Protocol C5041011 (APD334-210)

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

Statistical Analysis Plan (SAP)

Version: 3.0

Date: 23 July 2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. VERSION HISTORY

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
1.0/ 17 April 2024	2.0	N/A	N/A
2.0/ 25 April 2024	2.0	Align the definition of clinical remission for the mildly to moderately active UC population with the FDA Draft Guidance	Section 3.1.2 Second bullet was changed as: ▪ Definition per FDA Draft Guidance: clinical remission is defined as SF subscore = 0 or = 1 and no greater than baseline, RB subscore = 0, and ES ≤ 1 (excluding friability), based on the 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment
		Revise the criteria of potentially important PFT measurements.	Section 6.3.2.5 Potentially important PFT measurements criteria were changed from ">30%" to "≥30%" as: • Percent decrease from Baseline ≥ 30% in FEV1, ie, percent change from Baseline ≤ -30% • Percent decrease from Baseline ≥ 30% in FVC, ie, percent change from Baseline ≤ -30%
		Make the definition of analysis visit windows clearer	Appendix 9.1.4 Table 9 and footnotes were revised.
3.0/ Xx July 2024	2.0	Per FDA feedback of 03July2024 for the SAP	Sections 2.2.3, 2.2.4, 5.3.1.1, discontinuation of the study treatment for any reason will be treated as intercurrent event. Section 3.1 clarified that per FDA Draft Guidance, the SF and RB subscores will be calculated by averaging the most recent 3 consecutive days daily subscores (if not available, by averaging the most recent 4 nonconsecutive days daily subscores) within the 7 days prior to the day of bowel preparation, excluding the day of bowel preparation (for visit without endoscopy, prior to the date of visit). Sections 3.1 and 3.2 clarified that • RB/SF subscore derivation algorithm per protocol is the primary algorithm.

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
			• Clinical remission definition per FDA Draft Guidance will be used for the primary analysis. • Clinical remission based on clinical remission definition per FDA Draft Guidance using derivation of subscores for RB and SF per protocol will be used for the primary analysis. Section 5.1, the sequential testing order was changed. Complete symptomatic remission at Week 52 and complete symptomatic remission at Week 12 were moved to the end of the sequence. The order of key secondary endpoints in Section 2.2.2, Section 3.2.1, and Table 6 was changed as well to consistent with Section 5.1. Section 5.3.1.2, random seeds were added for multiple imputations. Section 6.3.1.5, severe TEAE was added to the comparative analysis list. Difference for EAIR between etrasimod and placebo with 95% CI was added as well. Section 6.4 added subgroup analysis for Sex (Male, Female), Race (White, Black or African American, Asian, Other), Ethnicity (Hispanic or Latino vs Not Hispanic or Latino).
		MedDRA version and WHO Drug dictionary update	Section 3.2.2.1, MedDRA version was updated as v27.0. Section 6.5.4, WHO Drug dictionary was updated as GlobalB3 March 2024.
		Week 20 is not a scheduled visit	Section 3.4.1, week 20 was removed from the list of visits for symptomatic remission and symptomatic response.
		Ukraine war	Sections 5.3.1 and 5.3.1.2.4, Ukraine war impact was added to hybrid imputation.
		Additional subgroup analyses	Section 6.4 added subgroup analysis for prior Oral 5-ASA Failure Only (yes, no), Naïve to biologic or JAK inhibitor therapy who failed corticosteroid and/or thiopurines (yes, no).
		Week 20 is not a scheduled visit	Appendix 9.1.4 Table 9, Week 20 was removed, windows for week 16 and week 24 were adjusted.

Table 1. Summary of Changes

Version/ Date	Associated Protocol Amendment	Rationale	Specific Changes
		Partial date convention was mixed for TEAE and prior and concomitant medication	Appendix 9.1.5 Table 10 was separated to 2 tables: Table 10 is for TEAE, Table 11 is for prior and concomitant medication.

2. INTRODUCTION

This statistical analysis plan (SAP) provides the detailed methodology for summary and statistical analyses of the data collected in Study C5041011 (APD334-210). This SAP is based on the Clinical Study Protocol Amendment 2.0, dated 4 August 2022.

The following analyses will be specified in separate analysis plans.

CCI

- A population PK analysis using plasma concentrations over time will be described in a separate PK analysis plan.

2.1. Modifications to the Analysis Plan Described in the Protocol

Table 2. Summary of Modification

Version/ Date	Associated Protocol Amendment	Rationale	Specific Modification
1.0/ 17 April 2024	Amendment 2.0, dated 4 August 2022	(1) The new definition of clinical remission is per FDA Draft Guidance, Ulcerative Colitis: Developing drugs for treatment (Section 8 References). (2) FDA's recommendation for C5041011 (APD334-210) Protocol Amendment 2.0 ("FDA Letter IND 125154 Advice-Information Request IND", received on 12/02/2022).	(1) Add new definition of clinical remission per FDA Draft Guidance, Ulcerative Colitis: Developing drugs for treatment (Section 8 References). (2) Add the alternative derivation method of rectal bleeding (RB) and stool frequency (SF) subscores per FDA Draft Guidance, Ulcerative Colitis: Developing drugs for treatment.

Table 2. Summary of Modification

Version/ Date	Associated Protocol Amendment	Rationale	Specific Modification
		(3) The change from baseline in RB/SF subscore is included in USPI Velsipity-etrasimod-tablets.	(3) Add analyses for modified Mayo score (MMS) and change from Baseline by visit, RB, SF, and composite RB/SF subscores and change from Baseline by visit.
		(4) To evaluate Cardiac Holter monitoring related assessments based on Holter recording, such as the treatment-emergent AV conduction changes, treatment-emergent bradycardia, recovery of nadir Mean hourly heart rate and short-term cardiac treatment-emergent adverse events.	(4) Add Cardiac Holter endpoints (Section 3.3.2) and corresponding planned analyses for the Cardiac Holter endpoints ("Cardiac Holter Monitoring" part of Section 6.3.2).
		(5) To be consistent with the language of definition for mucosal healing used in Full Prescribing Information of VELSIPITY™ (etrasimod), after VELSIPITY™ (etrasimod) was approved by FDA on October 12, 2023.	(5) Add the definition for mucosal healing, and rename it as Histologic-Endoscopic Mucosal Improvement (listed as key secondary efficacy endpoint at Week 52 and other secondary efficacy endpoint at Week 12) in Section 2.2.2, per Full Prescribing Information of VELSIPITY™ (etrasimod) (Section 8 References).
		(6) 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment.	(6) The primary efficacy analysis population is defined to be consistent with the FDA Draft Guidance for mildly to moderately active UC.
		(7) Efficacy analyses will be based on the estimand framework.	(7) Per protocol analysis set is irrelevant and is removed.
		(8) Per FDA comments.	(8) Endoscopic improvement at Week 12 and complete symptomatic remission at Week 12 and Week 52 are considered as key secondary endpoints and are included in type I error controlled family. Symptomatic remission at Week 52 is now considered as a secondary endpoint.
		(9) Add additional exploratory endpoints.	(9) Clinically meaningful improvement in bowel urgency at Week 12 and Week 52;

Table 2. Summary of Modification

Version/ Date	Associated Protocol Amendment	Rationale	Specific Modification
			bowel urgency remission at Week 12 and Week 52; complete bowel urgency remission at Week 12 and Week 52 are added as exploratory endpoints.
		(10) Add two secondary endpoints.	(10) 12-week corticosteroid-free clinical remission at Week 52 among participants receiving corticosteroids at baseline; 4-week corticosteroid-free clinical remission at Week 52.

2.2. Study Objectives, Endpoints, and Estimands

2.2.1. Study Objectives

- Primary objective: to assess the efficacy of etrasimod on clinical remission in participants with mildly to moderately active UC after 52 weeks of treatment.
- Secondary objective: to assess the efficacy of etrasimod on endoscopic improvement, histologic response, clinical response, and symptomatic remission, at timepoints up to 52 weeks of treatment.
- Safety objective: to assess the long-term safety of etrasimod after daily doses of 2 mg for up to 52 weeks.
- Other objectives: to evaluate etrasimod pharmacokinetics (PK) and the effect of etrasimod on health-related participant-reported outcomes and biomarkers.

2.2.2. Study Endpoints

This section lists the study endpoints. Their definitions are given in section 3.

The primary efficacy endpoint is:

- Clinical remission at Week 52.

The key secondary efficacy endpoints are:

- Endoscopic improvement at Week 52.
- 12-week Corticosteroid-free clinical remission at Week 52.
- Clinical remission at Week 12
- Endoscopic improvement at Week 12.
- Sustained clinical remission (clinical remission at both Weeks 12 and 52).

- Histologic-Endoscopic Mucosal Improvement at Week 52.
- Complete symptomatic remission at Week 52.
- Complete symptomatic remission at Week 12.

The other secondary efficacy endpoints and exploratory efficacy endpoints to be assessed in this study are listed in Section 3.2.1 and Section 3.4.1, respectively.

The safety endpoints are:

- Incidence and severity of adverse events
- Incidence and severity of laboratory abnormalities, and change from baseline in laboratory values (to include hematology, serum chemistry, coagulation, and urinalysis)
- Incidence of clinically significant vital sign abnormalities and changes from baseline

The Cardiac Holter endpoints to be assessed in this study are described in Section 3.3.2. The Pharmacokinetic (PK) endpoints to be assessed in this study are described in Section 3.2.3. The Health-related Quality of Life (HRQoL) endpoints to be assessed in this study are described in Section 3.4.2. The efficacy-related biomarker endpoints to be assessed in this study are described in Section 3.4.3.

2.2.3. Primary Estimand (Estimand 1)

The primary estimand of this study is a composite estimand, which estimates the effect of randomized treatment accounting for treatment adherence and response.

