STATISTICAL ANALYSIS PLAN

SUBSTUDY 1 – PHASE 2B Cohort 1

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: **V1.0, 09AUG2023**

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Statistical Analysis Plan V1.0 (Dated 09AUG2023) for Protocol C5041006 (APD334-202) (Substudy 1 – Phase 2b).

	Name	Signature	Date
Author:	PPD	See below	09-Aug-2023
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	10-Aug-2023
Position:	PPD		
Company:	IQVIA		
Approved By:	PPD, PhD	See below	09-Aug-2023
Position:	PPD		
Company:	Pfizer, Inc.		

MODIFICATION HISTORY

Version	Date of the Document Version	Author	Significant Changes from Previous Authorized Version
1.0	09AUG2023	PPD	Not applicable – first version

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 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 will focus on Substudy 1 – Phase 2b (SS1-P2b) Cohort 1.

Note, the Phase 2 substudy (Substudy A – Phase 2 [SSA-P2]) 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 have 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 dose-response relationship of 2 doses of etrasimod versus placebo as induction therapy in subjects with moderately to severely active Crohn's disease (CD)
- To select an oral etrasimod dose, based on efficacy and safety, for continued development

2.2. SECONDARY OBJECTIVES

- To evaluate the safety, tolerability, and efficacy of etrasimod in subjects with moderately to severely active CD
- To provide a subset of the target study population, etrasimod responders, to be evaluated in Substudy 3 - Maintenance

2.3. ESTIMANDS

Table 1 describes the primary analysis for the primary and secondary efficacy estimands for SS1-P2b.

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 or discontinued due to disease worsening will be treated as nonresponders on the event date and subsequent visits.	The difference in proportion of responders between etrasimod and placebo at Week 14
Secondary Endpoint	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 or discontinue due to disease worsening will be treated as nonresponders on the event date and subsequent visits.	The difference in the proportion of responders between etrasimod and placebo at Week 14
Secondary Endpoint	Efficacy of etrasimod on clinical remission by Patient-Reported Outcome 2 (PRO2) at Week 14	FAS	PRO2 < 8 at Week 14	Subjects who have received the protocol defined prohibited medications or discontinue due to disease worsening will be treated as nonresponders on the event date and subsequent visits.	The difference in the proportion of responders between etrasimod and placebo at Week 14

3. STUDY DESIGN

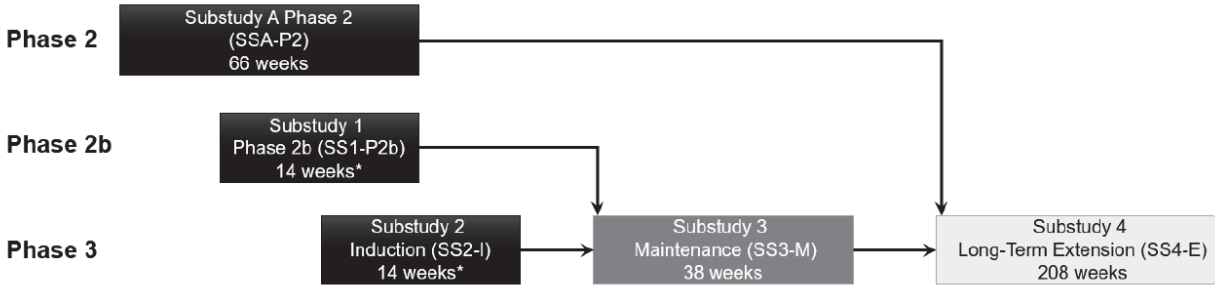
3.1. GENERAL DESCRIPTION

C5041006 (APD334-202 or APD334-202EU) is a seamless Phase 2/3, multicenter, randomized, double-blind, study that comprises of 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 at a stable dose (unless change required due to an adverse event) as a background therapy for CD.

Overall Study: Overall duration of the study is up to 282 weeks, including a 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 2b substudy.

The main focus of this SAP is Substudy 1 – Phase 2b 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

