STATISTICAL ANALYSIS PLAN

SUBSTUDY A – PHASE 2

C5041006 (APD334-202 or APD334-202EU)

A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL-GROUP STUDY TO ASSESS THE EFFICACY AND SAFETY OF ORAL ETRASIMOD AS INDUCTION AND MAINTENANCE THERAPY FOR MODERATELY TO SEVERELY ACTIVE CROHN'S DISEASE

AUTHOR: PPD

VERSION NUMBER AND DATE: **V3.0, 24May2023**

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V3.0 (Dated 24May2023 for Protocol C5041006 (APD334-202 or APD334-202EU) (Substudy A – Phase 2).

	Name	Signature	Date
Author:	PPD	See below	
Position:	PPD		
Company:	IQVIA		

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

	Name	Signature	Date
Approved By:	PPD	See below	
Position:	PPD		
Company:	IQVIA		
Approved By:	PPD, PhD	See below	
Position:	PPD		
Company:	Pfizer, Inc.		

MODIFICATION HISTORY

Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	27APR2022	PPD	Not applicable – first version
2.0	02MAY2023	PPD	Added Biologic Naïve as subgroup in Section 7.5. Added Biologic Naïve to Section 10 under Baseline characteristics summary. Updated table of contents Added section about removing the Study Drug Treatment analysis at Week 14 to Section 3.3 Removed IQR calculation from Section 6.6 Added Extension Period compliance calculation and description to Section 14.1 Added analysis section for subjects with CD-related hospitalizations to Section 15.3.1.3 Added section for occurrence of Treatment Emergent Adverse Events at Week 1 vs after Week 1 in the Induction Period to Section 17.1.1 Removed all references to Change from Extension Baseline analyses Added Absolute and Percent Change from Baseline analysis for TBNK parameters to Section 17.3.1 Changed Treatment Group presentation for outputs to account for Extension Period in Appendix 1 Updated APPENDIX 1 decimal places for CDAI and SES-CD Added Histopathology and Lipid panel assessments Updated for new dictionary versions Updated Signatures page Added Changes to Analyses from Protocol for TLFs being Induction Period only Updated Section 6.5 for Baseline considerations Updated Section 6.7 for only important data summarized in listings Updated Section 7.5 for Induction Period only outputs

			Removed Screen Failure line from Section 10 Removed Total Subject Years from Section 13 Updated Section 15.2.3 to remove P-values for CDAI/SESCD CFB outputs Added line about temporary interruptions to Section 17.1.3 Updated Lab summaries in Section 17.3.1 Removed Histopathology Section (Pfizer confirmed outputs not needed) Added overall Adverse Events and exposure section to Appendix 1
3.0	23May2023	PPD	Updated header to reflect Pfizer ownership of Arena Updated signature page for author Added information on previous unblinding to introduction Added that no hypothesis testing will be carried out in Final Analysis to Section 3.3 Updated study treatment exposure to reflect new outputs in Section 3.3 Removed references to Extension Baseline as they are no longer relevant Added Visit windowing upper limit to Section 6.4 Updated regions displayed in Section 9.1 Updated to add higher level Protocol Deviations to Section 9.2 Updated to use Extension Period throughout when referencing this analysis Added note to Section 15 for Endoscopic response subscores Added note of no comparison between Induction Analysis and Final Analysis in Section 15.2.3 Updated Section 16 to use Efficacy Evaluable Set for Final Analysis Updated to Extension Treatment groups in section 17 Added TEAE distinction by period to section 17.1 Updated AESI in section 17.1.5 Added additional Entire Study Tables to Section 17.1.6

			Removed comment about Short 14 week Induction Period from 17.1.6 Updated Section 17.3.1 for hepatic enzymes Added Day 2 change from baseline to Sections 17.4 and 17.5 Added new PR cat to Section 17.4.1 Removed PK output information from Section 18 to reflect sponsor changes Added lower bound for % weight change in Appendix 4 Updated for consistent casing throughout Updated to new protocol number throughout Added footnotes to Table 2 Added footnotes to Table 3 Specified that prohibited medications would only affect Induction Period efficacy outcomes as per sponsor instruction

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical rationale, methods, rules, and conventions to be used in the presentation and analysis of efficacy, safety, tolerability, pharmacokinetic, and efficacy-related biomarker data for Substudy SSA-P2 in Protocol C5041006 (APD334-202 or APD334-202EU). It describes the data to be summarized and analyzed, including specifics of the statistical analyses to be performed. This SAP is based on the protocol (Global Amendment 3.0, dated 22 December 2022 and Country-specific Amendment 2.6, dated 26 Apr 2021). In Amendment 2.5 and later, the endpoint "Proportion of subjects with endoscopic response" was reclassified from a secondary efficacy endpoint to a primary efficacy endpoint. Additionally, a pharmacokinetic secondary endpoint was added for SSA-P2 and the pharmacodynamic endpoints (biomarkers) previously listed as exploratory endpoints for SSA-P2 were reclassified as secondary endpoints. The analyses described in this SAP are aligned with the country-specific Amendment 2.5 and later. This study previously unblinded for Induction analyses under SAP version 1.0, dated 27Apr2022. SAP version 3.0 will cover the analyses provided in the Final Analyses.

Note, the Phase 2b substudy (Substudy 1 – Phase 2b [SS1-P2b]) and the three Phase 3 substudies (Substudy 2 - Induction [SS2-I], Substudy 3 – Maintenance [SS3-M], and Substudy 4 - Long-Term Extension [SS4-E]) will each be summarized in a separate SAP.

Analyses of the genetics (protocol Section 10.10) and exploratory biomarkers (protocol Section 10.11) data will be specified in a separate analysis plan.

2. STUDY OBJECTIVES AND ESTIMANDS

2.1. PRIMARY OBJECTIVES

- To evaluate the safety, tolerability, and efficacy of 2 doses of etrasimod as induction therapy in subjects with moderately to severely active Crohn's disease (CD)

2.2. SECONDARY OBJECTIVES

- To evaluate the long-term safety, tolerability, and efficacy of etrasimod in subjects with moderately to severely active CD
- To evaluate the pharmacokinetic (PK) and pharmacodynamic (PD) effects of etrasimod as induction and maintenance therapy, including changes in lymphocytes, C-reactive protein (CRP), and fecal calprotectin (FCP) in subjects with moderately to severely active CD

2.3. ESTIMANDS

Table 1 describes the primary analysis for the primary and secondary efficacy estimands for SSA-P2. For Secondary endpoint 2 and 3, a sensitivity analysis using LOCF for missing data will also be presented.

Table 1: List of Estimands

Estimand	Definition	Attributes			
		Population	Variable/ Endpoint	Intercurrent event handling strategy	Population-level summary measure
Primary Endpoint	Efficacy of etrasimod on endoscopic response at Week 14	FAS	Simple Endoscopic Score in Crohn's Disease (SES-CD) ≤ 4 and at least 2-point reduction from baseline with no subscore > 1 OR ≥ 50% decrease from baseline in SES-CD at Week 14	Subjects who have received the protocol defined prohibited medications prior to Week 14 efficacy assessment will be treated as nonresponders. Subjects with missing data for any reason will be treated as nonresponders	Proportion of responders and 90% exact confidence intervals by treatment at Week 14
Secondary Endpoint 1	Efficacy of etrasimod on clinical remission Crohn's Disease Activity Index (CDAI) at Week 14	FAS	CDAI < 150 at Week 14	Subjects who have received the protocol defined prohibited medications prior to Week 14 efficacy assessment will be treated as nonresponders.	Proportion of responders and 90% exact confidence intervals by visit and treatment

				Subjects with missing data for any reason will be treated as nonresponders	
Secondary Endpoint 2	Efficacy of etrasimod on SES-CD score at Week 14	FAS	SES-CD score at Week 14	None – observed data only	Change from baseline, including student t-test p-value by treatment at Week 14
Secondary Endpoint 3	Efficacy of etrasimod on CDAI score at Week 14	mFAS	CDAI score at Week 14	None – observed data only	Change from baseline, including student t-test p-value, by treatment

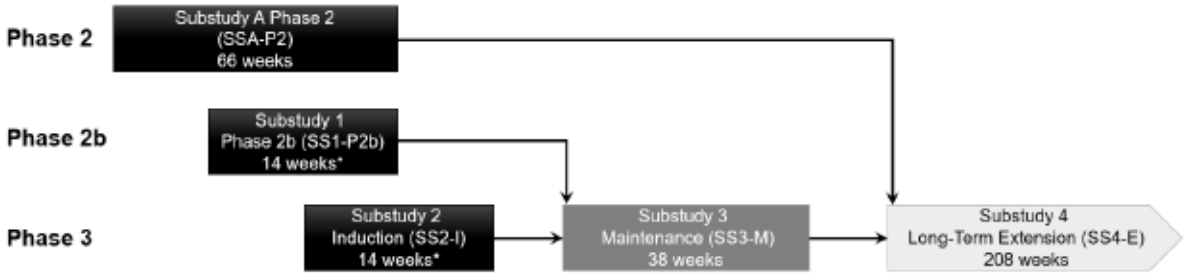
3. STUDY DESIGN

3.1. GENERAL DESCRIPTION

C5041006 (APD334-202 or APD334-202EU) is a seamless Phase 2/3, multicenter, randomized, double-blind, study that comprises 5 substudies designed to evaluate the efficacy, safety, and tolerability of etrasimod as therapy in subjects with moderately to severely active CD who are refractory or intolerant to at least one of the current therapies for CD (i.e., corticosteroids, immunosuppressants, or biologics). Subjects who are refractory or intolerant to corticosteroids and/or immunosuppressants may be either previously exposed to or naïve to biologics. Randomized subjects will be permitted to continue 5-aminosalicyclic acid (5-ASA) compounds, low-dose oral corticosteroids, and/or anti-diarrheal medications as background therapy for CD; however, corticosteroid tapering may be required in subjects who continue treatment beyond the Induction Period.

Overall Study: Overall duration of the study is up to 282 weeks, including a 28-Day Screening Period, Treatment Period of up to 274 Weeks, and the 4-Week Follow-up Period for safety assessments. Figure 1 displays the overall study design and additional information regarding the design of the Phase 2 substudy. The main focus of this SAP is Substudy A – Phase 2 only. The other substudies will be described in separate SAPs.