- Treatment: Etrasimod 2 mg vs Placebo
- Population: Participants with mildly to moderately active UC as defined in 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment (Section 8 References)
- Variable: primary efficacy endpoint of clinical remission at week 52 and all binary efficacy endpoints as described in Sections 3.1, 3.2.1 and 3.4.1
- Intercurrent events: the following intercurrent events are defined:
  1) Discontinue the study treatment during the double blind treatment period for any reasons
  2) Initiate a rescue medication for UC (Section 9.3)
  3) Have an increase in dose over baseline levels in their existing UC medication following the rules in Section 9.3

4) Have a rescue non-drug treatment (eg, colectomy, ileostomy, or sigmoidectomy) during the study (Section 9.3)

Participants having any of the intercurrent events as defined in 1) - 4) above will be considered as nonresponders, that is, have a known outcome of nonresponse at applicable timepoints. Missing measurements due to other reasons will be treated as missing. The method of dealing with missing data is presented in Section 5.3

- Population-level summary: Proportion of participants who are responders in each treatment group and differences in proportions of responders between Etrasimod 2 mg group and placebo group.

2.2.4. Secondary Estimand (Estimand 2)

The secondary estimand of this study is a hypothetical estimand, which estimates the effect of randomized treatment as if all participants maintain their randomized treatment.

- Treatment: Etrasimod 2 mg vs Placebo

- Population: Participants with mildly to moderately active Ulcerative Colitis (UC) as described in 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment.

- Variable: change from baseline in MMS and other continuous variables for exploratory efficacy endpoints, HRQoL endpoints and efficacy-related biomarker endpoints as described in Sections 3.4.1, 3.4.2 and 3.4.3, respectively.

- Intercurrent events: The following intercurrent events are defined:

  1) Discontinue the study treatment during the double blind treatment period for any reasons

  2) Initiate a rescue medication for UC (Section 9.3)

  3) Have an increase in dose over baseline levels in their existing UC medication following the rules in Section 9.3.

  4) Have a rescue non-drug treatment (eg, colectomy, ileostomy, or sigmoidectomy) during the study (Section 9.3)

  For participants having any of the intercurrent events as defined in 1) - 4) above, their data collected up to the intercurrent events will be used for analysis.

- Population-level summary: Least-squares mean (LSM) (Section 5.2.2) in each treatment group and differences in LSMs between Etrasimod 2 mg group and placebo group (Section 5.2.2).

2.3. Study Design

This is a multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of etrasimod 2 mg in participants with mildly to moderately active UC. The study consists of a 28-Day Screening Period, a 52 week treat-through design consisting of a 12-Week Treatment Period and a 40-Week Treatment Period, and a 4-Week Follow-Up Period (with 2 visits) (Figure 1. Study Design).

The target participant population will include:

1. Participants who have had an inadequate response to, loss of response to, or intolerance to conventional therapy and are naïve to biologic or Janus kinase (JAK) inhibitor therapy.

2. Participants who have had an inadequate response to, loss of response to, or intolerance to a biologic or JAK inhibitor (participants in this category may have received prior conventional therapy).

Participant eligibility will be determined during a Screening Period of 4 weeks (28 days). Entry criteria will be based on confirmation of mildly to moderately active UC. In order to align with the definition of mildly to moderately active UC based on the 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment, the disease-specific inclusion criteria in Protocol C5041011 (APD334-210) amendment 2.0 is updated as follows (Table 3):

Table 3. Disease Specific Inclusion Criteria

Protocol Version	Disease Specific Inclusion Criteria
Prior to amendment 2.0	Inclusion criteria 5: MMS of 4 to 6 and an ES $\geq$ 1.
Amendment 2.0	Inclusion criteria 5: MMS of 4 to 6 and an ES $\geq$ 2 and RB score $\geq$ 1.

Eligible participants 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 (12-Week Treatment Period + 40-Week Treatment Period). Due to the disease specific inclusion criteria update as described in Table 3 above, randomization stratification in Protocol C5041011 (APD334-210) amendment 2.0 is updated as follows (Table 4).

Table 4. Randomization Stratification

Protocol Version	Randomization Stratification
Prior to amendment 2.0	(a) naïve to biologic or JAK inhibitor therapy at study entry (yes or no) (b) baseline corticosteroid use (yes or no) and (c) baseline ES>1 (yes or no)
Amendment 2.0	(a) naïve to biologic or JAK inhibitor therapy at study entry (yes or no) and (b) baseline corticosteroid use (yes or no)

End of 12-Week Double-Blind Treatment Period

At the end of the 12-Week Treatment Period, participants will undergo the Week 12 efficacy and safety assessments and procedures. Participants whose UC condition in the opinion of

the Investigator is stable or improving compared with baseline (Week 0/Day 1) will continue with their double-blind treatment and move into the 40-Week Treatment Period.

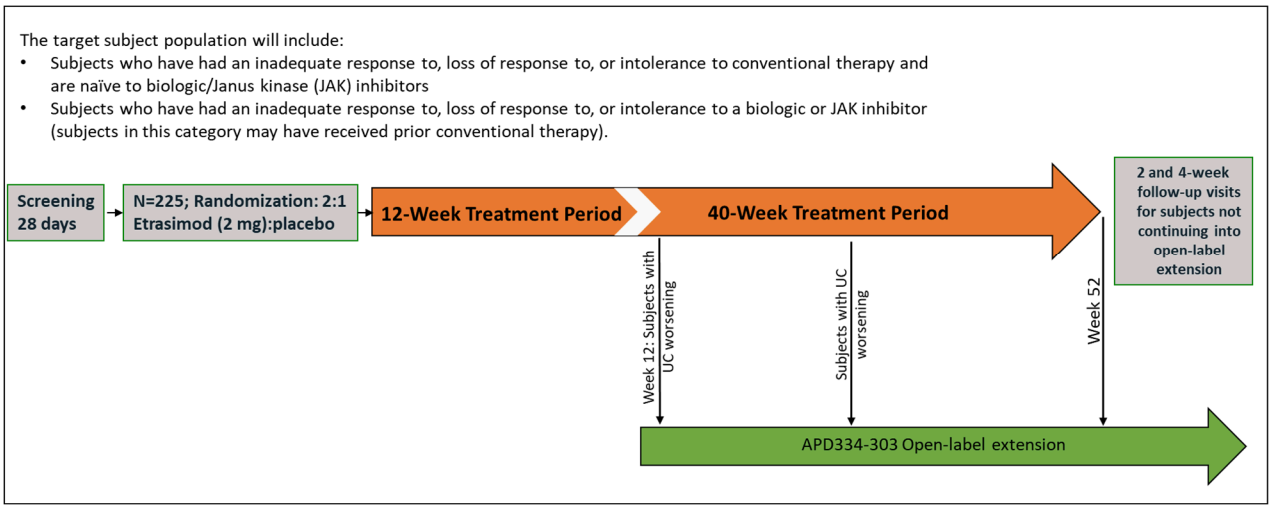
End of 40-Week Double-Blind Treatment Period (Week 52)

At the end of the 40-Week Treatment Period (ie, Week 52) and following completion of all study procedures, participants will have the option to enter into the Open-Label Extension (OLE) study (Study APD334-303) provided they meet all eligibility criteria.
Participants who do not participate in the OLE study will have 2-Week and 4-Week Follow-Up visits after their last treatment administration.

Open-Label Extension Study [Study C5041012 (APD334-303)]

Participants whose UC condition in the opinion of the Investigator has not improved or has worsened, compared with baseline (Week 0/Day 1), may be eligible to enroll in the OLE study. Participants must complete Week 12 to be considered for the OLE study.
Disease worsening will continue to be monitored by Investigators through the 40-Week Treatment Period (ie, from Weeks 13 to 52). Participants 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.

Figure 1. Study Design



Based on a 2-group Fisher's exact test, a two-sided significance level of 0.05, and a 2:1 randomization ratio, approximately 183 subjects (122 etrasimod, 61 placebo) are required to achieve at least 80% power to detect a difference of 20% in the primary endpoint of clinical remission (based on MMS) at Week 52 between the etrasimod treatment group (34%) and the placebo treatment group (14%). Alternatively, based on the same assumptions, using normal approximation to the binomial distribution, a two-sided test would achieve at least 90% power. The assumptions are based on the final analysis results in the Phase 3 Studies APD334-301 and APD334-302. Sample size is calculated using the EAST® software.

Based on the updated inclusion criteria in Protocol C5041011 (APD334-210) amendment 2.0, approximately 45 participants among the already randomized participants (assuming

approximately 2:1 ratio of etrasimod:placebo) did not meet the amended Inclusion Criterion 5, which includes ES score ≥ 2 and RB score ≥ 1. As a result, these participants will be excluded from the Primary Analysis Set (Section 4). Therefore, approximately 225 total participants (150 etrasimod, 75 placebo) were required to be enrolled into study C5041011 (APD334-210) under Protocol amendment 2.0.

3. ENDPOINTS AND BASELINE VARIABLES: DEFINITIONS AND CONVENTIONS

3.1. Primary Endpoint

3.1.1. Derivation of subscores for RB, SF and MMS

- Derivation 1 for RB and SF subscores (per protocol):
  - For RB and SF used to qualify for randomization. SF and RB subscores will be derived from the participant eDiary entries using the scores from the 3 most recent consecutive days within the 7 days prior to the day of bowel preparation, (excluding the day of bowel preparation), averaged and rounded to the nearest integer. Participants who do not have 3-consecutive days of eDiary data within that 7-day period and who do not have a minimum of 7 days of eDiary data prior to bowel preparation are not eligible for randomization.