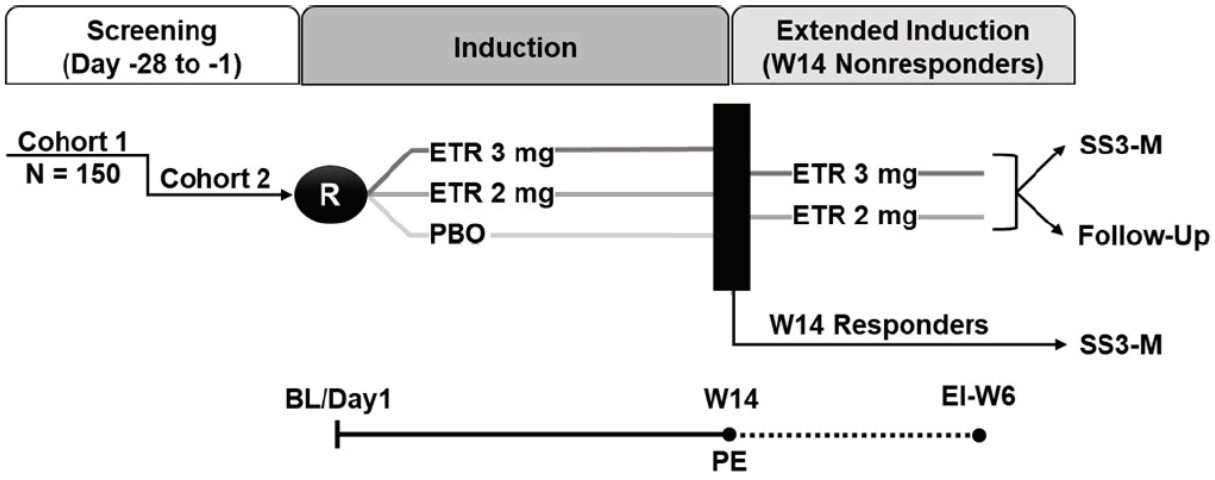
Substudy 1 - Phase 2b (SS1-P2b): As shown in Figure 1, SS1-P2b Cohort 1 is a Phase 2b randomized, double-blind, placebo-controlled, dose-ranging induction substudy to evaluate etrasimod as induction therapy and select an induction and maintenance dose(s) for continued evaluation in Phase 3. SS1-P2b Cohort 2 is a randomized, double-blind, placebo-controlled induction substudy to contribute to SS2-I

study to evaluate the selected etrasimod dose as induction therapy.

SS1-P2b will consist of 2 cohorts that will enroll consecutively. Cohort 1 will enroll and treat approximately 150 subjects. Efficacy and safety data from this cohort will be analyzed to select the dose(s) for continued evaluation in Phase 3 studies. To prevent a break in the enrollment of the study, it is expected approximately 75 subjects will be enrolled in Cohort 2 and treated while data for Cohort 1 are analyzed for dose selection. The actual number will depend on the enrollment rate and time needed to select the dose for further development.

Subjects will be randomized in a double-blinded fashion (1:1:1 ratio), with about 50 subjects expected in each of the treatment groups (etrasimod 2 mg, etrasimod 3 mg, or placebo) in Cohort 1 and about 25 subjects expected in each of the treatment groups in Cohort 2. Randomization will be stratified by biologic treatment failure status (yes/no) and baseline oral corticosteroid use (yes/no). Total duration is up to approximately 30 weeks, including a 40-Day Screening Period, 14-week double-blind Induction Period, a subsequent 6-week Extended Induction (EI) Period, and a 4-week Follow-up Period (for subjects who do not continue in SS3-M). Subjects who meet the response criteria at Week 14 will be eligible to SS3-M after completion of the Week 14 visit. Subjects who do not meet the response criteria at Week 14 will enter a 6-week EI Period for a total of up to 20 weeks induction treatment. Placebo subjects will receive either etrasimod 2 or 3 mg in the EI Period. At the end of the EI Period, subjects who show clinical improvement, based on the Investigator's judgement (globally), may be eligible to enter SS3-M; subjects who show no clinical improvement will be discontinued from the study. For EU region only, subjects with CD worsening, defined as ≥ 100-point increase in CDAI score from baseline and CDAI ≥ 220 points, will be discontinued from the study, subjects without CD worsening may be eligible for SS3-M. A schematic diagram of SS1-P2b is shown in Figure 2.

Figure 2: Substudy 1 – Phase 2b Schematic Diagram



BL, Baseline; EI, Extended Induction; ETR, etrasimod; PBO, placebo; PE, primary endpoint; R, randomization; SS3-M, Substudy 3 – Maintenance; W, Week

3.2. SCHEDULE OF ASSESSMENTS

The schedule of assessments for SS1-P2b can be found in Table 30 in Appendix 1 of the protocol.

3.3. CHANGES TO ANALYSIS FROM PROTOCOL

- The subgroups Disease localization (ileum only, colon only, both), baseline SES-CD score (≤ or > Median) and baseline fecal calprotectin (≤ or > Median) are not included in the protocol but will be summarized.
- Clarification of hypothesis test method and confidence interval method for common risk differences.
- For primary endpoint and secondary endpoints, multiple imputation becomes main method for missing handling and NRI becomes the alternative method.
- Multiple comparison adjustment was removed.