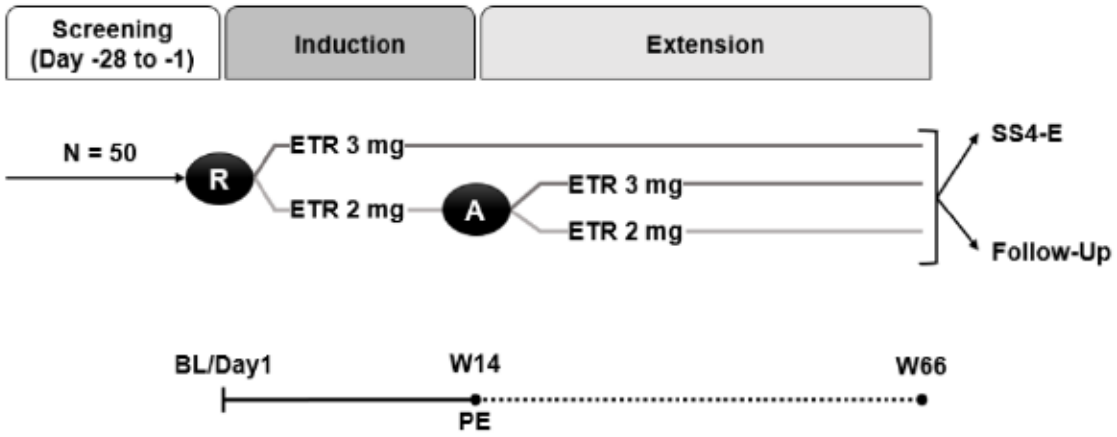
Figure 1: Schematic Diagram of C5041006 (APD334-202 or APD334-202EU) Study Design



*Subjects in SS1-P2b and SS2-I who do not meet prespecified clinical or endoscopic response criteria at Week 14 will enter a 6-week Extended Induction Period.
SSA-P2, Substudy A – Phase 2; SS1-P2b, Substudy 1 – Phase 2b; SS2-I, Substudy 2 – Induction; SS3-M, Substudy 3 – Maintenance; SS4-E, Substudy 4 – Long-Term Extension

Substudy A - Phase 2 (SSA-P2): As shown in Figure 2, this is a Phase 2 randomized, double blind, substudy to assess the safety, tolerability, and efficacy of oral etrasimod therapy in subjects with moderate to severe CD that supports the selection of an induction and maintenance dose(s) for Phase 3.
SSA-P2 plans to enroll and treat approximately 50 subjects, randomized 1:1, with about 25 subjects expected in each of the 2 treatment groups (etrasimod 2 mg and 3 mg). Randomization will be stratified by biologic treatment failure status (yes/no) and baseline oral corticosteroid use (yes/no). Total duration is up to 74 weeks, including a 28-Day Screening Period, 14-Week Induction Period, 52-Week Extension Period, and a 4-Week Follow-up Period (for subjects who do not continue in SS4-E). All subjects who complete the Induction Period may enter the Extension Period. Subjects will receive either 2 or 3 mg in the Extension Period, according to their Induction Period treatment and clinical response (as described in Table 1 of the Protocol). Subjects in the Extension Period who do not show clinical improvement, defined as a ≥ 70-point decrease from baseline for CDAI by Week 20, will be discontinued from study drug. Oral corticosteroids may be administered as a rescue therapy during the Extension Period. If subjects show no clinical improvement from rescue therapy, as determined by the Investigator, they will be discontinued from the study drug. Subjects who complete 66 weeks of treatment will be eligible to enter SS4-E upon completion of the Week 66 Extension visit.
A schematic diagram of SSA-P2 is shown in Figure 2.

Figure 2: Substudy A – Phase 2 Schematic Diagram



A, assigned; BL, Baseline; ETR, etrasimod; PE, primary endpoint; R, randomization; SS4-E, Substudy 4 – Long-Term Extension; W, Week

Protocol Section 3.1.1.1 describes the transitional subject enrollment flow from protocol Amendment 1.0 to Amendment 2.0. Subjects enrolled under Amendment 1.0 who have completed study visits beyond the End of Week 15 Visit will be reconsented under Amendment 2.0 at their next scheduled study visit and will continue in Extension Period for their remaining study visits.

3.2. SCHEDULE OF ASSESSMENTS

The schedule of assessments for SSA-P2 can be found in Table 30 in Appendix 1 of the protocol.

3.3. CHANGES TO ANALYSIS FROM PROTOCOL

- The protocol does not specify any statistical tests for SSA-P2. However, for change from baseline efficacy analyses for CDAI, SES-CD, Patient Reported Outcome 2 (PRO2), Abdominal Pain (AP), Stool Frequency (SF), and Abdominal Pain Stool Frequency (APSF), a t-test will be used. For change from baseline biomarker analyses for absolute lymphocyte count (ALC), fecal calprotectin (FCP), and C-reactive protein (hs-CRP), a Wilcoxon signed rank test will be used. These tests will be conducted for Induction Analysis data only. For the Final Analysis, some of the Induction analyses will be rerun along with some additional analysis for the Induction Period, and the Extension Analysis. There will be no hypothesis testing for the Final Analysis.
- Change from baseline in APSF score and proportion of subjects who achieve clinical remission APSF analyses are not specified in the protocol but will be summarized. These tests will be conducted for Induction Analysis data only. For final induction analysis rerun and Extension Period analysis there is no comparison tests.

- Three exploratory efficacy endpoints were added: Clinical response by PRO2, Clinical response by CDAI, and Endoscopic remission
- The Full Analysis Set (FAS) definition will be updated to include randomized subjects who receive at least one dose of study drug.
- Study Drug treatment during the study will be summarized separately for the Induction Period, Extension Period, and together for the entire study.
- No change from Extension baseline analyses will be presented.

4. PLANNED ANALYSES

The following analyses will be performed for this study:

- Induction analysis after all randomized subjects have either completed Week 14 or discontinued study drug treatment
- Final analysis after all subjects have completed protocol specified treatment, including Extension Period, or discontinued treatment.

Any, post-hoc, exploratory analyses completed to support planned study analyses, which were not identified in this SAP, will be documented and reported in the clinical study report (CSR). Any results from these unplanned analyses will also be clearly identified in the text of the CSR. Summaries from the Induction analysis may be provided to applicable regulatory authorities as required per national regulatory requirements and commitments.

4.1. INTERIM ANALYSIS

No formal interim analyses are planned for this substudy. There may be informal interim analyses performed periodically to monitor safety, PK, and efficacy. These analyses will be descriptive only and will be done on aggregate blinded data.

4.2. INDUCTION ANALYSIS

The induction analysis will occur after all subjects complete 14 weeks of double-blind induction treatment or discontinued treatment.

All planned analyses identified in this SAP will be performed following Arena's Authorization of the SAP, Induction analysis sets, and unblinding of Induction treatment.

After outstanding data queries have been resolved/closed, and the data have been cleaned and finalized, Arena will authorize breaking of the study blind and the analysis of the data will be performed. The SAP

must be approved and signed by Arena before the Induction Database Lock and Induction Period treatment assignment is unblinded.

A separate unblinded team will not be used for the induction analysis since the primary endpoint is assessed at Week 14. However, all investigators and subjects will remain blinded to study drug treatment during the entire Extension Period and until the end of the study. Study drug treatment for subjects who have been treated with etrasimod 2 mg and have not responded at Week 14 and have been randomized to receive etrasimod 2 mg or 3 mg, will be blinded to Sponsor, investigators and subjects, until the end of the study.

There will be no inferential comparisons in efficacy endpoints. Data from the Extension Period will be included in analysis datasets but may not be cleaned. Extension Period data will not be part of any analyses.

Subjects who were enrolled under Amendment 1 will be included in the induction analysis. The subjects who were randomized to receive placebo will be excluded from summary tables but will be included in data listings.

After initial induction analysis, Arena decided to change some table layouts, redefine regions, add, and remove some tables. In addition, endoscopic remission was derived incorrectly in the initial induction analysis (see detail in Section 15). So some induction analysis tables will be reproduced after the final database lock and unblinded. The reproduced tables will be used for the CSR.

4.3. FINAL ANALYSIS

The final analysis of SSA-P2 will occur when all subjects enrolled in the Extension Period have completed the substudy or discontinued treatment and had the safety follow-up.

After all subjects have completed the substudy, outstanding data queries have been resolved/closed, and the data have been cleaned and finalized, Arena will authorize breaking of the blind for the Extension Period (for subjects who were assigned 2 or 3 mg during the Extension Period based on response at Week 14), and the final Extension Analysis will be performed.

The final analyses will summarize the Extension Period alone and reproduce some of the Induction Period tables. Additionally, some adverse events and exposure tables will combine the Extension Period and Induction Period data. There will be no inferential comparisons in efficacy and safety endpoints for the Final analyses.

Subjects who were enrolled under Amendment 1 will be included in the final analysis. The subjects who were randomized to receive placebo will be summarized in tables and data listings.

Final analysis will consist of two sets of table, figures and listings (TFL) displays: reproduced TFLs for the Induction Period, including all subjects who have enrolled in SSA, and TFLs for the Extension Period, including subjects who have entered the Extension Period.

5. ANALYSIS SETS

Agreement and authorization of subjects included/excluded from each analysis set will be conducted prior to the unblinding of the substudy for the Induction Analysis and Final Analysis.

5.1. PROCESS FOR ANALYSIS SET ASSIGNMENT

Before the database is locked for the Induction Analysis and Final Analysis, the subjects excluded from each analysis set and the reason for exclusion will be sent to Arena for review. Any changes to the analysis sets will occur before the data is locked.

After the database is locked, Arena will approve the analysis sets which will be used for the corresponding analyses. Once the analysis sets are approved, the data used for each analysis (Induction or Final) will be unblinded.

5.2. SCREENED SET

The Screened Set will contain all subjects who provide informed consent for this substudy.

5.3. RANDOMIZED SET

The Randomized Set will contain all subjects in the Screened Set who were randomized to the study treatment.

5.4. INDUCTION PERIOD

Agreement and authorization of subjects included/excluded from each analysis set will be conducted prior to unblinding the Induction Period treatment.

5.4.1. Full Analysis Set [FAS]

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

5.4.2. Modified Full Analysis Set [mFAS]

The mFAS will include all randomized subjects who receive at least 1 dose of study treatment and have a baseline measurement and have at least 1 post-randomization measurement. Subjects will be summarized by the treatment to which they were randomized, regardless of treatment actually received. Note, the mFAS can vary with endpoints since some subjects may have the needed data for inclusion in the mFAS for some endpoints but others may not. The mFAS will be explicitly defined for the SES-CD and CDAI endpoints.

5.4.3. Per Protocol Set

The Per Protocol Set will include all subjects from the FAS without major protocol violations that might affect the evaluation of the effect of study drug on the primary endpoint. The Per Protocol Set will be used in supplementary analyses of the primary and secondary endpoints to evaluate the influences of major protocol violators and protocol deviators on the main efficacy results. Subjects may be excluded from the Per Protocol Set if they violate the eligibility criteria or significantly deviate from the study plan, including:

- Study treatment non-compliance (< 80 or > 120%) in the Induction Period
- Received incorrect study treatment for > 7 days in a row in the Induction Period
- Used prohibited medication (new or dose increase from the highest dose received during Screening) that may affect efficacy endpoint in the Induction Period
- Missed Week 14 SES-CD and CDAI assessment in pre-specified analysis window, per Section 6.4, for reason other than discontinued due to disease worsening

All protocol deviations will be reviewed in addition to the programmable deviations (e.g., study treatment non-compliance) to determine if any deviation is significant enough to exclude a subject from the Per Protocol Set. All exclusion flags for the Induction Period will be finalized before study unblinding for the induction analysis of dose-response. Subjects excluded from the Per Protocol Set will be flagged in the Protocol deviation listing. Subjects will be summarized by treatment to which they were randomized, regardless of treatment actually received.

5.4.4. Pharmacokinetic Set

The Pharmacokinetic Set will include all subjects in the FAS with at least 1 quantifiable postdose

pharmacokinetic measurement of etrasimod.

5.4.5. SAFETY SET

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

5.5. EXTENSION PERIOD

5.5.1. EFFICACY EVALUABLE SET

The Efficacy Evaluable Set will include all subjects who receive at least one dose of study treatment and have at least one efficacy measurement during the Extension Period; therefore, the Efficacy Evaluable Set is endpoint specific. Extension efficacy outcomes will be summarized using the Efficacy Evaluable Set.

5.5.2. SAFETY SET

The Safety Set will include all subjects who receive at least one dose of study treatment in the Extension Period. All extension safety outcomes will be analyzed in the Safety Set.

6. GENERAL CONSIDERATIONS

6.1. REFERENCE START DATE AND STUDY DAY

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

- If the date of the event is on or after the reference date then:
  - Study Day = (date of event – reference date) + 1
- If the date of the event is prior to the reference date then:
  - Study Day = (date of event – reference date)

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

6.2. BASELINE

Unless otherwise specified, the baseline is defined as the last non-missing measurement taken on or prior to the date of first dose (including unscheduled assessments). If measurements include time, the date/time will be used to define baseline. Otherwise, only dates will be compared. In the case where the last non-missing predose 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. For Quality-of-life assessments (IBDQ, EQ-5D-5L, WPAI-CD, and SF-36), assessments collected on Day 1 with the same assessment date as the date of first dose, will be considered baseline, regardless of time of collection.

6.3. RETESTS, UNSCHEDULED VISITS, AND EARLY TERMINATION DATA

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

The visit windowing will be applied before missing data are imputed. Listings will include scheduled, unscheduled, retest and early discontinuation data.