  - For RB and SF in treatment period. SF and RF subscores will be derived using the scores from the 3 most recent consecutive days within the 7 days prior to the day of bowel preparation (excluding the day of bowel preparation), averaged and rounded to the nearest integer. If 3 consecutive days are not available, then the average of the 2 most recent consecutive days within the 7 days will be used. On visits without endoscopy, the SF and RB are derived using the scores from the 3 most recent consecutive days within the 7 days prior to the date of visit, averaged and rounded to the nearest integer. If 3 consecutive days are not available, then the average of the 2 most recent consecutive days within the 7 days will be used.

  The above RB and SF subscore derivation algorithm per protocol will be the primary algorithm used for the primary analysis.

- Derivation 2 for RB and SF subscores (per 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment)

  - The SF and RB subscores will be calculated from the participant eDiary entries using the daily scores from the most recent 3 consecutive days or if not available, from the most recent 4 non-consecutive days within the 7 days prior to the day of bowel preparation, excluding the day of bowel preparation (for visit without endoscopy, prior to the date of visit), averaged and then rounded to the nearest integer. A minimum of 3 consecutive days of completed diary entries or 4 nonconsecutive days within 7-day period are necessary. Otherwise, the subscore for the visit will be missing.

The MMS includes the ES, RB, and SF components of the Mayo Clinic Score (MCS; and related composite scores). The total score range of the MMS is from 0 to 9, with each

component ranging from 0 to 3 (0 = normal, 1 = mild, 2 = moderate, 3 = severe). The MMS will be derived based on each of derivations for RB and SF above, respectively as follows:

- MMS Derivation 1 (per protocol), based on derivation 1 for RB and SF
- MMS Derivation 2 (per 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment), based on derivation 2 for RB and SF

3.1.2. Definition of Clinical Remission

- Definition per protocol: clinical remission is defined as SF subscore = 0 (or = 1 with a ≥ 1-point decrease from Baseline), RB subscore = 0, and endoscopic score (ES) ≤ 1 (excluding friability)
- Definition per FDA Draft Guidance: clinical remission is defined as SF subscore = 0 or = 1 and no greater than baseline, RB subscore = 0, and ES ≤ 1 (excluding friability), based on the 2022 FDA Draft Guidance; Ulcerative Colitis: Developing Drugs for Treatment.

The definition of clinical remission per FDA Draft Guidance will be used for the primary analysis.

3.1.3. Derivation of Clinical Remission

The primary efficacy endpoint is clinical remission at Week 52. Based on above different derivation algorithms for RB/SF subscores, and different definitions for clinical remission, the clinical remission will be derived based on the following 3 approaches, respectively:

- Clinical remission 1, based on clinical remission definition per protocol and RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
- Clinical remission 2, based on clinical remission definition per FDA Draft Guidance and RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
- Clinical remission 3, based on clinical remission definition per FDA Draft Guidance and RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

Clinical remission 2 will be used for the primary analysis.

3.2. Secondary Endpoint(s)

3.2.1. Efficacy Endpoint(s)

The study includes the following key secondary efficacy endpoints:

- Endoscopic improvement at Week 52. Endoscopic improvement is defined as ES ≤ 1 (excluding friability).
- 12-week corticosteroid-free clinical remission at Week 52: clinical remission at Week 52 and corticosteroid-free for ≥ 12 weeks immediately prior to Week 52. The 3 clinical remissions defined in Section 3.1.3 will be applied, respectively.

- Clinical remission at Week 12. The 3 clinical remissions described in Section 3.1.3 above will be applied.

- Endoscopic improvement at Week 12.

- Sustained clinical remission (clinical remission at both Weeks 12 and 52). The 3 clinical remission definitions defined in Section 3.1.2 will be applied, respectively.

- Histologic-Endoscopic Mucosal Improvement at Week 52. Histologic-Endoscopic Mucosal Improvement is defined as ES ≤ 1 (excluding friability) with Geboes Index score < 2.0.

- Complete symptomatic remission at Week 52. Complete symptomatic remission is defined as RB subscore = 0 and SF subscore = 0. Its derivation will be based on the following 2 approaches, respectively.

    - Complete symptomatic remission 1, based on RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
    - Complete symptomatic remission 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Complete symptomatic remission at Week 12. Two complete symptomatic remission definitions described above will be applied, respectively.

The study also includes the following other secondary efficacy endpoints.

- 12-week corticosteroid-free clinical remission at Week 52 among participants receiving corticosteroids at baseline: clinical remission at Week 52 and corticosteroid-free for ≥ 12 weeks immediately prior to Week 52. The 3 clinical remissions defined in Section 3.1.3 will be applied, respectively.

- 4-week corticosteroid-free clinical remission at Week 52: clinical remission at Week 52 and corticosteroid-free for ≥ 4 weeks immediately prior to Week 52. The 3 clinical remissions defined in Section 3.1.3 will be applied, respectively.

- 4-week corticosteroid-free clinical remission at Week 52 among participants receiving corticosteroids at baseline: clinical remission at Week 52 and corticosteroid-free for ≥ 4 weeks immediately prior to Week 52. The 3 clinical remissions defined in Section 3.1.3 will be applied, respectively.

- Clinical response at Week 12. Clinical response is defined as a ≥ 2-point and ≥ 30% decrease from baseline in MMS, and a ≥ 1-point decrease from baseline in RB subscore or an absolute RB subscore ≤ 1. Its derivation will be based on each of 2 derivations of RB and MMS in Section 3.1.1, respectively as follows:

    - Clinical response 1, based on RB and MMS derivation per protocol (Derivation 1, Section 3.1.1)

  - Clinical response 2, based on RB and MMS derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Clinical response at Week 52. Two clinical responses described above will be applied.

- Histologic-Endoscopic Mucosal Improvement at Week 12.

- Symptomatic remission at Week 52. Symptomatic remission is defined as SF subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline) and RB subscore = 0. Its derivation will be based on the following 2 approaches, respectively.

  - Symptomatic remission 1, based on RB and SF per protocol (Derivation 1, Section 3.1.1)

  - Symptomatic remission 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Symptomatic remission at Week 12. Two symptomatic remissions described above for symptomatic remission at Week 52 will be applied, respectively.

- Reduction from baseline in both ES and RB or in both ES and SF at Week 12. Derivations 1 and 2 for RB and SF described in Section 3.1.1 will be applied, respectively.

- Histologic response based on the Geboes Grading System at Week 12. Histologic response based on the Geboes Grading System is defined as Geboes Index score ≤ 3.1.

- Histologic response based on Robarts Histopathology Index (RHI) at Week 12. Histologic response based on RHI is defined as decrease in RHI of ≥7 points from baseline.

3.2.2. Safety Endpoint(s)

3.2.2.1. Adverse Event

AEs will be coded using MedDRA (v27.0). 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 Section 9.1.5 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.

Severity of TEAEs

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).

Relationship of TEAEs to Study Treatment

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

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 electronic case report form (eCRF).

Serious and Non-Serious TEAEs

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

TEAEs Leading to Death

TEAEs leading to Death are those events which are recorded with an outcome as "Fatal" on the Adverse Events eCRF.

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. The proposed candidate terms will be reviewed along with other supportive data (eg, vital signs, lab) to identify which events reflect AESI.
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

3.2.2.2. Laboratory Data

Detailed descriptions of the laboratory data and other tests that are collected per protocol can be found in Table 5 ("Clinical Laboratory Tests") of Protocol C5041011 (APD334-210) amendment 2.0.

3.2.3. PK Endpoints

Plasma concentrations of etrasimod will be assessed from samples collected pre-dose and 4 hours post-dose (after 12-lead ECG) on Week 0/Day 1, and pre-dose (trough) at Weeks 2, 4, 8, 12, 16, 24, 32, 48, 52, and ET (only required if ET visit ≤ 30 hours since last treatment administration), and at the 2-Week and 4-Week Follow-Up visits (for participants not enrolled into any extension study). A PK sample should also be drawn, if possible, at the time of any SAE or adverse event leading to study treatment discontinuation.

Plasma samples may also be used for profiling of drug binding proteins, bioanalytical method validation purposes, stability assessments, or to assess other actions of etrasimod with plasma constituents.

3.3. Other Safety Endpoint(s)

3.3.1. 12-lead ECGs

ECGs are recorded on a 12-lead ECG machine and read locally and 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)
- AV conduction abnormalities
  - First-degree AV Block
  - First-degree AV Block with PR>230ms
  - Second-degree AV Block Type 1
  - Second-degree AV Block Type 2
  - Third-degree AV Block or Higher

3.3.2. Cardiac Holter Endpoints

3.3.2.1. Binary Endpoints

- Incidence of mean hourly heart rate based on Holter recording (hHR) recovery.
  - Note: hHR was generated by the cardiac central laboratory for each 60-minute interval based on date/time of first dose and Holter recording date/time, with hour 1 starting at the time of dosing. See Section 9.1.7 for additional detail on how partial hourly interval is handled by the cardiac central laboratory and how hourly timepoint will be programmatically derived. For the analysis purpose, if hHR was associated with a time interval of <10 minutes, then its value will be set to missing for the summary. Per cardiac central laboratory, if the Holter recording was started < 60 seconds prior to dosing, then no pre-dose record is derived for the participant. In such case, all the Holter recording will contribute to post-dose calculation only.
- Incidence of low heart rate over the entire Holter monitoring period and at each post-dose hourly timepoint defined by:
  - hHR <40
  - hHR <50
  - hHR <50 bpm plus > 10 bpm decrease from baseline
- Incidence of first nadir hHR reached at each post-dose hourly timepoint

- Incidence of additional abnormal findings flagged by cardiac central laboratory
  - Mobitz I 2nd Degree AV Block
  - Mobitz II 2nd Degree AV Block
  - 2:1 AV Block
  - High Grade AV Block
  - Complete Heart Block
  - Varying Conduction Ratio AV Block
  - Evidence of new onset of transient bundle branch block (BBB) or Intraventricular conduction delay
  - Other specified (any abnormality detected other than the list above must be specified under "Other" and alert notified at discretion of reading Cardiologist)

3.3.2.2. Continuous Endpoints

- hHR and change from baseline by post-dose hourly timepoint
- Time to first nadir hHR (based on Holter hourly data)

3.3.3. Vital signs

The following vital signs measurements are reported for this study:

- Systolic blood pressure (SBP) (mm Hg)
- Diastolic blood pressure (DBP) (mm Hg)
- Heart rate (bpm)
- Respiratory rate (resp/min)
- Temperature (°C), where Temperature (°C) = (5/9) (Temperature (°F) – 32)

3.3.4. Pulmonary Function Tests

The following pulmonary function tests (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)

3.3.5. Ophthalmoscopy and Optical Coherence Tomography (OCT)

Ophthalmoscopy and OCT assessment will be performed according to the Schedule of Assessments (Table 6 in Protocol C5041011 (APD334-210) amendment 2.0 Appendix 1). The endpoint is the categorical result in ophthalmoscopy with OCT parameters by visit (the categories are listed on the eCRF).