4. PLANNED ANALYSES

The following analyses will be performed for this study:

- Induction analysis after all randomized subjects from Cohort 1 have either completed Week 14 or discontinued study drug treatment
- Extended induction analysis after all subjects have completed protocol specified treatment including extended induction period or discontinued treatment and had the safety follow-up

Results from unplanned analyses that are included in the clinical study report (CSR) will be documented in the CSR section of Change from Planned Analysis. Summaries from the Induction analysis may be provided to applicable regulatory authorities as required per national regulatory requirements and commitments. The CSR for Cohort 1 will be developed for the induction analysis and extended induction analysis after all Cohort 1 subjects have completed protocol specified treatment including EI period or discontinued treatment and the database is unblinded for Cohort 1. Cohort 2 will remain blinded until the database lock for SS2-I to avoid potential bias for SS2-I analysis. There is no hypothesis testing for Cohort 2 alone. Cohort 2 data of placebo and the selected etrasimod dose group for SS2-I study will be combined with data from SS2-I in testing the hypotheses for the primary and key secondary efficacy endpoints listed in SS2-I. Descriptive summary for Cohort 2 data will be described in the final CSR for SS1-P2b.

4.1. INTERIM ANALYSIS

No formal interim analyses are planned for this substudy.

4.2. INDUCTION ANALYSIS FOR COHORT 1

The induction analysis will occur after all subjects from Cohort 1 complete 14 weeks of double-blind induction treatment or discontinued treatment. The objective of the induction analysis is to determine the appropriate etrasimod dose(s) for Phase 3 trials.

All planned analyses identified in this SAP will be performed following Pfizer's Authorization of the SAP.

After outstanding data queries for the induction period of cohort 1 have been resolved/closed, and the data have been cleaned and finalized, Pfizer will authorize breaking of the study blind for cohort 1 and the analysis of the data will be performed. The SAP must be approved and signed by Pfizer before the Induction Database Lock and treatment assignment is unblinded.

Only Cohort 1 subjects will be unblinded during the Induction analysis. Sponsor and investigators will remain blinded to Cohort 2 treatments until database lock for Substudy 2 (SS2-I). Cohort 2 data will be analyzed in combination with data collected at SS2-I for Phase 3 induction studies for efficacy, safety, and PK analysis. Data from Cohort 2 will therefore not be included in SS1-P2b analysis datasets nor be presented in analysis Tables, Graphs and Listings.

4.3. EXTENDED INDUCTION ANALYSIS FOR COHORT 1

The extended induction analysis of SS1-P2b will occur when all subjects enrolled in Cohort 1 have completed the protocol specified treatment including the extended induction period or discontinued treatment and had the safety follow-up or entered SS3-M. The extended induction analysis is mainly for extended induction period. Some AE and exposure tables for all Cohort 1 subjects and overall study duration will be generated as well.

After outstanding data queries in cohort 1 have been resolved/closed, and the data have been cleaned and finalized, Pfizer will authorize breaking of the study blind for the extended induction period in cohort 1 and the analysis of the data will be performed. The SAP must be approved and signed by Pfizer before the Extended Induction Database Lock and Extended Induction Period treatment assignment is unblinded.

There is no hypothesis testing for extended induction analysis, only descriptive summary.

4.4. ANALYSIS FOR COHORT 2

Data from Cohort 2 will not be combined with Cohort 1 data and nor will it be analysed alone. The data from Cohort 2 will be combined with SS2-I for the 2 common treatment groups (selected etrasimod dose group and placebo group) and analysed at database lock of SS2-I. Cohort 2 data will be descriptively summarized and presented in SS1-P2b CSR.

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 Extended Induction Analysis.

5.1. PROCESS FOR ANALYSIS SET ASSIGNMENT

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

After the database is locked, Pfizer 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 Extended Induction) will be unblinded.

5.2. SCREENED SET

For each cohort, the Screened Set will contain all subjects who provide informed consent for this substudy in the corresponding cohort.

5.3. RANDOMIZED SET

For each cohort, the Randomized Set will contain all subjects in the Screened Set who were randomized to the study treatment in the corresponding cohort.

5.4. FULL ANALYSIS SET [FAS]

For each cohort, the FAS will include all randomized subjects who receive at least 1 dose of study

treatment in the corresponding cohort. The FAS will be used as the primary analysis set for primary and secondary efficacy endpoints.

5.5. MODIFIED FULL ANALYSIS SET [MFAS]

For each cohort, the mFAS will include all randomized subjects who receive at least 1 dose of study treatment in the corresponding cohort and have a baseline measurement and have at least 1 post-randomization measurement. 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.