6.4. WINDOWING CONVENTIONS

All scheduled study visits are defined relative to Study Day 1, the date of first dose in the Induction Period. Scheduled visit windows are defined in Appendix 1 of the protocol. 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. See Table 2 for specific visit windows.

Table 2: Substudy A – Phase 2 Visit Windows

For all assessments except Vital Signs, ECGs, and Holter monitoring	
Scheduled Study Visit[a] (Protocol Study Day)	**Analysis Visit Window[d] (Analysis Study Day)**
Baseline	≤ 1
Day 1	1
Week 1 (Day 7 ± 1)	2 to 11
Week 2 (Day 14 ± 3)	12 to 22

Week 4 (Day 28 ± 3)	23 to 36
Week 6 (Day 42 ± 3)	37 to 57
Week 10 (Day 70 ± 3)	58 to 85
Week 14 (Day 98 ± 3)	86 to 119 (86 to 130 for assessment impacted by COVID-19 or international conflict occurring in Europe)[b]
Week 20 (Day 140 ± 3)	120 to 155
Week 28 (Day 196 ± 7)	156 to 225
Week 36 (Day 252 ± 7)	226 to 281
Week 44 (Day 308 ± 7)	282 to 337
Week 52 (Day 364 ± 7)	338 to 414
Week 66 (Day 462 ± 7)	415 to 469
For Vital Signs, ECGs, and Holter monitoring	
Scheduled Study Visit[a] (Protocol Study Day)	**Analysis Visit Window for Vital signs, ECGs, and Holter monitoring[d] (Analysis Study Day)**
Baseline	≤ 1
Day 1	1
Day 2	2
Week 1 (Day 7 ± 1)	2 to 11, or 3 to 11[c]
Week 2 (Day 14 ± 3)	12 to 22
Week 4 (Day 28 ± 3)	23 to 36
Week 6 (Day 42 ± 3)	37 to 57
Week 10 (Day 70 ± 3)	58 to 85
Week 14 (Day 98 ± 3)	86 to 102
Week 15 (Day 105 ± 3)	103 to 109
End of Week 15 (Day 111 ± 3)	110 to 119
Week 20 (Day 140 ± 3)	120 to 155
Week 28 (Day 196 ± 7)	156 to 225
Week 36 (Day 252 ± 7)	226 to 281
Week 44 (Day 308 ± 7)	282 to 337
Week 52 (Day 364 ± 7)	338 to 414
Week 66 (Day 462 ± 7)	415 to 469

[a] Visits will only be displayed for assessments that were expected to be collected at these visits, according to the Schedule of Assessments, Table 30, outlined in the Protocol.
[b] Information based on CRF or Clinical Operation file. The extended visit window will only apply to the primary efficacy endpoint.

[c] Applicable only for subjects who require extended monitoring on Day 2 for vital signs and ECGs.
[d] The upper limit of visit windows cannot exceed the subjects last day in the corresponding study period

Windowing will be applied prior to any missing data calculations. The last non-missing measurement taken prior to the date of first dose or date of Week 15 dose (including unscheduled assessments) will be labeled as "Baseline" for the Induction and Extension Periods, respectively. Unscheduled or Screening visits that occurred before the baseline visit will not be assigned an analysis visit for each analysis period being summarized. The early discontinuation visit will be eligible for allocation to an analysis visit. The Follow-up visits will not be included in the visit windows and will be summarized separately without any window applied. For assessments taken less frequently (e.g., 12-lead ECG), the same visit windows apply but only visits where assessments were expected to be collected will be summarized. To ensure that assessments collected during Early Termination are summarized, an End of Treatment (EOT) visit will be included in all Induction analyses to summarize the last assessment on treatment.

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

6.5. STATISTICAL TESTS

For change from baseline in continuous efficacy measures (SES-CD, CDAI, AP, SF, APSF, etc.) and biomarker continuous measures (ALC, FCP, hs-CRP), p-values and 90% confidence intervals (CIs) will be presented. Efficacy measures will use a t-test, while biomarker measures will use Wilcoxon signed rank test. Summary statistics, including the number of subjects, mean, standard deviation (SD), median, minimum, and maximum will be presented for all continuous endpoints. For continuous measures, the interquartile range (IQR) will also be presented. Descriptive statistics, including number of subjects and percentages, will be presented for categorical endpoints, along with 90% CIs using the exact method. Unless otherwise specified, Induction analyses will only display data up to and including Week 14/ET. For Final Analysis, no comparison will be performed. Final analyses for Extension Period will display data for Baseline and Week 14 to Week 66/ET.

6.6. COMMON CALCULATIONS

For quantitative measurements:

- Change from baseline = Test Value at Visit X – baseline Value

- Percent change from baseline = ((Test Value at Visit X – baseline Value) / baseline Value) * 100
- Proportion at Visit X = Number of subjects satisfying criteria at Visit X / Total number of subjects at Visit X

6.7. GENERAL STUDY INFORMATION

A general table with summary of study information will be generated, including date of first subject signed ICF, last subject visit date and the Induction and Extension database lock dates. All analyses will be conducted using SAS® (version 9.4, SAS Institute Inc., Cary, NC).

Important data will be presented in listings.

7. SAMPLE SIZE JUSTIFICATION

This Phase 2 substudy is to evaluate the safety, tolerability, and efficacy of 2 doses of etrasimod with no inferential comparison between groups. With a 1:1 randomization ratio, a sample size of approximately 50 subjects (25 per treatment group) will be reasonable to provide numerical description of safety, tolerability, and efficacy at Week 14.

7.1. MULTICENTER STUDIES

This substudy will be conducted by multiple investigators at multiple centers internationally. Data from all sites will be pooled and statistical analyses will not be adjusted for investigational site, country, or geographic region. The pooling plan will be finalized before unblinding this substudy.
Randomization to treatment groups is stratified by:

- Baseline Biologic treatment failure (Yes/No)
- Baseline Oral corticosteroid use (Yes/No)

7.2. ADJUSTMENTS FOR COVARIATES AND FACTORS TO BE INCLUDED IN ANALYSES

The following covariates and factors, and/or endpoint baseline measure may be used in the analyses. For details of their inclusion in the analyses, see the specific analysis section.

- Baseline Biologic treatment failure (yes/no)
- Baseline Oral corticosteroid use (yes/no)

If a subject was assigned to an incorrect stratum at the time of randomization, their stratum based on the

Interactive Web Response System will be used for the statistical analyses. Analyses may be repeated using actual stratum if mis-stratification occurs in more than 10% of randomized subjects.

7.3. MISSING DATA

Missing adverse event (AE) relationship to study drug and AE seriousness will be imputed as described in Section 17.1.1 and 17.1.3, respectively. Partial or missing AE start dates, Concomitant Medication (CM) start/end dates and Crohn's disease diagnosis dates will also be imputed as described in APPENDIX 2. The Crohn's disease diagnosis date will be imputed to calculate the duration of disease only. No other missing safety data will be imputed.

Missing efficacy data will be handled as described in Sections 15.1.2, 15.2.2, and 15.3.2.

7.4. MULTIPLE COMPARISONS AND MULTIPLICITY

There are no adjustments for multiple comparisons or multiplicity.

7.5. EXAMINATION OF SUBGROUPS

The following subgroups will be assessed for the primary and secondary efficacy endpoints and the PRO2 remission and PRO2 change from baseline exploratory endpoints for the Induction Analysis only (specified in Sections 15.2.1 and 15.3.1):

- Sex (Female or Male)
- Race (White or Non-white)
- Age (≤ or > Median)
- Crohn's disease duration (≤ 5 or > 5 years, ≤ 15 or > 15 years)
- Disease localization (ileum only, colon only, both)
- Baseline CDAI score (≤ 300 or > 300)
- Baseline SES-CD score (≤ or > Median)
- Baseline CRP (≤ 3 mg/L or > 3 mg/L)
- Baseline biologic treatment failure status (Yes or No)*
- Prior Biologic Therapy (Yes or No)
- Baseline oral corticosteroid usage (Yes or No)*
- Baseline fecal calprotectin (≤ or > Median)
- Endoscopic response status at the end of the 14-week Induction Period**
- * Stratification reported by the interactive web response system (IWRS) will be used
- ** Subgroup for Final Analysis only

Additional subgroups may be assessed, if deemed necessary, and will be finalized before the Induction

analysis and Final analysis unblinding. Baseline subgroups will be based on Induction baseline values only.

8. OUTPUT PRESENTATIONS

APPENDIX 1 contains the conventions for presentation of data in outputs.

9. DISPOSITION AND WITHDRAWALS

All subjects who provide informed consent will be accounted for in this study. The following analyses will be performed.

9.1. DISPOSITION

Among the randomized subjects, the number and percent of subjects who completed/discontinued the treatment period, the number and percent of subjects who completed/discontinued the treatment period/ study, reasons for discontinuation, and the number and percent of subjects who continued to Extension Period will be summarized. This summary will also be provided by region and country. In the Final Analysis, the regions to be summarized include: North America, Eastern Europe, Western Europe, Eastern Asia, and Rest of World and will be derived based on country. The number and percent of subjects in each analysis set will be summarized in all randomized (or enrolled) subjects. An additional summary of the number of subjects screened, the number of screen-failed subjects, and reason for screen failure will be presented for all screened subjects. Listings of subject disposition, informed consent and reconsent will be presented.

Due to FDA Guidance (June 2020), to assess the impact that Coronavirus 2019 (COVID-19) had on subjects' participation in the substudy, a listing displaying the date of visit, type of visit (in-clinic or remote), and whether the assessments were impacted by COVID-19 will be presented. Only assessments impacted or not done due to COVID-19 will be displayed.

9.2. PROTOCOL DEVIATIONS

Protocol deviations will be summarized for the FAS in the Induction analysis and Safety Set for the Final analysis. After May 16th, 2022, protocol deviations were categorized under a slightly different set of categories and include an importance classification (Y/N). All deviations will include a severity of Major, Minor or Critical. Critical/major deviations collected before May 16th, 2022 will be classified as important. Details describing the protocol deviation categories before and after May 16th are described in the Protocol Deviations Management Plan (PDMP). Important deviations will be summarized for the

Final analysis by descending frequency in the total group, using a combination of the old and new categories as listed below.. All protocol deviations will be listed, seen here in alphabetical order.

- Administrative
- Blinding
    - Concomitant Medication
- Efficacy
- Eligibility and Entry Criteria
    - Inclusion Criteria
    - Exclusion Criteria
- Informed Consent and Process
- Investigational Product Compliance
    - IP Conditions
    - IP Preparation
    - IP Administration
    - Subject IP Compliance
    -
    - Laboratory assessment
- Other Criteria
- Patient Reported Outcomes
- PK/PD
- Randomization
- Safety
- Serious Adverse Event Criteria
    - Study Procedures
    - Subject Discontinuation
- Visit Schedule

10. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

The following demographic data and baseline characteristics will be summarized for the FAS in the Induction Analysis and Safety Set for the Extension Analysis.

- Age on consent (years) – including age groups <Median and ≥Median; ≤18, >18 to <65, and ≥65
- Sex
- Race
- Ethnicity
- Woman of childbearing potential (Yes/No)
- Height (cm)
- Weight (kg)

- BMI (kg/m^2)
- Alcohol consumption (Yes/No)
- Tobacco use (Yes/No)

The following baseline characteristics related to CD will be summarized:

- Duration of CD at consent (years, ≤ 5/ >5, ≤15/ >15)
- Severity of disease (Mild, Moderate, Severe)
- Disease localization (Ileum, Colon, Both)
- Hospitalizations for CD (Yes/No)
- Surgery for CD (Yes/No)
- Procedures for CD (Yes/No)
- Any acute exacerbations within past 12 months (Yes/No), including the number of acute exacerbations among those with any acute exacerbation = Yes
- Baseline biologic treatment failure randomization strata (Yes/No)
- Baseline oral corticosteroid use randomization strata (Yes/No)
- Baseline 5-ASA compound use (Yes/No)
- Baseline AP subscore
- Baseline SF subscore
- Baseline CDAI score
- Baseline SES-CD score
- Baseline PRO2 score
- Prior Biologic Therapy (Yes/No), including the number of prior biologic therapies

10.1. DERIVATIONS

- Duration of CD (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

11. MEDICAL HISTORY

Medical history will be collected on the Medical History eCRF and coded using Medical Dictionary for Regulatory Activities (MedDRA, version 25.1). All medical history will be summarized for the Safety Set by system organ class (SOC) and preferred term (PT).