3.4. Other Endpoints

3.4.1. Exploratory Efficacy Endpoints

- Symptomatic remission at each study visit (Weeks 2, 4, 8, 12, 16, 24, 32, 40, 48, and 52). Two symptomatic remission derivations described in Section 3.2.1 will be applied, respectively.
- Symptomatic response at each study visit (Weeks 2, 4, 8, 12, 16, 24, 32, 40, 48, and 52). Symptomatic response is defined as ≥ 30% decrease from baseline in composite RB and SF subscores. Its derivation will be based on the following 2 approaches, respectively.
    - Symptomatic response 1, based on RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
    - Symptomatic response 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)
- Clinical remission at Week 52 among participants achieving clinical response at Week 12. Three derivations for clinical remissions described in Section 3.1.3 will be applied, respectively.
- Reduction from baseline in both ES and RB or in both ES and SF at Week 52. Derivation 1 and 2 for RB and FS in Section 3.1.1 will be applied, respectively.
- Clinical response based on total Mayo score (TMS) at Week 52. The TMS includes all four components (ES, RB, SF and PGA) of the MCS. The total score range of the TMS is from 0 to 12 with each component ranging from 0 to 3 (0 = normal, 1 = mild, 2 = moderate, 3 = severe) and higher scores indicating more severe disease. Clinical response based on TMS is defined as a ≥ 3-point and ≥ 30% decrease from baseline in TMS, and a ≥ 1-point decrease from baseline in RB subscore or an absolute RB subscore ≤ 1. The TMS will be derived based on each of derivations for RB and SF (as defined in Section 3.1.1), respectively.
    - TMS clinical response 1, based on RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
    - TMS clinical response 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Histologic response based on Geboes Grading System at Week 52.
- Histologic remission based on the Geboes Grading System at Week 12. Histologic remission based on the Geboes Grading System is define as Geboes Score (GS) ≤ 2.0
- Histologic remission based on the Geboes Grading System at Week 52
- Histologic response based on the RHI at Week 52.
- Histologic remission based on the RHI at Week 12. Histologic remission based on RHI is defined as a RHI score ≤ 3 with lamina propria neutrophils subscore = 0 and neutrophils in epithelium subscore = 0
- Histologic remission based on the RHI at Week 52
- Clinical remission based on the Ulcerative Colitis 100 (UC-100) index at Week 12. UC-100 index is based on SF subscore, ES, and RHI score. These 3 components are incorporated into a calculable final index. The formula for the composite UC-100 score is: (1 + 16 × SF subscore [0 to 3] + 6 × ES [0 to 3] + 1 × RHI score [0 to 33]).

Clinical remission based on the UC-100 index is defined as UC-100 index ≤ 25. Its derivation will be based on the following 2 approaches, respectively.

  - UC-100 clinical remission 1, based on RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
  - UC-100 clinical remission 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Clinical remission based on the UC-100 index at Week 52. Two UC-100 clinical remissions described above will be applied, respectively.
- Clinical remission based on the TMS at Week 12. Clinical remission based on TMS is defined as a TMS ≤ 2 with all subscores ≤1. Its derivation will be based on the following 2 approaches, respectively.

  - TMS clinical remission 1, based on RB and SF derivation per protocol (Derivation 1, Section 3.1.1)
  - TMS clinical remission 2, based on RB and SF derivation per FDA Draft Guidance (Derivation 2, Section 3.1.1)

- Clinical remission based on the TMS at Week 52. Two TMS clinical remissions described above will be applied, respectively.
- MMS and Change from Baseline by visit. Derivations 1 and 2 for MMS in Section 3.1.1 will be applied, respectively.
- RB, SF, and composite RB/SF subscores and change from Baseline by visit. Derivations 1 and 2 for RB and SF in Section 3.1.1 will be applied.

- Clinically meaningful improvement in bowel urgency at Week 12 and Week 52: Decrease from baseline in Urgency NRS ≥ 3 points.
- Bowel urgency remission at Week 12 and Week 52: Urgency NRS ≤ 1 point.
- Complete bowel urgency remission at Week 12 and Week 52: Urgency NRS = 0 point.

3.4.2. HRQoL Endpoints

- Scores and change from baseline at Weeks 12, 32 (UC-PRO only), and 52 in the following:
  - Two Ulcerative Colitis Patient-Reported Outcomes (UC-PRO) modules. UC-PRO scoring is derived by calculating the average using at least 4 out of 7 daily scores, using the visit date to determine the days to be used. UC-PRO is available on the participants eDiary 14 days prior to and 14 days after the projected visit date. Participant should complete UC-PRO when it becomes available on the device for the visit, continuing until the visit occurs.
    - UC-PRO Signs and Symptoms (UC-PRO/SS). It is a 9-item questionnaire containing 2 domains: Bowel movement signs and symptoms (6 items) and functional symptoms (3 items). An average score is calculated for each domain; a higher score indicates worse symptoms.
    - UC-PRO Systemic Symptoms. The Systemic Symptoms measure includes 5 items, scored as a single scale, to address the presence and severity of systemic symptoms including pain, feeling tired, lack of appetite, feeling weak, and thirst.
  - Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F). It is a 13-item fatigue instrument utilized to assess fatigue/ tiredness and its impact on daily activities and functioning in a number of chronic diseases. The instrument includes items such as tiredness, weakness, listlessness, lack of energy, and the impact of these feelings on daily functioning (eg, sleeping, and social activities).
  - Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis (WPAI-UC). It consists of 6 questions asking about the effect of UC on the participant'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 derivation can be found in Section 9.2.1.
  - Urgency numeric rating scale (NRS). It 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).

- Inflammatory Bowel Disease Questionnaire total score (IBDQ). It 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. Details about scoring rules for IBDQ total score and 4 domain subscores can be found in Section 9.2.2.

- Medical Outcomes Study 36-Item Short Form Health Survey, version 2 physical and mental component and domain scores (SF-36). The 36-Item Short Form Health Survey (SF-36) is a 36-item, participant-reported survey of participant 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 participant'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. Details about scoring rules and derivations of domain scores can be found in Section 9.2.3

- Patient Global Impression of Change (PGIC). It is the PRO counterpart to the Clinical Global Impressions (CGI) scale, which was published in 1976 by the National Institute of Mental Health (US). It consists of one item taken from the CGI and adapted to the patient.

- UC-related hospitalizations.

- UC-related surgeries, including colectomy.

3.4.3. Efficacy-Related Biomarkers Endpoints

- Change from baseline in level of fecal calprotectin at Weeks 2, 4, 8, 12, 24, and 52.

- Change from baseline in level of high-sensitivity C-reactive protein (hs-CRP) at Weeks 2, 4, 8, 12, 16, 24, 32, 48, and 52.

- Change and percentage change from baseline in lymphocyte counts at Weeks 2, 4, 8, 12, 16, 24, 32, 48, and 52.

3.5. Baseline Variables

Unless otherwise specified, baseline is defined as the last non-missing measurement taken prior to the first dose (including unscheduled assessments). If measurements include time

(except for health-related quality of life instruments), the date/time will be used to define Baseline. Otherwise, only dates will be compared. For health-related quality of life 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.

Regarding cardiac analyses, for the analysis of hHR endpoints in Section 3.3.2, the last pre-dose hHR as derived by the cardiac central laboratory from Holter recording will be used as the baseline measurement (which may be based on recording from a partial hourly interval; see derivation of hourly timepoint in Section 9.1.7), except in the analysis of time to nadir hHR by baseline ECG HR, where baseline mean heart rate from pre-dose 12-lead ECG will be used.

The following demographic data and Baseline characteristics will be summarized:

- Age on consent (years)
- Sex
- Race
- Ethnicity
- Woman of childbearing potential (Yes or No)
- Height (cm)
- Weight (kg)
- Body mass index (BMI) (kg/m^2)
- Alcohol consumption (Yes or No)
- Caffeine consumption (Yes or No)
- Tobacco use (Yes or No)
- Region (North America, Western Europe, Easter Europe, Other)

The following Baseline characteristics related to UC will be summarized:

- Baseline MMS
- Baseline RB subscore
- Baseline SF subscore
- Baseline ES

- Baseline Physician's Global Assessment (PGA)
- Baseline TMS
- Duration of UC (years)
- Any acute exacerbations within past 12 months (Yes, No, Ongoing), 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
- 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)
- 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
- 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)

Additional baseline characteristics related to UC can be added, if deemed needed.

Prior treatment
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.

Medical history
Medical history will be collected on the medical history eCRF and coded using Medical Dictionary for Regulatory Activities (MedDRA, v26.1). The version used to code medical history will be displayed in the outputs.

4. ANALYSIS SETS (POPULATIONS FOR ANALYSIS)

Data for all participants will be assessed to determine if participants meet the criteria for inclusion in each analysis set prior to unblinding the database and classifications will be documented per standard operating procedures. The description of each analysis set is provided in Table 5 below.

