected to incorporate study site/center)
- Date (where applicable)

APPENDIX 2: PARTIAL DATE CONVENTIONS

Imputed dates will NOT be presented in the listings.

ALGORITHM FOR TREATMENT EMERGENCE OF ADVERSE EVENTS:

AE Start Date	AE Stop Date	Action
Known	Known, Partial, or Missing	If AE start date < study treatment first dose date, then not TEAE If AE start date ≥ study treatment first dose date, then TEAE
Partial, but known components show that it cannot be on or after date of first dose of study treatment	Known, Partial, or Missing	Not TEAE
Partial, could be on or after date of first dose of study treatment	Known	If AE stop date < study treatment first dose date, then not TEAE If AE stop date ≥ study treatment first dose date, then TEAE
	Partial	Impute AE stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < study treatment first dose date, then not TEAE If AE stop date ≥ study treatment first dose date, then TEAE
	Missing	Assumed TEAE
Missing	Known	If AE stop date < study treatment first dose date, then not TEAE If AE stop date ≥ study treatment first dose date, then TEAE
	Partial	Impute AE stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < study treatment first dose date, then not TEAE If AE stop date ≥ study treatment first dose date, then TEAE
	Missing	Assumed TEAE

ALGORITHM FOR PRIOR AND CONCOMITANT MEDICATIONS:

Start Date	Stop Date	Action
Known	Known	If medication stop date < study treatment first dose date, assign as prior If medication stop date ≥ study treatment first dose date, assign as concomitant
	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study treatment first dose date, assign as prior If medication stop date ≥ study treatment first dose date, assign as concomitant
	Missing	If medication stop date is missing could never be assumed a prior medication, assign as concomitant
Partial	Known	Impute start date as earliest possible date (ie, first day of month if day unknown or 1st January if day and month are unknown), then: If medication stop date < study treatment first dose date, assign as prior If medication stop date ≥ study treatment first dose date, assign as concomitant
	Partial	Impute start date as earliest possible date (ie, first day of month if day unknown or 1st January if day and month are unknown) and impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study treatment first dose date, assign as prior If medication stop date ≥ study treatment first dose date, assign as concomitant
	Missing	Impute start date as earliest possible date (ie, first day of month if day unknown or 1st January if day and month are unknown), then: If medication stop date is missing could never be assumed a prior medication, assign as concomitant
Missing	Known	If medication stop date < study treatment first dose date, assign as prior Else assign as concomitant
	Partial	Impute stop date as latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study treatment first dose date, assign as prior If medication stop date ≥ study treatment first dose date, assign as concomitant
	Missing	Assign as concomitant

ALGORITHM FOR STARTING DATE OF HOSPITALIZATION AND START/STOP DATES OF CONCOMITANT MEDICATIONS:

Start Date	Stop Date	Action
Partial	Partial	Impute start date as earliest possible date (ie, first day of month if day unknown or 1st January if day and month are unknown). Impute stop date to latest possible date (ie, last day of month if day unknown or 31st December if day and month are unknown). No partial stop date imputation for hospitalization.
Missing	Missing	Impute start date as informed consent date. Impute stop date as last visit date when the concomitant medication timepoint is not ongoing. No missing stop date imputation for hospitalization.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

APPENDIX 4: IBDQ SCORING RULES

The Inflammatory Bowel Disease Questionnaire (IBDQ) is a 32-item self-administered questionnaire which has 4 dimensions: Bowel symptoms (10 items), systemic symptoms (5 items), emotional health (12 items), and social function (5 items). Responses are graded on a 7-point Likert scale where 7 denotes “not a problem at all” and 1 denotes “a very severe problem”. Scores range from 32 to 224, a higher score indicates better quality of life.

The 4 dimensions are defined as:

- Bowel symptoms: Questions 1, 5, 9, 13, 17, 20, 22, 24, 26, 29
- Systemic symptoms: Questions 2, 6, 10, 14, 18
- Emotional health: Questions 3, 7, 11, 15, 19, 21, 23, 25, 27, 30, 31, 32
- Social function: Questions 4, 8, 12, 16, 28

APPENDIX 5: SF-36 SCORING RULES

The Medical Outcomes Study 36 Item Short Form Health Survey (SF-36) is a 36-item self-administered questionnaire which has 8 scales: Physical functioning (10 items), role limitations due to physical health (4 items), role limitations due to emotional problems (3 items), vitality (4 items), mental health (5 items), social functioning (2 items), bodily pain (2 items), and general health (5 items). Each item response is scored from 0 to 100 and items in the same scale are averaged together to create 8 subscores. Scores range from 0 to 100. A higher score indicates a more favorable health state.