5.6. PHARMACOKINETIC SET

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

5.7. SAFETY SET

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

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 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 1 – Phase 2b Visit Windows

Scheduled Study Visit[a] (Protocol Study Day)	**Analysis Visit Window[e] (Analysis Study Day)**
Baseline	≤ 1
Day 1	1
Day 2[b]	2
Week 1 (Day 7 ± 1)	2[b] to 11

Week 2 (Day 14 ± 3)	12 to 29
Week 6 (Day 42 ± 3)	30 to 57
Week 10 (Day 70 ± 3)	58 to 85
Week 14 (Day 98 ± 3)	86 to 130 (86 to EI First Dose Day for subjects who enter the Extended Induction Period)[c]
EI [d] – First Dose (Day 98 ± 3)	EI First Dose Day
EI – Week 1 (Day 105 ± 3)	EI First Dose Day +1 to EI First Dose Day +17
EI – Week 6 (Day 140 ± 3)	EI First Dose Day +18 to ET or EOS

[a] Visits will only be displayed for assessments that were expected to be collected at these visits, according to the Schedule of Assessments, Table 31, outlined in the Protocol.
[b] Applicable to subjects who require extended cardiac monitoring on Day 2 for vital signs and ECGs. If extended monitoring occurs, the analysis visit window for Week 1 will start on Day 3.
[c] Information based on CRF or Clinical Operation file.
[d] Extended induction visits will only occur for subjects who do not achieve the response criteria at Week 14 and enter Extended Induction period. EI – First Dose should occur within 1 week of the Week 14 visit but may be performed on the same day as the Week 14 visit. Due to an overlapping visit window with Week 14, a hierarchical algorithm will be used for the mapping. Week 14 will be mapped first, then assess records for the EI – First Dose mapping. If a scheduled Week 14 and EI – First Dose are both present in the same visit window, they will be mapped to separate visits.
[e] The upper limit of visit windows cannot exceed the subject's last day in the corresponding period.

Windowing will be applied prior to any missing data calculations. The last non-missing measurement taken prior to the date of first dose (including unscheduled assessments) will be labeled as "Baseline". Unscheduled or Screening visits that occurred before the baseline visit will not be assigned an analysis visit for summaries. 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.

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

Statistical hypothesis testing will be performed for Cohort 1 induction period only. Proportion-based analyses for the primary, secondary, and exploratory endpoints, will be analyzed using the Mantel-Haenszel (MH) weighted test, adjusted for randomization stratification factors. The common risk

difference and the two-sided 90% confidence interval will be obtained using the Mantel-Haenszel weighted method for the stratification. Continuous endpoints will be analyzed using a mixed model of repeated measures (MMRM) or an analysis of covariance (ANCOVA) model. Number of observations, mean, SD, median, 1st and 3rd quartiles, min, and max will be presented for all continuous endpoints descriptive summaries.

The default significance level will be CCI [redacted]; confidence intervals (CIs) will be 90% and all tests will be 2-sided, unless otherwise specified in the description of the analyses.

6.6. COMMON CALCULATIONS

For quantitative measurements:

- Change from baseline = Test Value at Visit X – Baseline Value
- Percent change from baseline = ((Test Value at Visit X – Baseline Value) / Baseline Value) * 100
- Proportion at Visit X = Number of subjects satisfying criteria at Visit X / Total number of subjects at Visit X
- IQR = Q3 (75th Percentile) – Q1 (25th Percentile)

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 Final database lock dates. All analyses will be conducted using SAS® (version 9.4, SAS Institute Inc., Cary, NC).

All data summarized will be presented in listings.

7. SAMPLE SIZE JUSTIFICATION

This substudy is powered to detect a dose-response relationship in the primary endpoint of endoscopic response at Week 14 based on pair-wise comparisons using 2-sample Fisher's exact test CCI [redacted] (Selinger 2018). With a 1:1:1 randomization ratio, a total of 150 evaluable subjects (50 per treatment group) in Cohort 1 will be needed CCI [redacted]. See Table 22 in the protocol for a range of testing powers to detect the treatment effect on different proportions of subjects achieving endoscopic response 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.

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 will 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 17.1.5, 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

As SS1-P2b Cohort 1 is a Phase 2, dose ranging sub-study to support a dose selection for Phase 3, there will be no multiplicity adjustment for the dose selection in this sub-study. All hypothesis testing will be performed at significance level of CCI [redacted].

7.5. EXAMINATION OF SUBGROUPS

The following subgroups will be assessed for the primary and secondary efficacy endpoints for the Induction Analysis (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)*
- Baseline oral corticosteroid usage (Yes or No)*
- Baseline fecal calprotectin (≤ or > Median)

* Stratification reported by the interactive web response system (IWRS) will be used

8. OUTPUT PRESENTATIONS

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

9. DISPOSITION AND WITHDRAWALS

For each cohort, 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 Extended Induction Period will be summarized. 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 will be presented for all screened subjects.

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

For each cohort, important protocol deviations will be summarized for the Safety Set. All protocol deviations will be listed.