Table 5. Summary of Analysis Sets

Analysis Set	Description
Screened Set	The Screened Set will consist of all participants who sign informed consent to participate in the study.
Randomized Set	The Randomized Set will consist of all participants who are randomized to study treatment.
Full Analysis Set (FAS)	The FAS will consist of all randomized participants, who receive at least 1 dose of study treatment. Under this approach, participants will be summarized by the treatment group to which they were randomized, regardless of the treatment actually received during the study.
Primary Analysis Set (PAS):	The Primary Analysis Set will consist of participants in FAS with baseline MMS 4-6, baseline ES ≥ 2 and baseline RB score ≥ 1. This analysis set will be the primary analysis set for analyses of all efficacy endpoints.
Modified Primary Analysis Set (mPAS)	The mPAS will consist of all participants in PAS, who receive at least 1 dose of study treatment and have a baseline and at least 1 post-randomization measurement. Under this approach, participants will be summarized by treatment to which they were randomized, regardless of the treatment received during the study. Note that the mPAS can vary with endpoints since some participants may have the needed data for inclusion in the mPAS for some endpoints but others may not.
Safety Set	The Safety Set will include all randomized participants who receive at least 1 dose of study treatment. For this set, participants will be analyzed according to the treatment received, regardless of randomization. The Safety Set will be used for all safety analyses
Modified Safety Set	The Modified Safety Set will include all participants who receive at least 1 dose of study treatment and have Holter recording data on Day 1. Unless stated otherwise, the analyses of Cardiac Holter endpoints (Section 3.3.2) will be performed for the Modified Safety Set. For this set, participants will be analyzed according to the treatment received, regardless of randomization.
Pharmacokinetic Set	The Pharmacokinetic Set will include all participants 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.

5. GENERAL METHODOLOGY AND CONVENTIONS

5.1. Hypotheses and Decision Rules

There are multiple null hypotheses for the comparison of etrasimod 2 mg versus placebo for the following primary endpoint and key secondary efficacy endpoints. The family-wise type-

I error rate will be controlled at the level of 0.05 using the fixed sequential testing approach in the order specified below.

(1) Clinical remission at Week 52 (Primary efficacy endpoint)

(2) Endoscopic improvement at Week 52

(3) 12-week corticosteroid-free clinical remission at Week 52

(4) Clinical remission at Week 12

(5) Endoscopic improvement at Week 12

(6) Sustained clinical remission (clinical remission at both Week 12 and Week 52)

(7) Histologic-Endoscopic Mucosal Improvement at Week 52

(8) Complete symptomatic remission at Week 52

(9) Complete symptomatic remission at Week 12

Hypotheses for all other efficacy endpoints not described above are to be tested at the nominal 5% (2-sided) significance level, without adjusting for multiple comparisons.

5.2. General Methods

In general, for descriptive analyses, number and percent will be presented for binary variables. Number, mean, standard deviation, median, minimum, maximum, first and third quartiles will be presented for continuous variables.

The general considerations for data derivation details can be found in Section 9.1 Data Derivation Details of Section 9. Appendix, such as derivations for some demographics and baseline characteristics related variables (Section 9.1.1), reference start date and study day (Section 9.1.2), retests, unscheduled visits, and early termination data (Section 9.1.3), definition and use of visit windows in reporting (Section 9.1.4), etc.

5.2.1. Analyses for Binary Efficacy Endpoints

The primary analysis of all binary efficacy endpoints will be performed based on the Mantel–Haenszel (MH) weighted method, stratified by (a) naïve to biologic or JAK inhibitor therapy at study entry (yes or no) and (b) baseline corticosteroid use (yes or no). Results will be presented as the number of participants with response, proportion of participants with response, estimated common risk difference and associated two-sided 95% confidence intervals (CIs), and p-values from the MH weighted method.
For (a) and (b) mentioned above that are used in MH weighted method, the reported strata (which are based on randomization system) will be used for the statistical analyses. Analyses may be repeated using actual strata (which are based on medications reported in eCRF) if the strata data recorded in the randomization system and the actual strata data recorded in the

CRF are inconsistent for more than 10% of the randomized participants. i.e., if there are more than 10% of randomized participants whose reported strata are not the same as actual strata.

The methods for handling with missing data are described in Section 5.3 below.

5.2.2. Analyses for Continuous Non-Cardiac Endpoints

All continuous exploratory efficacy endpoints, HRQoL endpoints and efficacy-related biomarker endpoints measured longitudinally will be summarized by visit using descriptive statistics (n, mean, standard deviation [SD], median, minimum, and maximum), and presented in a listing. They will also be analyzed using a linear mixed-effect model, with factors for naïve to biologic or JAK inhibitor therapy at study entry (yes or no), baseline corticosteroid use (yes or no), treatment, visit, treatment by visit interaction and baseline value. Reported strata will be used in the model. Unstructured covariance will be used. If there is a convergence issue in fitting the unstructured covariance, other covariance structures, such as a compound symmetry covariance structure will be considered as appropriate. LSM for each treatment group, standard errors (SEs), LSMs in treatment difference and associated 95% CIs, and p-values by visit will be reported.

Analyses above based on linear mixed-effect model may be repeated using actual strata if mis-classification occurs in more than 10% of randomized participants. i.e., if there are more than 10% of randomized participants whose reported strata are not the same as actual strata.

5.2.3. Analyses for Cardiac Endpoints

For binary cardiac endpoints, data will be summarized using descriptive statistics, including number of participants (n), percentage, and 2-sided 95% Clopper-Pearson CI. No missing data will be imputed.
For continuous cardiac endpoints, data will be summarized using descriptive statistics, including number of participants (n), mean, median, SD, 2-sided 95% CI, first quartile (Q1), third quartile (Q3), minimum, and maximum by timepoint. No missing data will be imputed.

For count cardiac endpoints, data will be summarized using the same descriptive statistics as for continuous endpoints. No missing data will be imputed.

5.3. Methods to Manage Missing Data

5.3.1. Binary Efficacy Endpoints

Participants with a missing binary efficacy outcome will be included in the analyses using the following method:

- Primary method: participants with missing data will be treated as nonresponders.
- Other methods for sensitivity analyses:
    1) Multiple imputation under missing at random (MAR) assumption.

2) Tipping point analysis.

3) multiple imputation with Copy Reference (CR) under missing not at random (MNAR) assumption.

4) Hybrid imputation with multiple imputation to handle missing endoscopy data due to the COVID-19 pandemic impact and Ukraine war impact, nonresponder imputation (NRI) for missing data not due to the COVID-19 pandemic impact and Ukraine war impact.

More details are provided in Sections 5.3.1.1 and 5.3.1.2 below.

5.3.1.1. Primary Missing Data Handling Method

For binary efficacy endpoints, participants having any of intercurrent events 1) discontinue the study treatment during the double blind treatment period for any reasons, or 2) initiate a rescue medication for UC, or 3) have an increase in dose over Baseline levels in their existing UC medication, or 4) have a rescue non-drug treatment (eg, colectomy, ileostomy, or sigmoidectomy) during the study will be considered as nonresponders, that is, have a known outcome of nonresponse at any applicable timepoints.

For primary analysis, all participants with missing data for other reasons will be treated as nonresponders. That is, all missing data will be treated as nonresponse regardless of reasons.

5.3.1.2. Sensitivity Missing Data Handling Methods

Other methods of handling missing data will be applied for primary and key secondary efficacy endpoints, which are all binary, assuming different missing mechanisms: multiple imputation under MAR assumption, tipping point analysis, and copy to reference under MNAR assumption. The analyses based on these methods will all be performed for PAS only.

5.3.1.2.1. Multiple Imputation Under MAR Assumption

Missing data (eg, component scores of MMS at the planned assessment timepoints) will be imputed under the MAR assumption. Binary endpoints will subsequently be computed from observed and imputed data and analyzed using the same method as in the primary endpoint analysis for each imputed dataset. Then the results will be combined using multiple imputation methodology (Rubin, 1987 Section 8 References).

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 (SAS Software v9.4. SAS Institute Inc, Section 8 References). The number of imputations will be 40. 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 participant is a prior UC treatment failure of oral 5-ASA only (ie, participant 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. The random seed **8238946** will be used.

Step 2

For each of the imputed datasets at Week 12 and Week 52, participants will be classified as responder or nonresponder. 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 MH weighted method, stratified by naïve to biologic or JAK inhibitor therapy at study entry (Yes or No) and baseline corticosteroid use (Yes or No) will be run for each of the 40 imputed datasets to obtain 40 estimators of interest at Week 12 and Week 52.

Step 4
Use Rubin's rule and SAS PROC MIANALYZE (Rubin, 1987; SAS Software v9.4. SAS Institute Inc Section 8 References) to produce an overall pooled estimate (mean of 40 estimates) with its associated SE, CI and p-value.

5.3.1.2.2. Tipping Point Analysis

The response status (yes/no) of a participant with missing data is multiply imputed by randomly drawing from a Bernoulli distribution with changing underlying probability independently for each treatment arm. The detailed steps are described below.

- **Step 1**: Assume a response proportion p_0 for all participants with missing data in the placebo group and a response proportion p_1 for all participants with missing data in the etrasimod group.

- **Step 2**: Impute the missing response for each participant from the Bernoulli distribution with response probability p_0 for placebo group and response probability p_1 for etrasimod group, independently. Combine with the participants with

nonmissing response to have a complete dataset. The random seed **8238946** will be used.