12. MEDICATIONS

Medications will be captured on the Concomitant Medications eCRF and coded using the WHO Drug dictionary (WHODDE01SEP2022). See APPENDIX 2 for handling of partial dates for medications. In the case where it is not possible to define a medication as prior or concomitant, the medication will be classified by the worst case (i.e. 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 and prior non-CD medications will be summarized by ATC Level 2 and Preferred Drug name for the Safety Set. Non-Drug Treatments used for CD will be flagged as such on the Concomitant Non-Drug Treatment eCRF and also summarized by ATC Level 2 and Preferred Drug name for the Safety Set. Prior treatment for CD will also be summarized by treatment. Drug class type, duration and reason for discontinuation will also be presented.

13. STUDY TREATMENT EXPOSURE

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

Dose interruptions will be recorded on the Dosing Administration eCRF and overdoses will be recorded on the Overdose eCRF. The frequency and percentage of subjects who had at least one dose interruption, who had at least one dose interruption of >7 days, who had at least one dose interruption of >14 days, who had one overdose, and who had >1 overdose will be summarized for the Induction Period and Extension Period. Dose interruption summaries will be cumulative (i.e. subjects with at least one dose interruption >14 days will be counted in the >7 day summary as well).

13.1. DERIVATIONS

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

14. STUDY TREATMENT COMPLIANCE

Total number of tablets expected, total number of tablets taken, total number of tablets missed, overall compliance to study treatment, frequency, and percentage of subjects with overall compliance < 80%, 80 – 120%, and > 120% will be summarized for the Safety Set for the Induction Period and Extension Period.

14.1. DERIVATIONS

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

- Induction/Extension Period compliance will be calculated as follows:

$$\frac{([\text{Total N of tablets dispensed in period}] - [\text{Total N of tablets returned in period}])}{2\,x\,([\text{Date of last dose in period}] - [\text{Date of first dose in period}] + 1)} \times 100$$

- Overall Study compliance will be calculated as follows:

$$\frac{([\text{Total N of tablets dispensed in the study}] - [\text{Total N of tablets returned in the study}])}{2\,x\,([\text{Date of last dose in the study}] - [\text{Date of first dose in the study}] + 1)} \times 100$$

If a bottle is not returned, all dispensed tablets will be assumed to be taken (i.e. number returned will be treated as 0). If a high percentage of bottles are not returned, additional analyses may be done where bottles not returned are excluded from the overall compliance calculation.

Both scheduled and unscheduled study treatment dispensations will be used in the compliance calculation. The compliance calculation will be performed and used in determining inclusion/exclusion of subjects in the respective Per Protocol Set for the Induction analysis. For subjects discontinued prior to Week 14, their study compliance for the Induction Period and the Entire Study will be identical. For subjects who remain in the study beyond Week 14, the Induction Period duration will be defined as the day before the Date of Week 15 visit – Date of first dose (instead of Date of Week 14 visit), to include subjects who discontinued between the end of Induction Period (Week 14) and start of Extension Period (Week 15). The Extension Period duration will be defined as Date of Week 66 visit – Date of Week 15 visit, since the Week 15 visit will be the first dose of administration in the Extension Period since the end of the Induction Period.

15. EFFICACY OUTCOMES

The following definitions will be used to assess efficacy outcomes:

- Endoscopic response**: endoscopic remission or ≥ 50% decrease from baseline in SES-CD

  - Endoscopic remission: SES-CD ≤ 4 and at least 2-point reduction from baseline with no subscore > 1. Note: subscore refers to variable subscore, not segmental subscore, not each variable score per segment. In the initial induction analysis, each variable score per segment was incorrectly used as subscore.

- Clinical response CDAI: Clinical remission CDAI or ≥ 100 point decrease from baseline in CDAI
- Clinical remission CDAI**: CDAI < 150
- Clinical response PRO2: Clinical remission PRO2 or ≥ 8-point decrease from baseline in PRO2
- Clinical remission PRO2: PRO2 < 8
- Clinical remission APSF: Unweighted average worst daily AP score ≤ 1 and unweighted average daily loose/watery SF score ≤ 3
- Time to PRO2 remission and FCP normalization: PRO2 < 8 and FCP ≤ 150 mg/Kg (subjects who do not achieve the goal or discontinue from study, are censored at 7 days after last study drug dose)
- Time to PRO2 response and FCP normalization: PRO2 < 8 or ≥ 8-point decrease from baseline) and FCP ≤ 150 (subjects who do not achieve the goal or discontinue from study, are censored at 7 days after last study drug dose)

**Indicates the response criteria for primary and secondary efficacy endpoints for SSA-P2

The SES-CD is an endoscopic grading system that consists of a composite score based on 4 components: the size of mucosal ulcers, the extent of the ulcerated surface, the endoscopic extension, and the presence of stenosis. Each of the four SES-CD components are assessed in the five segments of the ileum and colon: ileum, right, transverse, left, and rectum. The SES-CD is the sum of the individual scores of each of the components across the five segments. SES-CD scoring details can be found in APPENDIX 3.

The CDAI consists of 8 variables. Two of these, the abdominal pain (AP) and general well-being, are subjective and related to the disease, and each is weighted according to its ability to be predictive of disease activity. Stool frequency, hematocrit, use of anti-diarrheal medications, gender, body weight, height, assessment for EIMs of CD, and presence/absence of an abdominal mass and fever are also needed for the CDAI calculation. The total score ranges from zero to over 600. Details regarding the CDAI derivation can be found in APPENDIX 4.

The PRO2 is a patient-reported outcome measure based on stool frequency (SF, loose/watery stool for the loose/watery score is defined as type 6 [“fluffy pieces with ragged edges, a mushy stool”] or type 7 [“watery, no solid pieces, entirely liquid”] on the Bristol Stool Form Scale) and abdominal pain (AP, average rating graded from 0 [none] to 3 [severe] each day for 7 days) components of the CDAI. The

CDAI weighting factors for the SF and AP variables will be used to calculate the PRO2. PRO2 scoring details and weighting factors can be found in APPENDIX 4.

For all endpoints, subjects who meet the response criteria specified above will be referred to as responders. Subjects who do not meet the response criteria will be referred to as nonresponders. The following table specifies the analysis method(s) that will be used for each efficacy parameter.

Table 3: Efficacy Parameters and Analysis Method for Induction Analysis

Parameter	Analysis Set and Method[a]				Subgroup Analysis Set (Method)
	FAS	mFAS	PP	CP	
Primary endpoint					
Endoscopic Response	NRI	OBS	OBS	OBS	FAS (NRI)
Secondary endpoints					
Clinical remission CDAI	NRI	OBS	OBS	OBS	FAS (NRI)
SES-CD score[a]	OBS	OBS	OBS	OBS	mFAS (OBS)
CDAI score[a]		OBS	OBS	OBS	mFAS (OBS)
Biomarkers (ALC, CRP, FCP)		OBS			
Exploratory endpoints					
PRO2 score and subscores		OBS			mFAS (OBS)
APSF score and subscores		OBS			
Clinical remission PRO2		OBS[c]			mFAS (OBS)
Clinical response PRO2		OBS			
Clinical response CDAI		OBS			
Endoscopic remission		OBS			
Clinical remission APSF		OBS			
Time to loss of clinical remission/response CDAI[b]		OBS			
Time to remission/response by PRO2/Normalization of FCP		OBS			
Quality of life (IBDQ, SF-36, EQ-5D-5L, WPAI-CD)		OBS			
Fistulas, hospitalizations, surgeries	OBS				

Abbreviations: ALC, absolute lymphocyte count; APSF, abdominal pain stool frequency; BL, baseline; CDAI, Crohn's Disease Activity Index; CP, completers; CRP, c-reactive protein; EQ-5D-5L, EuroQoL-5 Dimensions 5-Level Version; FAS, full analysis set; FCP, fecal calprotectin; IBDQ, Inflammatory Bowel Disease Questionnaire; mFAS, modified full analysis set; NRI, nonresponse imputation; OBS, observed

data; PP, per protocol; SES-CD, Simple Endoscopic Score in Crohn's Disease; SF-36, Medical Outcomes Study 36 Item Short Form Health Survey; WPAI-CD, Work Productivity and Activity Impairment Questionnaire – Crohn's Disease.
Note: Blank cells indicate the analysis will not be presented.
[a] Adding sensitivity analysis for last observation carried forward by visit up to Week 14.
[b] Final analysis only.
[c] For the Final analysis, Clinical Remission in PRO2 will be reproduced for the Induction Period using FAS (NRI)

In final data analysis for Extension Periods, efficacy outcomes of endoscopic response, clinical remission CDAI, endoscopic remission, PRO2 remission, PRO2 response and
CDAI response will be analysed in efficacy evaluable set as following.

- Proportion of subjects achieving goal at Week 14;
- Proportion of subjects achieving goal at Week 52/Week 66;
- Proportion of subjects achieving goal at Week 14 and at Week 52/Week 66;
- Proportion of subjects achieving goal at Week 52/Week 66 among those who achieved goal at Week 14 (based on observed data only).

15.1. PRIMARY EFFICACY

The primary efficacy endpoint definition is defined in Section 15. Primary efficacy analyses will be presented by visit, unless specified otherwise.

15.1.1. PRIMARY EFFICACY VARIABLES AND DERIVATIONS

- The proportion of subjects with endoscopic response

The responder status for endoscopic response at Week 14 will be captured on the Responder-2 CRF page. Responder classification will be rederived using the SES-CD assessments. The derived responder status will be used in analyses. The responder status collected on the Responder-2 CRF page will be listed and discrepancies between the derived response and recorded response will be flagged.

15.1.2. MISSING DATA METHODS FOR PRIMARY EFFICACY VARIABLES

Subjects who took any prohibited medication during the Induction Period, as confirmed after blinded review by clinical and medical team members before unblinding, will be considered to have an outcome of nonresponse in the analysis of all efficacy endpoints at any subsequent timepoints in the Induction Period. The prohibited medications will be identified by the Arena clinical team during blinded data review.

Subjects with missing efficacy outcomes will be included in the analyses using the following missing data methods:

- Primary method: single imputation as nonresponder
- Supplementary analyses (Induction analysis): mFAS, Complete case analysis, Per Protocol Set analysis. These analyses are based on observed data only.

15.1.3. Primary Analysis of Primary Efficacy Variables

For the Induction Analysis, the primary analysis will be performed using the FAS. The endpoint will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals will be computed using the exact method. Missing data will be imputed as nonresponder.

For the Final Analysis, the analysis will be performed using the Efficacy Evaluable Set. Summaries will be displayed for the two Extension treatment groups (etrasimod 2 mg to 2 mg, etrasimod 2 mg to 3 mg, and etrasimod 3 mg to 3 mg) and two Induction treatment groups (etrasimod 2 mg and 3 mg). The etrasimod 2 mg treatment group will also be broken down by Week 14 Response. The endpoint will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals will be computed using the exact method. Missing data will be imputed as nonresponder.

15.1.4. Sensitivity Analysis of Primary Efficacy Variables

No sensitivity analyses will be performed.

15.1.5. Supplementary Analysis of Primary Efficacy Variables

For the Induction Analysis, the primary analysis will be repeated using the mFAS, Complete case analysis, and Per Protocol Set as supplementary analyses. The Complete case analysis will include subjects who complete 14 weeks of study treatment. Unless specified otherwise, supplementary analyses are based on observed data, ie, without missing data imputation.

15.2. Secondary Efficacy

The secondary efficacy endpoint definitions are defined in Section 15. All secondary efficacy analyses will be presented by visit, unless specified otherwise.