- Inclusion Criteria
- Exclusion Criteria
- Concomitant Medication
- Randomization
- Laboratory Assessment
- Study Procedures
- Safety
- Efficacy
- Investigational Product conditions
- Investigational Product administration
- Subject Investigational Product compliance
- Subject Discontinuation
- Visit Schedule
- Administrative
- Blinding
- Patient Reported Outcomes
- PK/PD
- Other Criteria
- Informed Consent

10. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

The following demographic data and baseline characteristics will be summarized for the Safety Set.

- Age on consent (years) – including age groups <Median and ≥Median; <18, ≥18 to <65, and ≥65
- Sex
- Race
- Ethnicity
- Height (cm)
- Weight (kg)

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

The following baseline characteristics related to CD will be summarized:

- Duration of CD at consent (years, ≤ 5/ >5, ≤15/ >15)
- 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
- Baseline CRP
- Baseline FCP
- 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 prior to the first dose of study treatment and were not ongoing OR ended prior to or on the date of 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, Extended 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 Extended Induction 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). The reasons for dose interruption will be summarized.

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 Extended Induction 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 Period will be calculated as follows:

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

- Extended Induction Period will be calculated as follows:

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

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. For subjects who remain in the study beyond Week 14, the induction period duration will be defined as Date of Week 14 visit – Date of first dose (instead of Date of Week 14 visit – Date of first dose + 1), since the last dose in the induction period should have been taken on the day before the Week 14 visit.

15. EFFICACY OUTCOMES

For both the Induction analysis and extended induction analysis, Cohort 1 and Cohort 2 data will be analyzed or summarized separately. Hypothesis testing will be performed for cohort 1 induction period efficacy data only. Data from Cohort 2 will be combined and analyzed with data collected in SS2-I. Cohort 2 will remain blinded until the database lock for SS2-I to avoid potential bias for SS2-I analysis and all Cohort 2 data will be described in the final CSR for SS1-P2b.

Subjects will be summarized by the treatment to which they were randomized, regardless of treatment actually received.

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.
- Clinical response CDAI: Clinical remission CDAI or ≥ 100-point decrease from baseline in CDAI
- Clinical response CDAI-70: Clinical remission CDAI or ≥ 70-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
- Time to PRO2 response and FCP normalization: (PRO2 < 8 or ≥ 8-point decrease from baseline) and FCP ≤ 150 mg/Kg

**Indicates the response criteria for primary and secondary efficacy endpoints for SS1-P2b

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	
Primary endpoint			
Endoscopic Response	NRI; multiple imputation	OBS	FAS (NRI; multiple imputation)
Secondary endpoint			
Clinical remission CDAI	NRI; multiple imputation	OBS	FAS (NRI; multiple imputation)
Clinical remission PRO2	NRI; multiple imputation	OBS	FAS (NRI; multiple imputation)
Exploratory endpoints			
Clinical response CDAI	NRI		
Clinical response CDAI-70	NRI		
CDAI score		OBS	
Clinical response PRO2	NRI		
PRO2 score		OBS	
APSF score and subscores		OBS	
Time to remission/response by PRO2/Normalization of FCP		OBS	
SES-CD score		OBS	
Endoscopic remission	NRI;		
Clinical remission APSF	NRI		
Endoscopic response/Clinical remission PRO2	NRI		
Endoscopic response/Clinical remission CDAI	NRI		

Quality of life (IBDQ, CD-PRO, SF-36, EQ-5D-5L, WPAI-CD, FACIT-F, AP NRS)		OBS	
PGIC	OBS		
Biomarkers (ALC, CRP, FCP, RHIS)		OBS	
Fistulas		OBS	
Hospitalizations, surgeries	OBS		
Corticosteroid-free clinical remission PRO2 and/or endoscopic response	NRI		

Abbreviations: ALC, absolute lymphocyte count; AP, abdominal pain; APSF, abdominal pain stool frequency; BL, baseline; CD-PRO, Crohn's Disease Patient-Reported Outcomes; CDAI, Crohn's Disease Activity Index; CRP, c-reactive protein; EQ-5D-5L, EuroQoL-5 Dimensions 5-Level Version; FACIT-F, Functional Assessment of Chronic Illness Therapy – Fatigue; FAS, Full Analysis Set; FCP, fecal calprotectin; IBDQ, Inflammatory Bowel Disease Questionnaire; mFAS, modified Full Analysis Set; NRI, nonresponse imputation; NRS, numeric rating scale; OBS, observed data; RHIS, Robart's Histopathology Index Score (RHIS); 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.

[a] Primary analysis. See Sections 15.1.4, 15.2.4, and 15.3.4 for sensitivity analyses.