- **Step 3**: Repeat Step 2 and generate N (eg, N=40) complete datasets.
- **Step 4**: For each complete dataset, obtain the estimated common risk difference and its standard error using the weighted MH weighted method, stratified by (a) naïve to biologic or JAK inhibitor therapy at study entry (yes or no) and (b) baseline corticosteroid use (yes or no).
- **Step 5**: Combine and obtain the final estimated common risk difference, its 95% 2-sided confidence intervals and test p-values based on the Rubin's rule and SAS PROC MIANALYZE
- **Step 6**: For each combination of (p_0, p_1) over a plausible region (say, p_0 from to 0.3, and p_1 from 0.0 to 0.5), repeat Step 1 through Step 5 to produce a significance tipping point map to help the determination of the tipping point boundary.

5.3.1.2.3. Copy Reference (CR) Multiple Imputation Under MNAR Assumption

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, participants cannot be modelled based on participants with observed data, and more assumptions are needed. Copy Reference approach (Carpenter et al, 2013) will be used. It is a type of missing data handling approach, assuming MNAR, where each missing value of MMS subscores for participants in the treatment group will be imputed using observed MMS subscores in the placebo group, and missing values of MMS subscores for participants in the placebo group will be imputed under the MAR approach. Once the missing MMS subscores are imputed, all binary efficacy endpoints (based on MMS) will be derived based on observed and imputed MMS subscores and analyzed using the same method described in Section 5.2.1 for each imputed dataset. Then the results will be combined using multiple imputation methodology based on the Rubin's rule. The random seed **8238946** will be used.

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 post discontinuation 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 participant 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 5.3.1.2.1 above.

Step 3
Same as Step 3 described in Section 5.3.1.2.1 above.

Step 4
Same as Step 4 described in Section 5.3.1.2.1 above.

5.3.1.2.4. Hybrid Imputation For Missing Data Due to COVID-19 Pandemic and Ukraine war

The COVID-19 pandemic and Ukraine war introduces unexpected and unknown impact on the clinical study due to unforeseen intercurrent events. As a sensitivity analysis for primary and key secondary efficacy endpoints, missing endoscopy data will be handled using a hybrid approach combining nonresponder Imputation (NRI) and multiple imputation, where participants with missing endoscopy data due to COVID-19 pandemic and Ukraine war will be imputed using multiple imputation method under a MAR assumption, as described in Section 5.3.1.2.1**.** The random seed **8238946** will be used. Missing data for any other reason will be considered a nonresponse.

5.3.2. Continuous Non-Cardiac Endpoints

For all non-cardiac continuous (i.e., exploratory efficacy endpoints, HRQoL endpoints and efficacy-related biomarker endpoints) endpoints measured longitudinally, missing values post baseline will not be imputed explicitly. For such endpoints, assuming that the missing data mechanism is missing at random (MAR), the data will be analyzed based on a linear mixed-effect model for these continuous endpoints (Section 5.2.2)

5.3.3. Cardiac Endpoints

For all cardiac endpoints, missing values will not be imputed explicitly.

5.4. Supplementary Analyses for Primary and Key Secondary Efficacy Endpoints

For primary endpoint analysis, the primary model as described in Section 5.2.1 will be repeated using the mPAS (with data as observed) as a supplementary analysis.
For key secondary efficacy endpoint analyses, the primary model as described in Section 5.2.1 will be repeated using the mPAS (with data as observed) as supplementary analyses.

6. ANALYSES AND SUMMARIES

6.1. Summaries and Analyses for Efficacy, HRQoL and Efficacy-Related Biomarker Endpoints

The detailed summaries of analyzing the efficacy, HRQoL and efficacy-related biomarker endpoints listed in Sections 3.1, 3.2.1, 3.4.1, 3.4.2 and 3.4.3 are provided in Table 6 and Table 7 below. All outputs will be based on the PAS, unless otherwise specified.

Table 6. Planned Analyses for Primary and Key Secondary Efficacy Endpoints

Endpoint	Main Analysis	Main Missing Data Handling	Sensitivity Missing Data Handling	Supplementary Analyses
Primary endpoint				
Clinical remission at Week 52	MH weighted method (Section 5.2.1) based on PAS	NRI	MI assuming MAR, Tipping point analysis, CR MI assuming MNAR, hybrid imputation	mPAS (observed data)
Key secondary efficacy endpoint				
Endoscopic improvement at Week 52	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
12-week corticosteroid-free clinical remission at Week 52	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
Clinical remission at Week 12	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
Endoscopic improvement at Week 12	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
Sustained clinical remission (clinical remission at both Week 12 and Week 52)	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
Histologic-Endoscopic Mucosal Improvement at Week 52	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint

Table 6. Planned Analyses for Primary and Key Secondary Efficacy Endpoints

Endpoint	Main Analysis	Main Missing Data Handling	Sensitivity Missing Data Handling	Supplementary Analyses
Complete symptomatic remission at Week 52	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint
Complete symptomatic remission at Week 12	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint	Same as primary endpoint

Table 7. Planned Analyses for Other Secondary Efficacy, Exploratory Efficacy, HRQoL and Efficacy-Related Biomarker Endpoints

Endpoint	Main Analysis	Main Missing Data Handling	Sensitivity Missing Data Handling	Supplementary Analyses
Other secondary efficacy endpoints (Section 3.2.1)	Same as primary endpoint	Same as primary endpoint	N/A	N/A
Exploratory efficacy endpoints: binary endpoints (Section 3.4.1)	Same as primary endpoint	Same as primary endpoint	N/A	N/A
Exploratory efficacy endpoints: continuous endpoints (Section 3.4.1)	linear mixed-effect model based on PAS (Section 5.2.2)	Data will not be imputed explicitly, assuming MAR (Section 5.3.2)	N/A	N/A
HRQoL endpoints (Section 3.4.2)	Same as continuous exploratory efficacy endpoint	Same as continuous exploratory efficacy endpoint	N/A	N/A
Efficacy-related biomarker endpoints (Section 3.4.3)	Same as continuous exploratory efficacy endpoint except that the analysis for lymphocyte counts will be based on Safety Set, not PAS	Same as continuous exploratory efficacy endpoint	N/A	N/A

For continuous endpoint, line plots over time for LS mean and SE by treatment will be provided if deemed appropriate.

6.2. Summaries and Analyses for PK Endpoints

Unless otherwise specified, PK summaries will use the Pharmacokinetic Set.

Plasma concentrations of Etrasimod will be assessed from PK samples collected prior to dosing and 4 hours (± 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 participants 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.

Individual participant etrasimod plasma concentrations (and etrasimod metabolite concentrations, if applicable) will be presented in the data listings, including participant ID, treatment received, 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 participants are expected to achieve steady-state plasma concentration by Week 2. For each participant, the average steady-state trough (predose) plasma concentration based on Week 2 to Week 52 trough concentrations ($C_{trough,ss}$,W2-W52) per participant will also be calculated and presented in a data listing and summarized using descriptive statistics.

Mean (± SD) etrasimod concentration versus nominal time (Weeks 0 to 52) will be plotted on a linear scale.

The pharmacokineticist will determine the strategy for dealing with data affected by protocol deviations or events which may impact the quality of PK concentration data on a case-by-case basis with input from the study physician and clinical pharmacologist, as needed. Examples for protocol deviations or events include, but may not be limited to, vomiting 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.

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

6.3. Safety Summaries and Analyses

All outputs for safety outcomes will be based on the Safety Set.

6.3.1. Adverse Events

6.3.1.1. 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 participants with the AE divided by the total exposure in participant-years.

Severity

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

Relationship to Study Treatment

All related TEAEs will be summarized by SOC, PT, and maximum severity, with SOC and PT presented by descending frequency. If a participant reports a TEAE more than once within a SOC/PT, the AE with the worst-case severity will be used in this summary. A "related TEAE" for the purpose of this summary is defined as a TEAE with relationship to study treatment of "probably related" or "related".

6.3.1.2. TEAEs Leading to Discontinuation of Study Treatment and TEAEs Leading to Interruption of Study Treatment

All TEAEs leading to discontinuation of study treatment will be summarized by SOC and PT. TEAEs leading to interruption of study treatment will be summarized by SOC and PT as well.

6.3.1.3. Serious and Non-Serious TEAEs

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. All nonserious TEAEs will be summarized by SOC and PT.

6.3.1.4. TEAEs of Special Interest

TEAEs of special interest will be summarized by category, subcategory, and PT by descending frequency by treatment group.

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, Section 8 References). The stratification variables, naïve to biologic or JAK inhibitor therapy status at study entry (Yes or No) and Baseline corticosteroid use (Yes or No), will be included in the model. A participant with K events contributes (K+1) observations to the input data set; here K might be 0. The k-th observation of the participant 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 (Etrasimod group vs placebo group), 95% CIs, and 2-sided p-values for treatment effect comparison will be presented.

6.3.1.5. Overall Summary of Adverse Events

In addition to the summaries above, an overview of TEAEs will be summarized by number and frequency of participants 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 the following:

- TEAEs leading to Deaths
- Severe TEAEs (CTCAE grade ≥ 3)
- Serious TEAEs
- TEAEs leading to study treatment discontinuation.
- TEAEs of special interest
- Treatment-emergent Serious infections
- Treatment-emergent Opportunistic infections (Narrow)
- Treatment-emergent Malignancies

Risk differences between etrasimod and placebo along with 95% CIs using the Wilson's score method will be presented. Differences between "Exposure-Adjusted Incidence Rates" (EAIR) comparing treatment to placebo along with 95% CIs based on large-sample approximation assuming Poisson standard errors for the estimated incidence rates as described in Liu et al. will be presented as well.

6.3.2. Other Safety Summaries and Analyses Endpoint(s)

6.3.2.1. Laboratory Data

In general, 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 unless specified otherwise. Local laboratory assessments will be listed.

For the sites and/or countries where central laboratory assessments are not available (due to reason such as war in Ukraine) and local laboratory assessments are collected, the local laboratory assessments could be considered for laboratory related analyses, if deemed appropriate.