15.2.1. Secondary Efficacy Variables and Derivations

15.2.1.1. Efficacy

- The proportion of subjects with clinical remission CDAI
- Change from baseline in SES-CD score
- Change from baseline in CDAI score

The responder status for clinical remission CDAI at Week 14 will be captured on the Responder-2 CRF page. Responder classification will be rederived using the CDAI assessments. The derived responder status will be used in analyses. Discrepancies between the derived response and collected response at Week 14 will be flagged in the listings.

15.2.1.2. Pharmacokinetic

- Etrasimod plasma concentrations at 4 hours postdose and at steady-state trough concentrations ($C_{trough,ss}$)

15.2.1.3. Pharmacodynamic

- Change and percentage change from baseline in ALC
- Change and percentage change from baseline in FCP concentration
- Change and percentage change from baseline in CRP

15.2.2. Missing Data Methods for Secondary Efficacy Variables

Missing data methods will be applied to endpoints in Section 15.2.1.1 15.2.1.1only. The other secondary efficacy endpoints will be based on observed data only.

Subjects with missing efficacy outcomes will be included in the analyses using the following missing data methods:

- Primary method (proportion-based endpoint): single imputation as nonresponder
- Primary method (continuous endpoints): Observed data only
- Sensitivity analyses (continuous endpoints for Induction analysis): Last observation carried forward (LOCF)**
- Supplementary analyses (Induction analysis): mFAS, Complete case analysis, Per Protocol Set analysis using observed data only
- Supplementary analyses (Final analysis): observed data only for proportion-based endpoint
    - ** For SES-CD, the baseline assessment is eligible to be carried forward since assessments are only collected at baseline, Week 14 and Week 52. For CDAI, see APPENDIX 4 for details regarding CDAI missing data imputation.

15.2.3. Primary Analysis of Secondary Efficacy Variables

The following analyses apply to the efficacy and pharmacodynamic endpoints only. The pharmacokinetic analyses are described in Section 18.

For the Induction Analysis, proportion-based endpoints will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals will be computed using the exact method. Missing data will be imputed as nonresponder. Continuous endpoints will be summarized using the number of observations, mean, standard deviation, median, minimum, maximum, p-values and 90% CI of mean for change from baseline summaries. P-values will be calculated using the student's t-test. No comparison will be performed for Induction Period in the Final Analysis.

For the Extension Period, the analyses will be performed using the Efficacy Evaluable Set for CDAI and SES-CD score endpoints. Summaries will be displayed for the two Extension treatment groups (2 mg and 3 mg) and two Induction treatment groups (2 mg and 3 mg). The etrasimod 2 mg treatment group will also be broken down by Week 14 Response. Proportion based endpoints will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals will be computed using the exact method. Missing data will be imputed as nonresponder. Continuous endpoints will be summarized using the number of observations, mean, standard deviation, median, minimum, maximum and 90% CI of mean for change from baseline summaries. Change from baseline summaries will be displayed for change from Induction Period baseline.

Value, change, and percent change from baseline in level of fecal calprotectin, hs-CRP, and lymphocyte counts at protocol-specified visits will be summarized. For change and percent change from baseline analyses (Induction Period only), p-values will be presented, using the Wilcoxon signed rank test. mFAS will be used for the Induction Analysis and Safety Set will be used for the Final Analysis. The exploratory efficacy-related biomarker analyses for immunophenotyping, proteomics, RNA transcriptomics, fecal microbiome, and endoscopic healing index will be described in a separate plan.

15.2.4. Sensitivity Analysis of Secondary Efficacy Variables

For Change from baseline in SES-CD and CDAI endpoints, the primary analysis will be repeated using LOCF as a sensitivity analysis for the Induction Analysis.

15.2.5. Supplementary Analysis of Secondary Efficacy Variables

For the Induction Analysis, the primary analysis for the SES-CD and CDAI endpoints will be repeated using the mFAS, Complete case analysis, and Per Protocol Set as supplementary analyses. The Complete case analysis will include subjects who complete 14 weeks of study treatment and have no missing data.

For the Final Analysis, clinical remission CDAI will be repeated using observed data only.

15.3. EXPLORATORY EFFICACY

Exploratory efficacy endpoints are defined in Section 15. All exploratory efficacy analyses will use the mFAS/Efficacy Evaluable Set and will be presented by visit, unless specified otherwise.

15.3.1. EXPLORATORY EFFICACY VARIABLES AND DERIVATIONS

15.3.1.1. CDAI/PRO2

- Change from baseline in PRO2 score, including weighted averaged AP and SF subscores
- Proportion of subjects with clinical remission by PRO2
- Proportion of subjects with clinical response by PRO2
- Change from baseline in Abdominal pain score and Stool Frequency Score (APSF), un-weighted averaged AP and SF subscores
    - Proportion of subjects with clinical remission by APSF
    - Time to loss of clinical CDAI remission (Final analysis only), defined as remission lost date – remission achieved date at Week 14 + 1. Remission lost date is the first assessment date since subjects who had achieved clinical remission at Week 14 (CDAI < 150) no longer meet the remission criteria for two or more consecutive visits. The assessment date from the first week where remission was lost will be used as the remission lost date in analyses.
- Time to loss of clinical CDAI response (Final analysis only), defined as response lost date – response achieved date at Week 14 + 1. Response lost date is the first assessment date since subjects who had achieved clinical response at Week 14 (CDAI < 150 or ≥ 100-point decrease from CDAI baseline) no longer meet the response criteria for the two or more consecutive visits. The assessment date from the first week where response was lost will be used as the response lost date in analyses.
- Time to remission by PRO2 and normalization of FCP, defined as first assessment date when both remission by PRO2 and normalization of FCP occur.
- Time to response by PRO2 and normalization of FCP, defined as first assessment date when both response by PRO2 and normalization of FCP occur.
- Proportion of subjects with clinical response by CDAI
    - Proportion of subjects with endoscopic remission by SES-CD

15.3.1.2. Other PROs

- Change from baseline in IBDQ, SF-36, WPAI-CD, and EQ-5D-5L

- Change from baseline in IBDQ Bowel Symptom domain score
- Change from baseline in IBDQ fatigue item score

15.3.1.3. Fistulas, Hospitalizations, and Surgery

- The proportion of subjects with draining fistulas by visit up to the end of treatment, among subjects with draining fistulas at induction study baseline (Safety Set for Final Analysis)
- Proportion of subjects with CD-related hospitalizations or surgery (Safety Set for Final Analysis)

15.3.2. Missing Data Methods for Exploratory Efficacy Variables

No missing data will be imputed for the exploratory efficacy variables. Data will be analyzed as observed.

15.3.3. Analysis of Exploratory Efficacy Variables

Same analyses described in Section 15.2.3 will be used for the exploratory efficacy variables. CD-related hospitalizations and surgeries will be summarized for the FAS for the Induction analysis and Safety Set for Final analysis. Subjects who experience more than 1 hospitalization or surgery during the substudy will have the duration of the events summed so the total duration is summarized.

Time to event analyses will be summarized including descriptive statistics, Kaplan-Meier estimates of 25th, 50th, and 75th time to event percentiles, and 90% confidence intervals. Kaplan-Meier estimates will be presented as survival curves by treatment. Normalization of FCP is defined as FCP $\leq$ 150 mg/kg. Time to event summaries will use the mFAS for the Induction analysis and Efficacy Evaluable Set for the Final analysis. Time to loss of event summaries will use the subset of subjects in the Efficacy Evaluable Set who were responders at Week 14.

For the Induction Analysis, subjects who do not achieve the event and are not discontinued will be censored at 7 days after their Week 14 assessment date. Subjects who do not achieve the event and discontinued from the study will be censored at the date of discontinuation.

For the Final Analysis, subjects who do not achieve the event or loss of event and are not discontinued will be censored at 7 days after their last dose of study drug. Subjects who do not achieve the event or loss of event and discontinued from the study will be censored at the date of discontinuation.

15.3.4. Sensitivity Analysis of Exploratory Efficacy Variables

No sensitivity analyses will be performed.

15.3.5. Supplementary Analysis of Exploratory Efficacy Variables

No supplementary analyses will be performed.

16. Health-Related Quality of Life (HRQoL) Analysis

Subject-reported HRQoL instruments will be electronically captured and used in support of the efficacy outcomes. All HRQoL are administered at Baseline (except PGIC), Week 4, Week 10, Week 14, Week 28, and Week 66/ET. Even though the protocol Schedule of Assessments (SoA) doesn't specify that HRQoL assessments are collected at Week 10, data was collected and will be included in analyses. The HRQoL analyses will be performed using the mFAS for the Induction Analysis and Efficacy Evaluable Set for the Extension Analysis, with data as observed, unless otherwise specified.

16.1. Variables and Derivations

16.1.1. Inflammatory Bowel Disease Questionnaire (IBDQ)

The IBDQ is a validated 32-item self-administered questionnaire which has 4 dimensions: bowel systems (10 items), systemic systems (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". Overall score ranges from 32 to 224, a higher score indicates better quality of life. The IBDQ total score and 4 domain subscores are derived on the eCRF directly. Details about scoring rules can be found in 0. The IBDQ total score, Bowel systems subscore, Emotional health subscore, Systemic systems subscore, Social function subscore, and Fatigue item score will be summarized. Fatigue item score corresponds to question 2 on the questionnaire.

16.1.2. Abdominal Pain Numeric Rating Scale (NRS)

The abdominal pain NRS is a single item that measures the "worst abdominal pain in the past 24 hours" using an 11-point NRS ranging from 0 (no pain) to 10 (pain as bad as can imagine). Change from baseline in abdominal pain NRS will be summarized by visit and treatment group.

16.1.3. Medical Outcomes Study 36 Item Short Form Health Survey (SF-36)

The SF-36 is a 36-item, subject-reported survey of subject health. The SF-36 consists of 36 questions measuring 8 health domains: physical functioning, bodily pain, role limitations due to physical problems, role limitations due to emotional problems, general health perceptions, mental health, social function, and vitality. The subject's responses are solicited using Likert scales that vary in length, with 3 to 6 response

options per item. The SF-36 will be scored using 2 overall summary scores: physical component summary and mental component summary scores. A higher score indicates better health status.
The Physical Component Summary Measure (norm-based), the Mental Component Summary Measure (norm-based), 8 domain subscores (0-100 based), and the SF-6D health utility index score are derived by Optum and integrated into the SF-36v2 eCRF. Details about scoring rules and derivations of domain scores can be found in APPENDIX 6. The SF-36 health utility index score will be summarized.

16.1.4. EuroQoL-5 Dimensions 5-Level Version (EQ-5D-5L)

The EQ-5D 5-level version includes one question for each of the five quality of life dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression with 5 levels of severity ranging from 1 = "no problems" to 5 = "extreme problems". The EQ-5D-5L questionnaire also includes a visual analog scale (VAS), by which respondents can report their perceived health status with a grade ranging from 0 (the worst possible health status) to 100 (the best possible health status). Change from baseline for the EQ-5D-5-L VAS score will be summarized. Additionally, the shift from baseline to End of Treatment in each dimension score will be summarized by treatment group for the Induction Analysis only.

16.1.5. Work Productivity and Activity Impairment Questionnaire – Crohn's Disease (WPAI-CD)

The WPAI-CD consists of 6 questions asking about the effect of CD on the subject's ability to work and perform regular activities.
The percent work time missed due to problem (absenteeism), percent impairment while working due to problem (presenteeism), percent overall work impairment due to problem, and percent activity impairment due to problem are derived on the eCRF directly. Details about derivations can be found in APPENDIX 7.

16.1.6. Patient Global Impression of Change (PGIC)

The PGIC is a two-item scale designed to assess a patient impression of the overall change in CD symptoms and whether the change in CD symptoms is meaningful. Questions include a 7-point Likert scale based on current CD symptoms. The higher the score, the worse the symptoms. The number and percent of subjects in each PGIC category will be summarized by visit in the FAS for the Induction Analysis and Efficacy Evaluable Set for the Extension Analysis. The number and percent of subjects with meaningful change (as reported by the subject) differences will also be summarized.