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 at Week 14

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 study, as confirmed after blinded review by clinical and medical team members before unblinding, or who discontinued due to disease worsening will be considered to have an outcome of nonresponse in the analysis of all binary efficacy endpoints on the event date and at any subsequent timepoints. The prohibited medications will be identified by the Pfizer clinical team during blinded data review. Disease worsening will be determined based on discontinuation

reasons collected in CRF.

Subjects with missing response status will be included in the analyses using the following missing data methods:

- Main method: multiple imputation for SES-CD component sub-scores under Missing At Random and then derive the binary response. Details will be described in Appendix 11.
- Alternative method: treated all missing data as nonresponder
- Supplementary analyses (Induction analysis): mFAS, using observed data only.

15.1.3. Main Analysis and alternative analysis of Primary Efficacy Variables

For the Induction Analysis, the main analysis will be performed using the FAS. Missing data will handled using multiple imputation as describe in Section 15.1.2. Each MI dataset will be analyzed using Mantel-Haenszel weighted analysis adjusted for randomization stratification factors. The confidence interval for each treatment group will use the normal approximation. The results will be combined using Rubin's rule to report the final response rate for each treatment group, common risk difference between etrasimod vs placebo with 90% CI and p-value.

For the Induction Analysis, the alternative analysis will be performed using the FAS. Summaries will be displayed for the three Induction Analysis treatment groups (placebo, etrasimod 2 mg and 3 mg). The endpoint will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals for treatment group proportions will be computed using the exact method.. Missing data will be imputed as nonresponder. For Cohort 1, p-values from MH weighted test, adjusted for randomization stratification factors for comparison between etrasimod doses vs placebo, the common risk difference and the two-sided 90% confidence interval from the MH method adjusted for the stratification will be presented as well.

15.1.4. Supplementary Analysis of Primary Efficacy Variables

For the Induction Analysis, the alternative analysis will be repeated using the mFAS. Unless specified otherwise, supplementary analyses are based on observed data, i.e., 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

- The proportion of subjects with clinical remission CDAI at Week 14

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.

- Proportion of subjects with clinical remission by PRO2 at Week 14

15.2.2. Missing Data Methods for Secondary Efficacy Variables

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

- Main method: Multiple imputation under missing at random (MAR) for CDAI scores and PRO2 scores and then derive the binary efficacy endpoints. Details can be found in Appendix 11.
- Alternative method: treated as nonresponder (NRI)
- Supplementary analyses (Induction analysis): mFAS, using observed data only.

15.2.3. Primary Analysis of Secondary Efficacy Variables

For the Induction Analysis, the main analysis will be performed using the FAS. Missing data will be handled using multiple imputation as describe in Section 15.1.2. Each MI dataset will be analyzed using Mantel-Haenszel weighted analysis adjusted for randomization stratification factors. The confidence interval for each treatment group will use the normal approximation. The results will be combined using Rubin's rule to report the final response rate for each treatment group, common risk difference between etrasimod vs placebo with 90% CI and p-value.

For the Induction Analysis, the alternative analysis will be performed using the FAS. The secondary efficacy endpoint will be summarized using frequency counts, percentages and 90% CIs. Confidence intervals for treatment group proportions will be computed using the exact method. Missing data will be imputed as nonresponder. P-values from MH weighted test, adjusted for randomization stratification factors for comparison between etrasimod doses vs placebo, the common risk difference and the two-sided 90% confidence interval from the MH method adjusted for the stratification will be presented as well.

For the Extended Induction Analysis, the analyses will be performed using FAS subjects who continued to the Extended Induction Period. Summaries will be displayed for the four treatment groups in extended induction period (Induction placebo to Extension etrasimod 2 mg, Induction placebo to Extension etrasimod 3 mg, Induction etrasimod 2 mg to Extension etrasimod 2 mg and Induction etrasimod 3 mg to Extension etrasimod 3 mg). The secondary efficacy 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.

15.2.4. Supplementary Analysis of Secondary Efficacy Variables

For the Induction Analysis, the alternative analysis will be repeated using the mFAS. Unless specified otherwise, supplementary analyses are based on observed data, i.e., without missing data imputation.

15.3. Exploratory Efficacy

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

15.3.1. Exploratory Efficacy Variables and Derivations

15.3.1.1. CDAI/PRO2

- Proportion of subjects with clinical remission CDAI by visit up to Week 14 (FAS & NRI, mFAS & OBS)
- Proportion of subjects with clinical remission PRO2 by visit up to Week 14 (FAS & NRI, mFAS & OBS)
- Proportion of subjects with clinical response CDAI by visit up to Week 14 (FAS & NRI)
- Proportion of subjects with clinical response CDAI-70 by visit up to Week 14 (FAS & NRI)
- Change from baseline in CDAI score by visit up to Week 14
- Proportion of subjects with clinical response by PRO2 by visit up to Week 14 (FAS & NRI)
- Change from baseline in PRO2 by visit up to Week 14
- Change from baseline in Abdominal pain score and Stool Frequency Score (APSF), un-weighted AP and SF subscores by visit up to Week 14
- Proportion of subjects with clinical remission by APSF by visit up to Week 14 (FAS & NRI)
- 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.