Safety Laboratory Evaluations

Hematology, serum chemistry, coagulation, and urinalysis will be 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 Table 5 ("Clinical Laboratory Tests") of Clinical Study Protocol C5041011 (APD334-210) amendment 2.0.

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 9.1.4.

The following summaries will be provided for laboratory data:

- Value and change from Baseline by visit (e.g., hematology, serum chemistry, and coagulation)
- Incidence of markedly abnormal lymphocytes and neutrophils at anytime post-baseline, End of Treatment, any Follow up.
    - Markedly Abnormal lymphocytes and neutrophil is defined as Lymphocytes < 0.2×10^9/L, 0.5×10^9/L, or Neutrophils < 0.5×10^9/L, 1×10^9/L.
- Shift from end of treatment to each follow-up visit in incidence of lymphocytes for participants with abnormal low value at end of treatment.
- 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

  The eDISH plots above will be repeated with adjustment for elevated Baseline.
- Incidence of hepatic enzyme elevations by visit and overall post baseline
    - > 1 ×, 2 ×, 3 ×, 5 ×, 8 ×, 10 ×, 20 × upper limit of normal (ULN) elevation in ALT
    - > 1 ×, 2 ×, 3 ×, 5 ×, 8 ×, 10 ×, 20 × ULN elevation in AST
    - > 3 ×, 5 ×, 10 ×, 20 × ULN elevation in either ALT or AST
    - > 1 ×, 1.5 ×, 2 ×, 3 × ULN elevation in total bilirubin
    - > 1 ×, 1.5 ×, 2 ×, 3 ×, 5 ×, 8 × ULN elevation in ALP
    - > 1 ×, 2 ×, 3 ×, 5 ×, 8 × ULN elevation in GGT
    - > 3 × ULN elevation in either ALT or AST and > 1.5 × ULN elevation in total
    - bilirubin
    - > 3 × ULN elevation in either ALT or AST and > 2 × ULN elevation in total
    - bilirubin

- > 3 × ULN elevation in either ALT or AST and > 1.5 × ULN elevation in ALP
- > 3 × 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 participant's baseline, whichever is higher.
Participant'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 participants who ever had lymphocytes < 0.5×10^9/L or neutrophils < 1×10^9/L will also be provided.

Pregnancy Tests

Pregnancy tests are performed throughout the study (Serum β-hCG test at Screening; urine pregnancy test at all other visits) in female participants of childbearing potential. All pregnancy test results (positive or negative) will be listed in chronological order. Number of female participants with positive urine/serum pregnancy test will be summarized.

Other Screening Laboratory Assessments

Other screening laboratory assessments conducted per Table 5 ("Clinical Laboratory Tests") of Protocol C5041011(APD334-210) amendment 2.0 will be listed.

6.3.2.2. ECG Evaluations

The ECG parameters reported for this study per Protocol C5041011 (APD334-210) amendment 2.0 will be summarized by visits.

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

ECG Evaluations

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 "**ECG Markedly Abnormal Criteria**" above) and AV blocks 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 participants meeting markedly abnormal criteria will also be provided. In this listing,

- For each participant, only ECG parameters ever meeting markedly abnormal criteria will be included.
- For each participant, if there is at least one assessment meeting markedly abnormal criteria for a given ECG parameter, then all assessments collected of this ECG parameters over time will be included.

6.3.2.3. Cardiac Holter Monitoring

hHR

Value and change from baseline in hHR will be descriptively summarized by hourly timepoint.
Time to first nadir hHR post-dose will be descriptively summarized as a continuous outcome measure in the Modified Safety Set. The number and percentage of participants who reached their first nadir hHR at each nominal hourly timepoint (eg, hour 1, hour 2, etc.) will also be descriptively summarized.

Cardiac Abnormal Findings

Incidence of participants with abnormal findings flagged by cardiac central laboratory will be descriptively summarized.

Incidence of participants with low heart rate as defined in Section 3.3.2.1 will be descriptively summarized.

Data Listings

Holter recording data on Day 1 will be listed.

6.3.2.4. Vital Sign

The vital sign parameters reported for this study per Protocol C5041011 (APD334-210) amendment 2.0 will be summarized by visits. Table 8 below lists the markedly abnormal criteria for vital signs.

Table 8. Markedly Abnormal Criteria for Vital Signs

Vital Sign Parameter	Unit	Low	High
SBP	mm Hg	≤ 90 mm	> 150 mm Hg
DBP	mm Hg	≤ 50 mm	> 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

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 8 above).
- Listing of participants 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.

For participants 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.

Vital signs derivation for temperature: temperature (°C) = (5/9) (temperature (°F) – 32).

6.3.2.5. Pulmonary function tests

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:

- Percent decrease from Baseline > 20% in FEV1, ie, percent change from Baseline < -20%
- Percent decrease from Baseline > 20% in FVC, ie, percent change from Baseline < -20%
- Percent decrease from Baseline >20% in DLCO (if available), ie, percent change from Baseline < -20%
- Percent decrease from Baseline ≥30% in FEV1, ie, percent change from Baseline ≤ -30%
- Percent decrease from Baseline ≥30% in FVC, ie, percent change from Baseline ≤ -30%
- Percent decrease from Baseline ≥30% in DLCO (if available), ie, percent change from Baseline ≤ -30%

The PFT parameters reported for this study per Protocol C5041011 (APD334-210) amendment 2.0 will be summarized by visits.

The following summaries will be provided for PFT data descriptively:

- Value and change from Baseline by visit.
- Incidence of markedly abnormal and/or potentially important PFT values by visit per "**PFT Markedly Abnormal and Potentially Important Criteria**" above.

All PFT data will be listed. Abnormal PFTs, markedly abnormal and/or potentially important PFT values will be flagged in the listing.

6.3.2.6. 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.
- Incidence of markedly abnormal change in central foveal thickness (CFT) (> 40 µm increase from Baseline in CFT in either eye) by Visit.
- Categorical result in ophthalmoscopy with OCT parameters by visit (the categories are listed on the eCRF).

A listing of all ophthalmoscopy and OCT assessments will be provided. A listing of participants 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 participants who experienced an AE related to eye disorders.

6.3.2.7. Tuberculosis Questionnaire

Tuberculosis Questionnaire at screening and post screening will be listed.

6.4. Subset Analyses

The following major subgroup analyses for the primary and key secondary efficacy endpoints will be performed based on PAS, in order to explore whether the treatment effects are consistent across different subgroups.

- Age (years):
  - > or ≤ median value (based on PAS)
  - ≥ or < 65 years
- Sex (Male, Female)
- Race (White or Non-White)
- Race (White, Black or African American, Asian, Other, where Other captures the remaining races)
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- 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 fecal calprotectin > or ≤ median value (based on PAS)
- Baseline CRP > or ≤ median value (based on FAS)
- Prior Oral 5-ASA Failure Only (yes or no)

- Naïve to biologic or JAK inhibitor therapy who failed corticosteroid and/or thiopurines (yes or no)

*actual stratum (based on medications reported on the eCRF)

The main subgroup analyses for primary and key secondary efficacy endpoints will be based on the same MH weighted method as mentioned in Section 5.2.1 except that

- For subgroup baseline oral corticosteroid usage (yes or no), the reported stratum of baseline oral corticosteroid usage (yes or no) will not be used in MH weighted method.

- For subgroup naïve to biologic or JAK inhibitor therapy at study entry (yes or no), the reported stratum of naïve to biologic or JAK inhibitor therapy at study entry (yes or no) will not be used in MH weighted method.

Additional subgroups may be assessed, if deemed necessary. If any subgroup includes < 5% of all participants, no inferential statistics will be generated.
Graphical display (eg, forest plots) of the treatment differences between treatment groups will be presented for each of the primary and key secondary efficacy endpoints. The treatment differences between treatment groups, along with 95% CIs, that are obtained from MH weighted method mentioned above will be displayed.

6.5. Baseline and Other Summaries and Analyses

6.5.1. Baseline Summaries

Demographics and medical history including variables defined in Section 3.5 will be summarized by treatment group.

6.5.2. Study Conduct and Participant Disposition

Among the randomized participants, the number and percent of participants who completed/discontinued study treatment, reasons off treatment, the number and percent of participants 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 participants in each analysis set will be summarized for all randomized participants. An additional summary of the number of participants screened, the number of screen-failed participants, and reason for screen failure will be presented for all screened participants. If there are participants whose blind was broken during the study, a listing of these participants will be provided. The number of participants whose visit was impacted by the COVID-19 pandemic will also be summarized.

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 participant's rights, safety, or well-being (ICH 1996, ICH 1998). All-important protocol deviations will be identified before study unblinding. The identified important protocol deviations will be summarized by protocol

deviation categories (Section 9.1.6). All protocol deviations will be listed, including whether a protocol deviation that occurred during by the COVID-19 pandemic was impacted by the COVID-19 pandemic (Yes or No) or the war between Russia and Ukraine (Yes or No).

6.5.3. Study Treatment Exposure

6.5.3.1. Duration of 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.

Duration of exposure (weeks) = (date of last study treatment administration – date of first study treatment administration + 1) / 7. For participants 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.

6.5.3.2. Dose interruption and Overdose

For dose interruptions, the frequency and percentage will be provided for the following:

- participants who had at least 1 dose interruption
- participants who had at least 1 dose interruption of > 7 days
- participants who had at least 1 dose interruption of > 14 days

For overdose, the frequency and percentage will be provided for the following:

- participants who had 1 overdose
- participants who had > 1 overdose will be summarized.

6.5.3.3. Study Treatment Compliance

For 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 participants with overall compliance of < 80% or > 120% will be summarized for the Safety Set for the entire study.

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 participant is expected to have taken between their first and last study treatment administration and is numerically identical to the participant'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.

- 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 participant, if a high percentage (> 25%) of bottles were not returned by a participant, additional analyses may be done where bottles not returned are excluded from the overall compliance calculation for the participant. 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 the entire study and will be used in determining inclusion/exclusion of participants in the respective Per Protocol Set.