16.1.7. CD-Related Hospitalizations and Surgeries

The CD-related hospitalizations are captured on the Hospitalization for Crohn's Disease eCRF. The CD-

related surgeries, including colectomy, are captured on the Concomitant Non-Drug Treatment eCRF. The number and percent of subjects with at least 1 CD related hospitalization or surgery along with the total duration will be summarized by treatment group using the FAS for the Induction Analysis and Efficacy Evaluable Set for the Extension Analysis. Total duration will be calculated as the sum of the hospitalization discharge date – hospitalization admission date + 1.

16.2. MISSING DATA METHODS

No missing data will be imputed for HRQoL endpoints.

17. SAFETY OUTCOMES

All outputs for safety outcomes will be based on the Safety Set. There will be no statistical comparisons between the treatment groups for safety data.
For the Final Analysis, summaries will be presented for the Extension Period treatment (etrasimod 2 mg to 2 mg, etrasimod 2 mg to 3 mg, and etrasimod 3 mg to 3 mg) and Induction Period treatment (etrasimod 2 mg and 3 mg).

17.1. ADVERSE EVENTS

AEs will be coded using MedDRA (version 25.1). Treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first dose of study treatment. TEAEs that started or worsened in severity before the first dose of Extension Period will be counted to Induction Period. TEAEs that started or worsened in severity after or on the first dose of Extension Period will be counted to Extension Period.
See APPENDIX 2 for handling of partial dates for AEs for the purpose of assigning treatment-emergent flags. In the case where it is not possible to define an AE as treatment-emergent or not, the AE will be classified as treatment-emergent.

17.1.1. ALL TEAES

All TEAEs will be summarized by system order class (SOC) and preferred term (PT). This summary table and all other TEAE summaries by SOC and PT will be presented by descending frequency in the total etrasimod treatment group. All AEs (TEAEs and non-TEAEs) will be listed.

Additionally, a summary of exposure-adjusted incident rate TEAEs by SOC and PT will be presented. Exposure is defined as the sum of either time (year) from first dose to the onset of first such event for those who experienced the adverse event, or time from first dose to last participation for those who did

not experience the adverse event.

The occurrence of TEAEs in Week 1 will be compared to their occurrence after Week 1 across the treatment arms in the Induction Period and will be summarized by SOC and PT.

17.1.1.1. Severity

Severity is classified as Grade 1: Mild, Grade 2: Moderate, Grade 3: Severe, Grade 4: Life-Threatening, Grade 5: Death Related to AE, using the Common Terminology Criteria for Adverse Events, version 5.0 (CTCAE v5.0).

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

17.1.1.2. Relationship to Study Treatment

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

All related TEAEs will be summarized by SOC and PT. A "related TEAE" for the purpose of this summary is defined as a TEAE with relationship to study drug of "probably related" or "related".
All TEAEs will be summarized by SOC, PT, and highest relationship (as reported on the eCRF, not grouped), with SOC and PT presented alphabetically. If a subject reports a TEAE more than once within a SOC/PT, the AE with the worst-case relationship to study treatment will be used in this summary. TEAEs with a missing relationship to study treatment will be regarded as "related" to study treatment in the summary. Summaries will present all categories of relatedness for Induction Analysis.

17.1.2. TEAEs Leading to Discontinuation of Study Treatment

TEAEs leading to permanent discontinuation of study treatment will be identified by action taken being recorded as "Drug withdrawal" on the Adverse events eCRF.
All TEAEs leading to discontinuation of study treatment will be summarized by SOC and PT.

17.1.3. Serious and Non-Serious TEAEs

Serious adverse events (SAEs) are those events recorded as "Serious" on the Adverse Event eCRF.
All serious TEAEs will be summarized by SOC and PT. If the seriousness is missing, the AE will be considered as "Serious". A listing of serious TEAEs will also be presented. All non-serious TEAEs will also be summarized by SOC and PT. All TEAEs leading to temporary interruption of study treatment will

be summarized by SOC and PT for Extension Period.

17.1.4. TEAEs Leading to Death

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

TEAEs leading to death will be summarized by SOC and PT.

17.1.5. TEAEs of Special Interest

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

- Cardiovascular Events
  - Bradycardia
  - AV conduction delay
  - Increased Blood Pressure
- Macular Edema
- Pulmonary Disorders
  - Airflow obstruction (forced expiratory volume in 1 second [FEV_1], 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

17.1.6. Overall Summary of Adverse Events

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

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

* 'Related' TEAEs refers to TEAEs related or probably related to study drug.

In addition to AE summaries by Induction Period and Extension Period, 4 AE tables will be generated for overall study:

1. Summary of Treatment-Emergent Adverse Events (Safety Set);
2. Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Set);
3. Treatment-Emergent Adverse Events by System Organ Class and Preferred Term - Exposure Adjusted Incidence Rate (Safety Set);
4. Related Treatment-Emergent Adverse Events by System Organ Class and Preferred Term (Safety Set)

17.2. Deaths

Deaths during the study (e.g., date of death, primary cause of death) will be presented in a data listing.

17.3. Laboratory Evaluations

17.3.1. Safety Laboratory Evaluations

Hematology, serum chemistry, coagulation, and urinalysis are analyzed and reported by central laboratory and sometimes by local laboratory. Results out of reference range are flagged by the performing laboratory (e.g., low, high).

Presentations will use SI Units. Quantitative laboratory measurements reported as "< X" or "> X", where X may be the lower limit of quantification (LLQ) or the upper limit of quantification (ULQ), respectively, will be converted to X for the purpose of quantitative summaries, but will be presented as recorded, i.e. as "< X" or "> X" in the listings. For urinalysis, only pH and specific gravity are considered as quantitative tests. If local and central laboratory assessments are available from the same day, central laboratory assessments will be used in the summary as per Section 6.4.

The following summaries will be provided for laboratory data:

- Value and change from baseline by visit (for hematology, serum chemistry, quantitative urinalysis [pH and specific gravity], and coagulation)
    - Incidence of abnormal values according to laboratory reference ranges by visit
- Incidence of lymphocytes < 0.2 x 10^9/L, < 0.5 x 10^9/L, or neutrophils < 0.5 x 10^9/L, < 1 x 10^9/L at end of treatment in Induction Period and anytime within this period
- Shift from baseline to End of Treatment reference range category (for quantitative measurements and categorical measurements), for Induction Analysis only
- eDISH plots for the following laboratory assessments (using values after the first administration of study treatment):
    - Maximum Aspartate Transaminase (AST) vs Same-Day Total Bilirubin
    - Maximum Alanine Aminotransferase (ALT) vs Same-Day Total Bilirubin
    - Maximum Alkaline Phosphatase (ALP) vs Same-Day Total Bilirubin (Induction Analysis only)

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/ALP assessment occurs at two 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 upper limit of normal, and the other one showing values as multiples of upper limit of normal, or subject's baseline, whichever is higher.

- Incidence of hepatic enzyme elevations at baseline and post baseline
    - > 1, ≥ 2, 3, 5x ULN elevation in ALT (8, 10, 20 x ULN elevation will also be presented for Induction Analysis, not for Final Analysis)
    - > 1, ≥ 2, 3, 5x ULN elevation in AST (8, 10, 20 x ULN elevation will also be presented for Induction Analysis, not for Final Analysis)
    - > 3, 5, 10, 20 x ULN elevation in either ALT or AST (Induction Analysis only)
    - > 1, ≥ 2, 3 x ULN elevation in total bilirubin (1.5 x ULN elevation will also be presented for Induction Analysis, 5 x ULN elevation will also be summarized for the Extension Period)

- > 1, ≥ 2, 3, 5 x ULN elevation in ALP (1.5, 8 x ULN elevation will also be presented for Induction Analysis, not for Final Analysis)
- > 3 x ULN elevation in either ALT or AST and > 1.5 x ULN elevation in total bilirubin
- > 3 x ULN elevation in either ALT or AST and > 2 x ULN elevation in total bilirubin
- > 3 x ULN elevation in either ALT or AST and > 1.5 x ULN elevation in ALP
- ≥ 3 x ULN in ALT or AST and ≥ 2 x ULN in Bilirubin and <2 x ULN in ALP (Hy's Law)

- Confirmed incidence of hepatic enzyme elevations at postbaseline will be generated as well for ALT, AST, ALP and Bilirubin. A confirmed case is defined as ≥2 consecutive postbaseline assessment values meet the criteria.
- Absolute and percent change from baseline in mean TBNK counts by visit will be presented as scatter plots with the standard errors for the mean for Induction Period only
- Summary statistics for value and change from baseline in Lipid panel results by visit

Subject's laboratory assessments, central or local, at all timepoints in chronological order. Values outside of the laboratory reference range will be flagged. Values obtained from local laboratory will be flagged. Listing of lymphocytes and neutrophils over time in subjects with lymphocytes < 0.5 x 10^9/L or neutrophils < 1 x 10^9/L will also be provided for the Induction Analysis only.

17.3.2. Pregnancy Tests

Urine β-hCG and/or serum β-hCG pregnancy tests are performed in female subjects of childbearing potential. All pregnancy test results in subjects will be listed in chronological order for the Induction Analysis only.

17.3.3. Other Screening Laboratory Assessments

Screening laboratory assessments for virology, drug screen/toxicology, QuantiFERON TB Gold, stool pathogens, *C. diff*, VZV IgG will be listed for the Induction Analysis only. Analyses of genetics and exploratory efficacy-related biomarkers data based on samples collected at Screening in subject who provided consent will be described in a separate plan.

17.4. ECG Evaluations and Holter Monitoring

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

- HR (bpm)

- PR interval (msec)
- RR interval (msec)
- QRS interval (msec)
- QT interval (msec)
- QTcB interval (msec)
- QTcF interval (msec)
- Overall interpretation of ECG (Investigator's judgment):
    - Normal
    - Abnormal, Not Clinically Significant (Abnormal NCS)
    - Abnormal, Clinically Significant (Abnormal CS)
- Overall interpretation of ECG (central reader):
    - Normal
    - Abnormal, Not Clinically Significant (Abnormal NCS)
    - Abnormal, Clinically Significant (Abnormal CS)

The following summaries will be provided for ECG data:

- Value and change from baseline by visit (for quantitative measurements)
- Value and change from Baseline (predose) on Day 1 and Day 2 (in subjects with extended cardiac monitoring) (for quantitative measurements)
    - Incidence of markedly abnormal values by visit
- Shift in normal/abnormal NCS/abnormal CS in the overall interpretation from baseline to end of induction treatment and to the worst-case post-baseline (Induction Analysis only)
- Shift in markedly abnormal categories from baseline to post-baseline by visit for Induction Analysis only

All ECG results, including first dose cardiac monitoring, will be listed. Subjects not satisfying the discharge criteria for first dose cardiac monitoring will also be listed (Table 18 in the Protocol). A listing of all ECG assessments over time in subjects meeting markedly abnormal criteria will also be provided. For each subject, only ECG parameters ever meeting markedly abnormal criteria will be included.

Holter monitoring assessments will be collected to assess ectopic ventricular beats, pauses, supraventricular tachycardia, atrioventricular block and other important arrhythmias. Descriptive statistics for hourly data and summary data will be presented. If there are more than 10 Holter events, then descriptive statistics for events will be included as well. Holter monitoring data will be listed separately for hourly, summary and event assessments. If there is more than 1 assessment with the same date and time, the assessment from the central read will be used in analyses.

17.4.1. ECG MARKEDLY ABNORMAL CRITERIA

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

- Absolute values in QT, QTcF and PR:
  - ≥ 450 msec (male) or ≥ 470 msec (female) in QTcF
  - > 500 msec in QT
  - > 230 msec in PR
- Change from baseline in QT and QTcF:
  - > 30 msec increase from baseline
  - > 60 msec increase from baseline

In shift tables, subjects will be classified according to each binary category (i.e., Markedly abnormal vs. Not markedly abnormal, with markedly abnormal defined in the footnote).