15.3.1.2. Endoscopic/SES-CD

- Change from baseline in SES-CD scores at Week 14
- Proportion of subjects with endoscopic remission at Week 14 (FAS & NRI)
- Proportion of subjects with endoscopic response and clinical remission PRO2 at Week 14 (FAS & NRI)

- Proportion of subjects with endoscopic response and clinical remission CDAI at Week 14 (FAS & NRI)

15.3.1.3. Other PROs

- Change from baseline in health-related patient reported outcomes by visit up to Week 14 (IBDQ, CD-PRO/SS, SF-36, EQ-5D-5L, WPAI-CD, FACIT-F, AP NRS)
- Proportion of subjects by PGIC score category at Week 14
- Change from baseline in the CD-PRO/SS Bowel and Abdominal Function domain scores at Week 14
- Change from baseline in IBDQ Bowel Symptom domain score by visit up to Week 14
- Change from baseline in IBDQ fatigue item score by visit up to Week 14

15.3.1.4. Biomarkers

- Change and percentage change from baseline in FCP concentration by visit up to Week 14
- Change and percentage change from baseline in CRP concentration by visit up to Week 14
- Change and percentage change from baseline in ALC by visit up to Week 14

15.3.1.5. Fistulas, Hospitalizations, and Surgery

- Proportion of subjects with draining fistulas at Week 14, among subjects with draining fistulas at Baseline
- Proportion of subjects with at least a 50% reduction in draining fistulas at Week 14, among subjects with draining fistulas at Baseline
- Rate of CD-related hospitalizations and surgery

15.3.1.6. Other

- Proportion of subjects who are not on concomitant corticosteroids with clinical remission PRO2 and/or endoscopic response at Week 14

15.3.2. Missing Data Methods for Exploratory Efficacy Variables

For endpoints specified in previous section using NRI for FAS, missing data will be imputed as nonresponders. For mFAS and OBS, 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

Similar analyses described in Section 15.2.3 will be used for the exploratory proportion-based endpoints.

Continuous variables measured at scheduled visits up to Week 14 will be analyzed using a MMRM model. The MMRM model will include treatment group, visit, and interaction of treatment-by-visit as factors, and baseline measure and randomization stratification factor as covariates. An unstructured covariance matrix will be specified for the MMRM model. LS means at visit and LS mean differences between treatment group with p-values and corresponding 90% CIs will be reported. This method will also be applied to other score based or continuous measures. No missing data will be imputed.

Continuous variables only measured at Baseline and Week 14 will be analyzed using analysis of covariance (ANCOVA) model. The ANCOVA model will include treatment group as factor, and baseline measure and randomization stratification factor as covariates. LS means and LS mean differences between treatment group with p-values and corresponding 90% CIs will be reported. No missing data will be imputed.

Continuous endpoints will be summarized using the number of observations, mean, standard deviation, median, minimum, maximum. 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, p-values will be presented, using the Wilcoxon rank sum test. Analysis will be performed using the mFAS. The exploratory efficacy-related biomarker analyses for immunophenotyping, proteomics, RNA transcriptomics, fecal microbiome, and endoscopic healing index will be described in a separate plan.

CD-related hospitalizations and surgeries will be summarized for the FAS. 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 25^{th}, 50^{th}, and 75^{th} time to event percentiles, and 90% confidence intervals. Kaplan-Meier estimates will be presented as survival curves by treatment. Time to event summaries will use the mFAS.

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 discontinue 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

For induction period, proportion of subjects with clinical remission CDAI by visit and proportion of subjects with clinical remission PRO2 by visit will be analyzed for FAS using non-responder imputation for missing data, as well as mFAS using observed cases without imputation for missing data.

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 14 and EI - Week 6/ET. The HRQoL analyses will be performed using the mFAS, 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. Crohn’s Disease Patient-Reported Outcomes (CD-PRO)

The CD-PRO is a validated instrument designed to assess the signs, symptoms, and impact of CD through 6 modules: Module 1 (Bowel Signs and Symptoms), Module 2 (Abdominal Symptoms), Module 3 (Systemic Symptoms), Module 4 (Coping Strategies), Module 5 (Daily Life Impact), and Module 6 (Emotional Impact). The CD-PRO will be administered if and when available. The CD-PRO module scores are derived on the eCRF directly. For each CD-PRO module score change from baseline will be summarized by visit and treatment group. Details about scoring rules and derivations of domain scores can be found in APPENDIX 8.