The 8 dimensions are defined as:

- Physical functioning: Questions 3 to 12
- Role limitations due to physical health: Questions 13 to 16
- Role limitations due to emotional problems: Questions 17 to 19
- Vitality: Questions 23, 27, 29, 31
- Mental health: Questions 24 to 26, 28, 30
- Social functioning: Questions 20, 32
- Bodily Pain: Questions, 21, 22
- General health: Questions 1, 33 to 36

APPENDIX 6: WPAI-UC SCORING RULES

The Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis (WPAI-UC) questionnaire was developed to measure the effect of general health and symptom severity on work productivity and regular activities. The questionnaire includes the following questions.

Questions:

- 1 = Are you currently employed?
- 2 = During the past seven days, how many hours did you miss from work due to your problems associated with UC?
- 3 = During the past seven days, how many hours did you miss from work because of any other reasons?
- 4 = During the past seven days, how many hours did you actually work?
- 5 = During the past seven days, how much did your UC affect your productivity while you were working?
- 6 = During the past seven days, how much did your UC affect your ability to do your regular daily activities, other than work at a job?

Derivation:

- 1) Calculate work time missed score as: $Q2/(Q2 + Q4)$
- 2) Calculate impairment while working score as: $Q5/10$
- 3) Calculate overall work impairment score as: $Q2/(Q2 + Q4) + [(1 - Q2/(Q2 + Q4)) \times (Q5/10)]$
- 4) Calculate activity impairment score as: $Q6/10$

Multiply each score by 100 in order to express as percentages.

APPENDIX 7: CLASSIFICATION OF RESCUE THERAPY

This appendix outlines the algorithm for the Arena clinical team/medical reviewers to classify rescue therapies for ulcerative colitis (UC) in a blinded manner. This will help establish 1) intercurrent events for the efficacy estimands, and 2) whether a subject will be excluded from the Per Protocol Set(s). All rescue therapies identified by the medical reviewers per the algorithm below will be imported in programming. This process will be repeated until database lock and the list of all rescue therapies identified will be finalized before study unblinding.

Only medications and medical procedures reported on the electronic case report forms (eCRFs) can be assessed whether they are rescue therapy for UC. If the exposure happens in the follow-up period (beginning on or after the date of last study treatment administration), then it would not be considered as a rescue therapy. Impact of rescue therapy use in the analysis is timing-dependent, eg, if a subject starts a rescue therapy between their Week 12 and Week 52 endpoint assessments, then it may have potential impact on Week 52 endpoint analysis but will have no impact on Week 12 endpoints. The rules outlined below apply to both new use and increase in dose from Baseline.

Biologics with immunomodulatory properties

- Rule:
 - Any exposure after first dose
- List of medications:
 - AntiTNF α antibodies:
 - ADALIMUMAB
 - CERTOLIZUMAB
 - CERTOLIZUMAB PEGOL
 - GOLIMUMAB
 - INFLIXIMAB
 - Other Biologics:
 - USTEKINUMAB
 - VEDOLIZUMAB

Nonbiologics with immunomodulatory properties

- Immunosuppressants
 - Rules to consider for Week 12 efficacy endpoints:
 - After first dose and up to and including Week 8: any increase from baseline for more than 5 days
 - After Week 8: Any dose above baseline

-
- Rules to consider for Week 52 efficacy endpoints:
 - After Week 12 and up to and including Week 40: any increase from baseline for more than 5 days
 - After Week 40: Any dose above baseline
 - List of medications:
 - MERCAPTOPYRINE
 - AZATHIOPRINE
 - TIOGUANINE
 - METHOTREXATE
 - METHOTREXATE SODIUM
 - 5-ASA COMPOUNDS
 - Rules to consider for Week 12 efficacy endpoints:
 - After first dose and up to and including Week 8: any increase from baseline for more than 5 days
 - After Week 8: Any dose above baseline
 - Rules to consider for Week 52 efficacy endpoints:
 - After Week 12 and up to and including Week 40: any increase from baseline for more than 5 days
 - After Week 40: Any dose above baseline
 - List of medications:
 - MESALAZINE
 - BALSALAZIDE
 - BALSALAZIDE DISODIUM DIHYDRATE
 - BALSALAZIDE SODIUM
 - OLSALAZINE SODIUM
 - SULFASALAZINE
 - BECLOMETASONE W/MESALAZINE
 - Routes:
 - ORAL
 - RECTAL
 - Other small molecule immunomodulatory active agents
 - Rules to consider for Week 12 efficacy endpoints:

-
- After first dose and up to and including Week 8: any increase from baseline (including new use) for more than 5 days
 - After Week 8: Any dose above baseline
 - Rules to consider for Week 52 efficacy endpoints:
 - After Week 12 and up to and including Week 40: any increase from baseline (including new use) for more than 5 days
 - After Week 40: Any dose above baseline
 - List of medications:
 - CICLOSPORIN
 - TACROLIMUS
 - TOFACITINIB
 - TOFACITINIB CITRATE
 - Systemic glucocorticoids
 - Systemic glucocorticoids given via oral or rectal routes of administration
 - Rules to consider for Week 12 efficacy endpoints:
 - After first dose and up to and including Week 8: any increase from baseline for more than 7 days
 - After Week 8: Any dose above baseline
 - Rules to consider for Week 52 efficacy endpoints:
 - After Week 12 and up to and including Week 40: any increase from baseline for more than 7 days
 - After Week 40: Any dose above baseline
 - List of medications:
 - BETAMETHASONE
 - BETAMETHASONE DIPROPIONATE
 - BETAMETHASONE SODIUM PHOSPHATE
 - DEXAMETHASONE
 - DEXAMETHASONE SODIUM PHOSPHATE
 - DEXAMETHASONE VALERATE
 - METHYLPREDNISOLONE
 - METHYLPREDNISOLONE SODIUM SUCCINATE
 - PREDNISOLONE

-
- PREDNISOLONE SODIUM PHOSPHATE
 - PREDNISOLONE METASULFOBENZOATE SODIUM
 - PREDNISON
 - TRIAMCINOLONE
 - HYDROCORTISONE
 - HYDROCORTISONE ACETATE
 - HYDROCORTISONE BUTYRATE
 - HYDROCORTISONE SODIUM SUCCINATE
 - Routes:
 - ORAL
 - RECTAL
 - Systemic glucocorticoids given via parenteral routes of administration
 - Rule for Week 12 endpoint:
 - Any exposure after first dose up to Week 12
 - Rules for Week 52 endpoint:
 - After Week 12 and up to and including Week 40: more than one dose
 - Any exposure after Week 40
 - List of medications:
 - BETAMETHASONE
 - BETAMETHASONE DIPROPIONATE
 - BETAMETHASONE SODIUM PHOSPHATE
 - DEXAMETHASONE
 - DEXAMETHASONE SODIUM PHOSPHATE
 - DEXAMETHASONE VALERATE
 - METHYLPREDNISOLONE
 - METHYLPREDNISOLONE SODIUM SUCCINATE
 - PREDNISOLONE
 - PREDNISOLONE SODIUM PHOSPHATE
 - PREDNISOLONE METASULFOBENZOATE SODIUM
 - PREDNISON
 - TRIAMCINOLONE

-
- HYDROCORTISONE
 - HYDROCORTISONE ACETATE
 - HYDROCORTISONE BUTYRATE
 - HYDROCORTISONE SODIUM SUCCINATE
 - Routes:
 - INTRAVENOUS
 - INTRAMUSCULAR
 - Topical Glucocorticoids
 - Rules for Week 12 endpoints:
 - After first dose and up to and including Week 8: any exposure above baseline (or new use) for more than 5 days
 - After Week 8: Any dose above baseline
 - Rules for Week 52 endpoints:
 - After Week 12 and up to and including Week 40: any increase from baseline for more than 5 days
 - After Week 40 any increase above baseline
 - List of medications:
 - BUDESONIDE
 - Routes:
 - ORAL
 - RECTAL
 - Beclomethasone
 - Rules for Week 12 endpoints:
 - After first dose and up to and including Week 8: Any increase above baseline for more than 5 days
 - After Week 8: Any dose above baseline
 - Rules for Week 52 endpoints:
 - After Week 12 and up to and including Week 40: Any increase from baseline for more than 5 days
 - After Week 40 any increase above baseline
 - List of medications:
 - BECLOMETASONE

- BECLOMETASONE DIPROPIONATE
- BECLOMETASONE W/MESALAZINE
- Routes:
 - ORAL
 - RECTAL

Medical procedures

- Leukocyte apheresis, other apheresis, and plasma exchange
 - Rule:
 - Any exposure after first dose
 - List of medical procedures
 - APHERESIS
 - LEUKAPHERESIS
 - COLECTOMY (partial or total)
 - SIGMOIDECTOMY
 - COLOSTOMY
 - ILEOSTOMY

Statistical Analysis Plan - V1.0 - SAP or PK-PD Sections Thereof - 19-Jan-2022

Electronic Signature Manifestation

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

Signer Full Name	Meaning of Signature	Date and Time
██████████	Document Approval (I certify that I have the education, training and experience to perform this task)	19 Jan 2022 20:47:12 UTC
██████████	Document Approval (I certify that I have the education, training and experience to perform this task)	19 Jan 2022 20:49:35 UTC
██████████	Document Approval (I certify that I have the education, training and experience to perform this task)	19 Jan 2022 20:50:35 UTC
██████████	Document Approval (I certify that I have the education, training and experience to perform this task)	19 Jan 2022 20:52:31 UTC