17.5. VITAL SIGNS

The following vital signs measurements (and unit) are reported for this study:

- Systolic blood pressure (mmHg)
- Diastolic blood pressure (mmHg)
- Heart rate (bpm)
- Respiratory rate (resp/min)
- Temperature (^{o}C)
- Weight (kg)
- Height (cm) [Screening only]

The following summaries will be provided for vital signs data:

- Value and change from baseline by visit
  - Value and Change from Predose on Day 1 and Day 2 (in subjects with extended cardiac monitoring) (Induction analysis only)
- Incidence of markedly abnormal values by visit
- Incidence of minimum heart rate on Day 1 and Day 2 (Induction analysis only)

All vital signs, including first dose cardiac monitoring, will be listed. Subjects not satisfying the discharge criteria for first dose cardiac monitoring will also be listed (Table 18 in the Protocol). A listing of all vital signs assessments over time in subjects meeting markedly abnormal criteria will also be provided. For each subject, only vital sign parameters ever meeting markedly abnormal criteria will be included.

17.5.1. VITAL SIGNS SPECIFIC DERIVATIONS

- Temperature (^{o}C) = (5/9) (Temperature (^{o}F) – 32)

17.5.1.1. Vital Signs Markedly Abnormal Criteria

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

Variable	Unit	Low	High
SBP	mmHg	≤ 90 mmHg	> 150 mmHg
DBP	mmHg	≤ 50 mmHg	> 90 mmHg
Heart rate	bpm	< 40 bpm < 50 bpm < 50 bpm and decrease from predose (baseline) of > 10 bpm at 4 hour postdose on Day 1 or Day 2	> 100 bpm or

17.6. PHYSICAL EXAMINATION

All physical examination assessments will be listed for subjects who had at least 1 abnormal assessment, for the Induction Analysis only.
All abdominal mass assessments will be listed, for Induction Analysis only.

17.7. OTHER SAFETY ASSESSMENTS

17.7.1. PULMONARY FUNCTION TESTS (PFTS)

The following PFT measurements will be reported for this study:

- Forced Expiratory Volume at a second (FEV_1), including actual, Predicted and % Predicted
- Forced Vital Capacity (FVC), including actual, Predicted and % Predicted
- FEV_1/FVC, including actual, Predicted and % Predicted
- Forced Expiratory Flow (FEF) 25 – 75, included actual, Predicted and % Predicted
- Total Lung Capacity (TLC) (if available), including actual, Predicted and % Predicted
- DLCO (if available), including hemoglobin adjusted DLCO

The following summaries will be provided for PFT data:

- Value and change from baseline by visit
    - Incidence of markedly abnormal values by visit
    -

Due to carbon monoxide binding to hemoglobin, DLCO values will be corrected for hemoglobin

concentration. The following conversion will be used for DLCO analyses:

$DLCO_{cor} = DLCO_{pred} *[1.7*Hgb/(AS + Hgb)]$

where $DLCO_{cor}$ is the corrected DLCO value, $DLCO_{pred}$ is the Predicted DLCO value collected on the Pulmonary Function Test and DLCO eCRF page, Hgb is the hemoglobin value collected on the same date as the DLCO assessment, and AS is the Age-Sex-Factor (9.38 for Females, 10.22 for Males).

All PFT data will be listed. A listing of subjects with at least 1 abnormality in PFT will also be provided, including a flag for whether markedly abnormal criterion is met. Frequencies and percentages of patients with a clinically notable decrease in PFT values will be summarized.

17.7.1.1. PFT Markedly Abnormal 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 < 50%
    - Potentially important PFT measurements will also be identified using the criteria below:
- Decrease from baseline > 20% in FEV1
- Decrease from baseline > 20% in FVC
- Decrease from baseline > 20% in DLCO

17.7.2. Ophthalmoscopy and Optical Coherence Tomography (OCT)

All ophthalmoscopy and OCT assessments will be listed. A listing of subjects with at least 1 abnormality in OCT will be provided. Visual acuity and central foveal thickness as measured by OCT will be summarized by presenting summary and descriptive statistics of observed data and change from baseline values.

17.7.3. Tuberculosis Screening/Questionnaire and Chest X-Ray

Abnormal screening tuberculosis results and chest X-rays will be listed (Induction analysis only).
All tuberculosis questionnaire assessments will be listed (Induction analysis only).

17.7.4. ENDOSCOPY

18. ALL ENDOSCOPY BIOPSY ASSESSMENT RESULTS WILL BE LISTED. PHARMACOKINETICS

Plasma concentrations of etrasimod, will be assessed from samples collected prior to dosing and 4 hours (± 15 minutes) postdose (after ECG) on Day 1. Additionally, PK samples will be collected prior to dosing (trough) at Weeks 1, 2, 6, and 14 of the Induction Period, and at Weeks 28 and 52 of the Extension Period. Additionally, samples will be collected at early termination and at both visits of the 4-Week Follow-up period (at FU1 and FU2).

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 Population Modeling Analysis Plan (PMAP).

19. DATA NOT SUMMARIZED OR PRESENTED

The other variables and/or domains not summarized or presented are:

- Comments

These domains and/or variables will not be summarized or presented but will be available in the clinical study database, SDTM and/or ADaM datasets.

20. REFERENCES

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

EQ-5D-5L User Guide Version 3.0. September 2019: https://euroqol.org/wp-content/uploads/2019/09/EQ-5D-5L-English-User-Guide_version-3.0-Sept-2019-secured.pdf

ICH E3. Structure and Content of Clinical Study Reports. Dated 30 November 1995. Available at: 7 (accessed 14 August 2018).

ICH E9. Statistical Principles For Clinical Trials. Dated 5 February 1995. Available at: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf (accessed 14 August 2018).

ICH E9 R(1). Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. Dated 20 November 2019. Available at: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

Rubin, D.B. (1987). *Multiple Imputations for Nonresponse in Surveys*. Wiley.

Selinger C, Sandborn W, Panes J, et al. Etrolizumab as induction therapy in moderate to severe Crohn's disease: Results from bergamot cohort 1. *Gut.* 2018;67(1):A53.

U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH), & Center for Veterinary Medicine (CVM). (2020, June). *Statistical Considerations for Clinical Trials During the COVID-19 Public Health Emergency Guidance for Industry*. U.S Food & Drug Administration. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-considerations-clinical-trials-during-covid-19-public-health-emergency-guidance-industry

APPENDIX 1. PROGRAMMING CONVENTIONS FOR OUTPUTS

OUTPUT CONVENTIONS

Outputs will be presented as shown in the Output shells.

DECIMALS AND PERCENTAGES

- If the original data has N decimal places, then the summary statistics should have the following decimal places:
  - Minimum and maximum: N
  - Mean, median: N + 1
  - SD: N + 2
- Percentages and confidence intervals will be reported to 1 decimal place. Where counts are zero, percentages will not appear in the output.
- SES-CD and CDAI scores will display:
  - 1 decimal place for mean, median
  - 2 decimal places for SD and SE
- SES-CD will display 0 decimal places for min, max
- CDAI will display 1 decimal places for min, max
- Kaplan-Meier estimates will be reported to 2 decimal places.
  -

DATES AND TIMES

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

SPELLING FORMAT

English US.

PRESENTATION OF TREATMENT GROUPS

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

Treatment Group	For Tables, Graphs and Listings
Etrasimod 2 mg	Etrasimod 2 mg
Etrasimod 3 mg	Etrasimod 3 mg
Screen Failure	Screen Failure
Not Treated*	Not Treated

*To be used for subjects in safety listings who are randomized but do not receive study drug.

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

- For Demographics and Safety analysis
  - Induction 2 mg and Extension 2 mg
  - Induction 2 mg and Extension 3 mg
  - Induction 3 mg and Extension 3 mg
  - Total subjects with 3mg at Extension
  - Overall total
- For Efficacy and Biomarker Analysis
  - Induction 2 mg and Extension 2 mg (Week 14 responders)
  - Induction 2 mg and Extension 2 mg (Week 14 non-responders)
  - Induction 2 mg and Extension 3 mg
  - Induction 3 mg and Extension 3 mg
- For overall AE and exposure tables:
  - Induction 2 mg and Extension 2 mg (including subjects discontinued in Induction Period)
  - Induction 2 mg and Extension 3 mg
  - Induction 3 mg and Extension 3 mg (including subjects discontinued in Induction Period)
  - Overall total

Treatment groups will be presented by Extension Treatment, Induction Treatment and Week 14 Clinical Response. Summaries will also include any Etrasimod treatment by dose and overall.

PRESENTATION OF VISITS AND STUDY PERIOD

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

Visit Name	Study Period
Screening	Screening
Baseline, Day 1, Day 2 (VS/ECG only), Week 1, Week 2, Week 4, Week 6, Week 10, Week 14	Treatment
Early Termination	Dependent on Analysis Visit assigned

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

Visit Name	Study Period
Baseline, Day 1, Day 2 (VS/ECG only), Week 1, Week 2, Week 4, Week 6, Week 10, Week 14, Week 15 and End of Week 15 (VS/ECG/Holter monitoring only), Week 20, Week 28, Week 36, Week 44, Week 52, Week 66	Treatment
Early Termination	Dependent on Analysis Visit assigned
2-Week Follow-up, 4-Week Follow-up	Follow-up

LISTINGS

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

- Induction analysis: Randomized treatment group (or treatment received if it's a safety output), first by Etrasimod 2 mg, then Etrasimod 3 mg, then Screen Failure and then Not Treated (only in safety listings if there are any randomized subjects who did not receive study drug)
- Final analysis: Randomized treatment group (or treatment received if it's a safety output) in order of :
    - Induction 2 mg and Extension 2 mg
    - Induction 2 mg and Extension 3 mg
    - Induction 3 mg and Extension 3 mg
    - Induction Placebo and Extension X mg
- Subject number (which is expected to incorporate study site/center)
- Date (where applicable)
- Visit (where applicable) – Note: if assessments are missing a date (due to being Not Done), the assessment should be ordered so that the visit is in chronological order

    - CONVENTIONS RELATED TO PHARMACOKINETIC DATA
- All PK concentrations will be reported and analyzed with the same precision as the source data provided by the bio-analytical 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.

APPENDIX 2. PARTIAL DATE CONVENTIONS

Imputed dates will NOT be presented in the listings.

ALGORITHM FOR TREATMENT EMERGENCE OF ADVERSE EVENTS:

AE START DATE	AE STOP DATE	ACTION
Known	Known, Partial or Missing	If AE start date < study drug first dose date, then not TEAE If AE start date ≥ study drug first dose date, then TEAE
Partial, but known components show that it	Known, Partial or Missing	Not TEAE

AE START DATE	AE STOP DATE	ACTION
cannot be on or after date of first dose of study drug		
Partial, could be on or after date of first dose of study drug	Known	If AE stop date < study drug first dose date, then not TEAE If AE stop date ≥ study drug first dose date, then TEAE
	Partial	Impute AE stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < study drug first dose date, then not TEAE If AE stop date ≥ study drug first dose date, then TEAE
	Missing	Assumed TEAE
Missing	Known	If AE stop date < study drug first dose date, then not TEAE If AE stop date ≥ study drug first dose date, then TEAE
	Partial	Impute AE stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If AE stop date < study drug first dose date, then not TEAE If AE stop date ≥ study drug first dose date, then TEAE
	Missing	Assumed TEAE

ALGORITHM FOR PRIOR AND CONCOMITANT MEDICATIONS:

START DATE	STOP DATE	ACTION
Known	Known	If medication stop date < study med first dose date, assign as prior If medication stop date ≥ study med first dose date, assign as concomitant
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study med first dose date, assign as prior If medication stop date ≥ study med first dose date, assign as concomitant

START DATE	STOP DATE	ACTION
	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 (i.e. first day of month if day unknown or 1st January if day and month are unknown), then: If medication stop date < study med first dose date, assign as prior If medication stop date ≥ study med first dose date, assign as concomitant
	Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown) and impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study med first dose date, assign as prior If medication stop date ≥ study med first dose date, assign as concomitant
	Missing	Impute start date as earliest possible date (i.e. 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 med first dose date, assign as prior Else assign as concomitant
	Partial	Impute stop date as latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown), then: If medication stop date < study med first dose date, assign as prior If medication stop date ≥ study med first dose date, assign as concomitant
	Missing	Impute start date to be informed consent date. Assign as concomitant