6.5.4. Concomitant Medications and Nondrug Treatments

Medications will be captured on the Concomitant Medications eCRF and coded using the WHO Drug dictionary (GlobalB3_March 2024). Refer to Section 9.1.5 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.

7. INTERIM ANALYSES

There are no interim analyses planned for this study.

8. REFERENCES

Cao D, He X. Statistical analysis of adverse events in randomized clinical trials using SAS. PharmaSUG Conference Proceedings; 2011.

Carpenter JR, Roger JH, Kenward MG. Analysis of longitudinal trials with protocol deviation: A framework for relevant, accessible assumptions, and inference via multiple imputation. J Biopharm Stat. 2013;23(6):1352-71.

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.

FDA. (2022). Ulcerative colitis: Developing drugs for treatment. US Department of Health and Human Services, Food and Drug Administration, Silver Spring, MD. Guidance for Industry, https://www.fda.gov/media/158016/download.

FDA. (2022). Multiple endpoints in clinical trials. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry. https://www.fda.gov/media/162416/download

FDA. (2020). Conduct of clinical trials of medical products during the COVID-19 public health emergency. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for Industry, https://www.fda.gov/media/136238/download.

FDA (to Pfizer). (2022) FDA Letter IND 125154 Advice-Information Request IND

ICH. (1996). Structure and content of clinical study reports, E3. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Guideline for Industry, https://www.fda.gov/media/71271/download.

ICH. (1998) Statistical principles for clinical trials, E9. International Council for Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Guidance for Industry, https://database.ich.org/sites/default/files/E9_Guideline.pdf. Accessed: 05 July 2023.

Liu GF, Wang J, Liu K, et al. Confidence intervals for an exposure adjusted incidence rate difference with applications to clinical trials. Stat Med. 2006;25(8):1275-86.

Rubin DB. Multiple imputation for nonresponse in surveys. New York, NY: John Wiley and Sons; 1987. 296 p.

Pfizer (to FDA). (2023) Response to FDA-Clinical IR received 02 December 2022, Study C5041011/APD334-210 (GLADIATOR) Pfizer (to FDA). (2023) Etrasimod UC_Type C Meeting_FDA_GLADIATOR_Proposed Analysis Plan_August 2023

Pfizer (2023). Full Prescribing Information of VELSIPITY™ (etrasimod), https://www.velsipity.com

SAS Software v9.4. SAS Institute Inc., Cary, NC, USA.

9. APPENDIX

9.1. Data Derivation Details

9.1.1. Demographics and Baseline Characteristics Related variables

- 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/m^2) = Weight (kg) / Height $(m)^2$

9.1.2. 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:

Study Day = (date of event – reference start date) + 1

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

Study Day = (date of event – 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.

9.1.3. 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 9.1.4 below. 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.

9.1.4. Definition and Use of Visit Windows in Reporting

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 APD334-210 Amendment 2.0 Appendix 1: SCHEDULE OF ASSESSMENTS. 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 9 below for specific visit windows.

Table 9. Analysis 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 141
Week 24 (Day 169 ± 7)	142 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]	66 to 113 (66 to 141 for assessments impacted by COVID-19 pandemic, or impacted by the war in Ukraine) [c]
Week 32 (Day 225 ± 7) [PFT only]	198 to 253 (198 to 281 for assessments impacted by COVID-19 pandemic, or impacted by the war in Ukraine) [c]
Week 52 (Day 365 ± 14) [b]	337 to 393 (337 to 421 for assessments impacted by COVID-19 pandemic, or impacted by the war in Ukraine) [c]

a. Applicable only for participants 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.
c 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. Assessments impacted by the war in Ukraine are reported in the protocol deviation document.

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". 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 labeled as "Baseline", not "Week 0". 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 two or more measurements for an endpoint fall in a same analysis visit window, the measurement with the date closest to the protocol scheduled visit day will be used in the analysis. If 2 measurements in the same analysis visit window are equidistant from the protocol scheduled visit day, the earlier one will be used in the analysis. If multiple measurements are available on the same day, then the average of the measurements 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.

9.1.5. Partial Date Convention

Imputed dates will NOT be presented in the listings.

Table 10. Algorithm for Treatment Emergence of Adverse Events

Start Date	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

Table 10. Algorithm for Treatment Emergence of Adverse Events

Start Date	Stop Date	Action
		• 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

Table 11. 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

Table 11. Algorithm for Prior and Concomitant medications

Start Date	Stop Date	Action
	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

Table 12. Algorithm for Starting Date of Hospitalization and Start/Stop Dates of Concomitant Medication

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

Table 12. Algorithm for Starting Date of Hospitalization and Start/Stop Dates of Concomitant Medication

Start Date	Stop Date	Action
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.

9.1.6. Protocol Deviation Categories

Protocol deviation categories include but are not limited to the following:

- Informed consent and Process
- Inclusion criteria
- Exclusion Criteria
- Concomitant Medication
- PK/PD
- Laboratory Assessment
- Study Procedures
- Patient Reported Outcome
- Safety
- Efficacy
- IP Administration
- Visit Schedule
- IP Conditions
- Participant IP Compliance
- Participant Discontinuation
- Randomization
- Administrative

- Other

9.1.7. Hourly Timepoint Determination in Cardiac Analysis

Based on dosing date/time, the cardiac central laboratory divided the collection period into 60-minute intervals (eg, from dosing time to dosing time + 60 minutes, from dosing time + 60 minutes to dosing time + 120 minutes, etc.). In some cases, recording from a partial hourly interval was used to derive hourly data if the Holter recording end time didn't coincide with a multiple of 60 minutes from dosing. For Holter recording data before dosing, per study design, each participant is expected to have one pre-dose Holter record with a duration of about 15 minutes. However, there may exist multiple pre-dose Holter records derived by the cardiac central laboratory for the same participant, based on every 60-minute increment backwards from dosing time until the overall Holter start time is reached. Similarly, pre-dose (eg, hour 0) data may be based on recording from a partial hourly interval, depending on Holter recording start/stop time relative to dosing.

The hourly timepoint will be programmatically calculated as the Holter hourly data stop time minus dosing time, converted to hours and rounded to the nearest whole number. If there are more than 1 hourly timepoint from this derivation, the latter one will be assigned to the next whole hourly timepoint. This applies to both pre-dose and post-dose timepoints.

9.2. Endpoint Derivations

9.2.1. 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)) x (Q5/10)]

4) Calculate activity impairment score as: Q6/10

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

9.2.2. 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

A IBDQ total score can also be calculated as the sum of all 32 items.

9.2.3. 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

9.3. Classification of Rescue Therapy

This appendix outlines the algorithm for the Pfizer clinical team/medical reviewers to classify rescue therapies for ulcerative colitis (UC) in a blinded manner. This will help establish intercurrent events for the efficacy estimands. 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 participant 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:
  - AntiTNFα antibodies:
    - ADALIMUMAB
    - CERTOLIZUMAB
    - CERTOLIZUMAB PEGOL
    - GOLIMUMAB
    - INFLIXIMAB
  - Other Biologics:
    - USTEKINUMAB
    - VEDOLIZUMAB
    - MIRIKIZUMAB

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:
    - Thiopurines
      - MERCAPTOPURINE
      - 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
  - MESALAMINE
  - 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

- UPADACITINIB

- 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
      - PREDNISONE
      - TRIAMCINOLONE
      - HYDROCORTISONE

        - HYDROCORTISONE ACETATE
        - HYDROCORTISONE BUTYRATE
        - HYDROCORTISONE SODIUM SUCCINATE
    - Routes:
        - ORAL
        - RECTAL

- Systemic 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
        - PREDNISONE

  - 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

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

9.4. 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 DDMMMYYYY or DDMMMYYYY:hh:mm.

SPELLING FORMAT
English US.

PRESENTATION OF TREATMENT GROUPS
For outputs, treatment groups will be represented as follows and in that order:

Table 13. Treatment Group Display

Treatment Group	For Tables, Figure, and Listings
Etrasimod	Etrasimod
Placebo	Placebo
Screen Failure	Screen Failure
Not Treated[a]	Not Treated

a. To be used for participants 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

Table 14. Visit and Study Period Display

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	Dependent on Analysis Visit assigned, described in Section 9.1.4
2-Week Follow-Up, 4-Week Follow-Up	Follow-Up

LISTINGS

All listings will be ordered by the following, unless otherwise specified:

- 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 participants who did not receive study treatment)
- Participant number (which is expected to incorporate study site/center)
- Date (where applicable)

9.5. Abbreviations

Table 15. List of Abbreviations

Abbreviation	Term
5-ASA	5-aminosalicylic acid
AE	adverse event
AESI	adverse events of special interest
ALC	absolute lymphocyte count
ALP	alkaline phosphatase
ALT	alanine aminotransferase
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	bpm beats per minute
CI	confidence interval
CM	concomitant medication

Table 15. List of Abbreviations

Abbreviation	Term
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
Ctrough,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
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
FEV1	forced expiratory volume in 1 second
FVC	forced vital capacity
GGT	gamma-glutamyl transferase
HR heart rate	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
MMS	modified Mayo score
MNAR	missing not at random
NCS	not clinically significant
NHI	Nancy Histological Index
NRI	nonresponder imputation
NRS	numeric rating scale

Table 15. List of Abbreviations

Abbreviation	Term
OCT	optical coherence tomography
OLE	open-label extension
PFT	pulmonary function test
PAS	Primary Analysis Set
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	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
TNFα	tumor necrosis factor alpha
UC	ulcerative colitis
ULN	upper limit of normal
ULQ	upper limit of quantitation
USPI	United States Prescribing Information
WHO	World Health Organization
WPAI-UC	Work Productivity and Activity Impairment Questionnaire – Ulcerative Colitis