ALGORITHM FOR CROHN'S DISEASE DIAGNOSIS DATE:

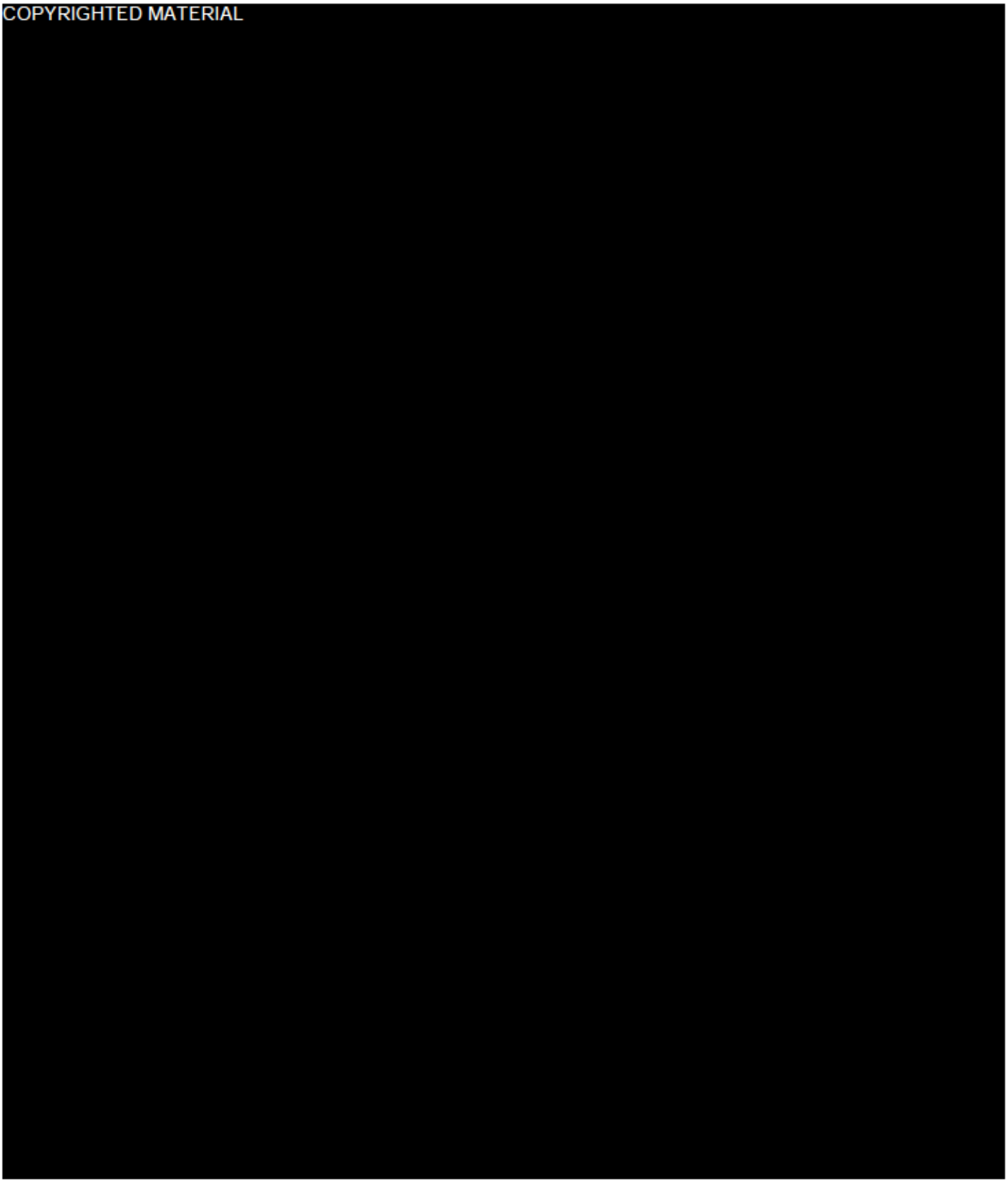
START DATE	ACTION
Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown)

ALGORITHM FOR CONCOMITANT START/STOP DATES:

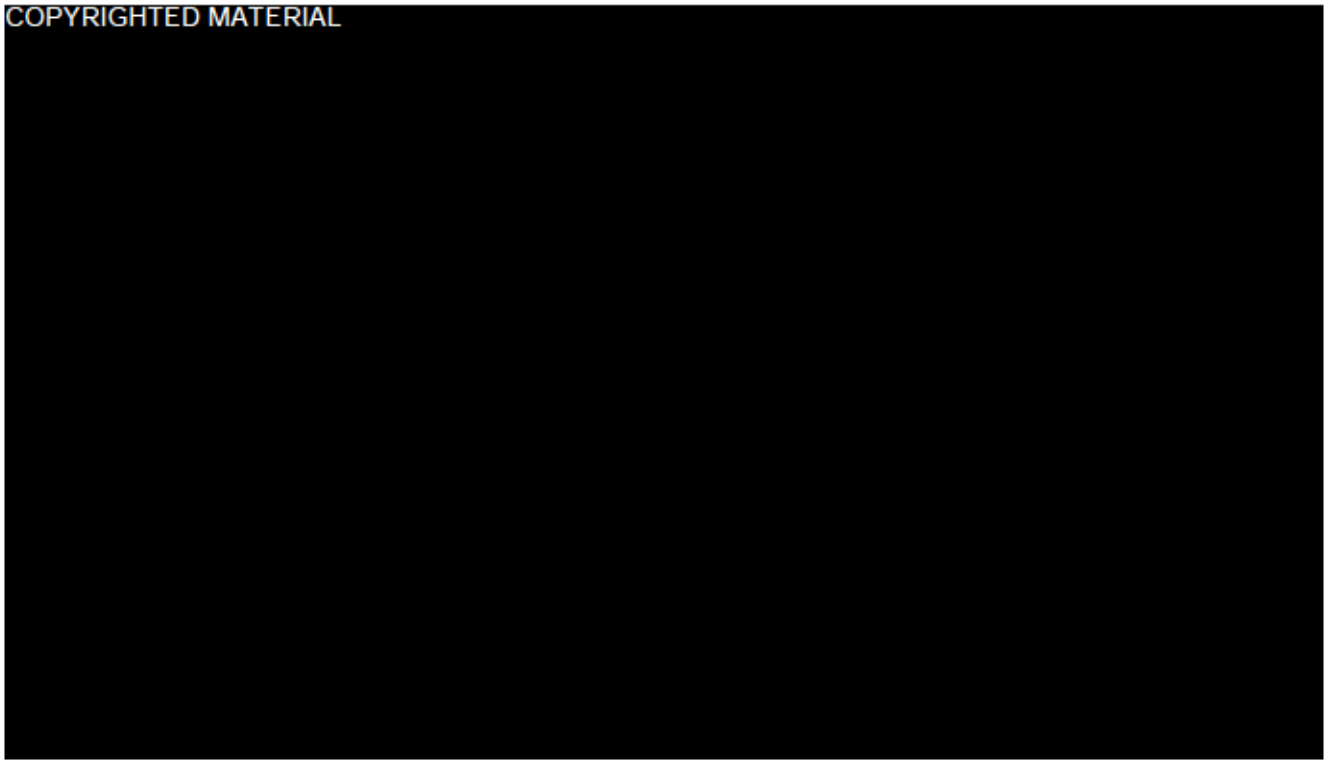
START/STOP DATE	ACTION
Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown) Impute to latest possible date (i.e. last day of month if day unknown or 31st December if day and month are unknown)
Missing	Impute medication start date to informed consent date. Impute medication end date to last visit date.

APPENDIX 3. SIMPLE ENDOSCOPIC SCORE IN CROHN'S DISEASE

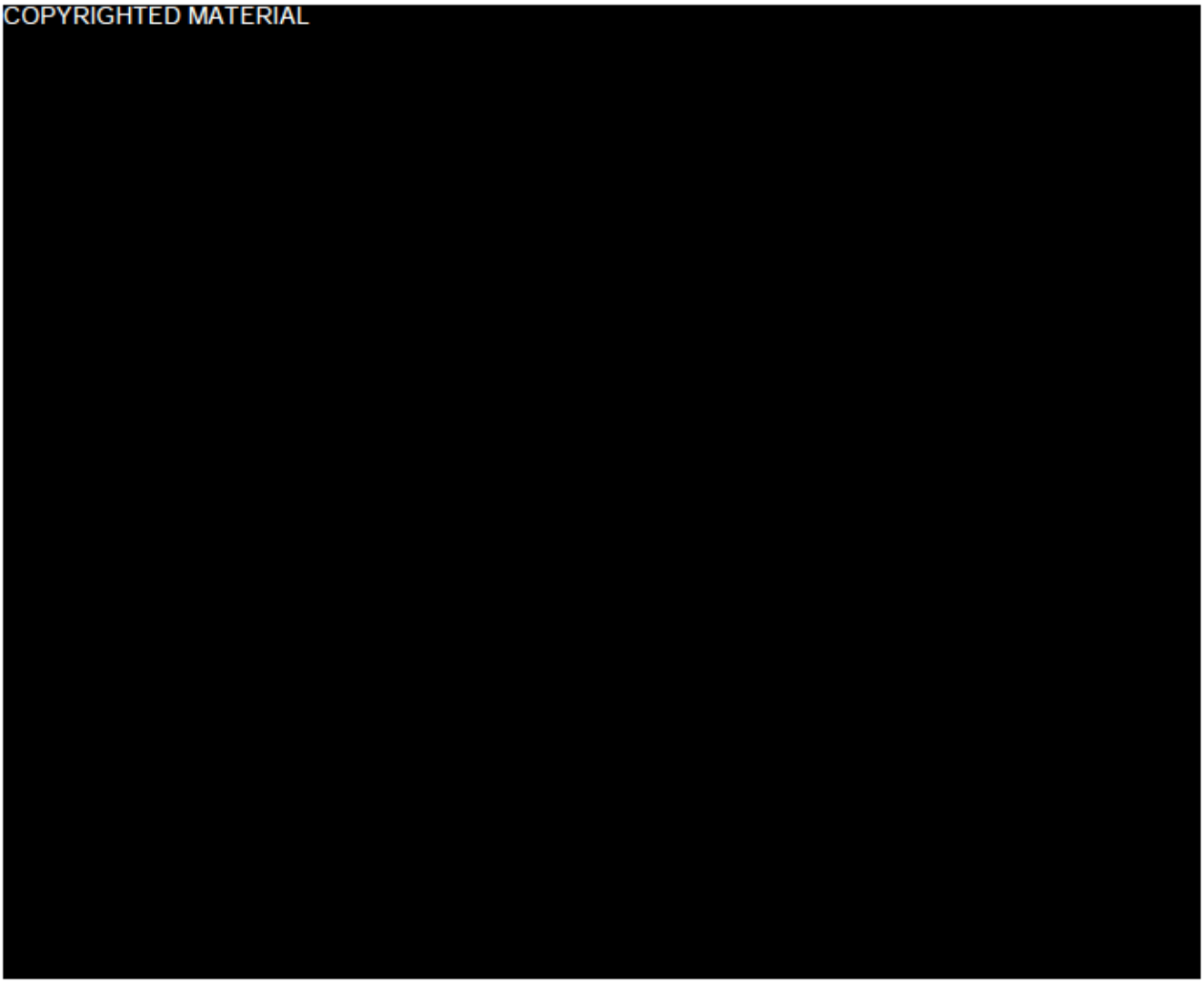
APPENDIX 4. CDAI VARIABLES AND WEIGHTING FACTORS



APPENDIX 5. IBDQ SCORING RULES



APPENDIX 6. SF-36 SCORING RULES



APPENDIX 7. WPAI-CD SCORING RULES

The WPAI-CD 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 CD?
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 CD affect your productivity while you were working?
6 = During the past seven days, how much did your CD 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.