16.1.3. 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.

16.1.4. 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.

16.1.5. 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.

16.1.6. 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.

16.1.7. 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. The number and percent of subjects with meaningful change (as reported by the subject) differences will also be summarized.

16.1.8. 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. Total duration will be calculated as the sum of the hospitalization discharge date – hospitalization admission date + 1.

16.1.9. Functional Assessment of Chronic Illness Therapy - Fatigue (FACIT-F)

The FACIT-F scale is a 13-item instrument designed to assess fatigue/tiredness and its impact on daily activities and functioning in a number of chronic diseases. The instrument includes items such as tiredness, weakness, listlessness, lack of energy, and the impact of these feelings on daily functioning (eg, sleeping, and social activities). The FACIT-F subscale score is derived on the eCRF directly. Change from baseline in FACIT-F subscale score will be summarized by visit and treatment group. Details about derivations can be found in APPENDIX 9 and APPENDIX 10.

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 both the Induction and Extended Induction analysis, only Cohort 1 data will be summarized. Data from Cohort 2 will be combined and analyzed with data collected in SS2-I.

17.1. ADVERSE EVENTS

AEs will be coded using MedDRA. The most up to date MedDRA version at the time of analysis will be used. Treatment-emergent adverse events (TEAEs) are defined as AEs that started or worsened in severity on or after the first dose of study treatment.
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 etrasimod 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.

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.

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. TEAEs Leading to Interruption of Study Treatment

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

17.1.4. TEAEs Leading to Discontinuation from Study

TEAEs leading to discontinuation from study will be identified by primary reason off study being recorded as "Adverse Event" on the End of Study eCRF.
All TEAEs leading to discontinuation from study will be summarized by SOC and PT.

17.1.5. 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.

17.1.6. 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.

17.1.7. 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 by Pfizer clinicians to identify which events reflect AESI.

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

- Cardiovascular Events
  - Bradycardia
  - AV conduction delay
  - Hypertension
- Macular Edema
- Pulmonary Disorders
  - Airflow obstruction (forced expiratory volume in 1 second [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

17.1.8. 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.

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 selected hematology parameters, selected serum chemistry parameters, 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 substudy and anytime in the study
- 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

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 by normal baseline, and overall
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in ALT
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in AST
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in total bilirubin
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in ALP
  - >= 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 elevation in either ALT or AST and >= 2 x ULN elevation in total bilirubin and < 2 x ULN elevation in ALP
- Confirmed Incidence of hepatic enzyme elevations at post-baseline by normal baseline, abnormal baseline, and overall. A confirmed case is defined as >=2 consecutive test values meet the criteria.
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in ALT
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in AST
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in total bilirubin
  - > 1, >= 2, >= 3, >= 5 x ULN elevation in ALP
- 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
- Summary statistics for value, change from baseline and percent change from baseline in Lipid panel results by visit

Subject's laboratory assessments, central or local, at all timepoints in chronological order will be listed. 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.

17.3.2. Pregnancy Tests

Not applicable.

17.3.3. OTHER SCREENING LABORATORY ASSESSMENTS

Not applicable.

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)
- 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
- Shift in markedly abnormal categories from baseline to post-baseline by visit

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 and QTcF:
  - ≥ 450 msec (male) or ≥ 470 msec (female) in QTcF
  - > 500 msec in QT
- 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 (^{0}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 (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 (^{0}C) = (5/9) (Temperature (^{0}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

17.6. PHYSICAL EXAMINATION

Not applicable.

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 in 1 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

Not applicable.

17.7.4. Endoscopy

All endoscopy biopsy assessment results will be listed.

18. 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 Week 6 of the Extended Induction 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, min, max and 2 decimal places for SD and SE
- 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**
Placebo	Placebo
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, Safety and Biomarker analysis
    - Induction Placebo to Extension 2 mg

  - Induction Placebo to Extension 3 mg
  - Induction 2 mg to Extension 2 mg
  - Induction 3 mg to Extension 3 mg
  - Overall total
- For Efficacy analysis
  - Induction Placebo to Extension 2 mg
  - Induction Placebo to Extension 3 mg
  - Induction 2 mg to Extension 2 mg
  - Induction 3 mg to Extension 3 mg

Treatment groups will be presented by Extended Induction Treatment and Induction Treatment. Summaries will also include Placebo, 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 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 6, Week 10, Week 14, EI - First Dose, EI - Week 1, EI - Week 6	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):

- Randomized treatment group (or treatment received if it's a safety output), first by Placebo, then 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)
- 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 (PK 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 cannot be on or after date of first dose of study drug	Known, Partial or Missing	Not TEAE
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

AE START DATE	AE STOP DATE	ACTION
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
	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

START DATE	STOP DATE	ACTION
	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:

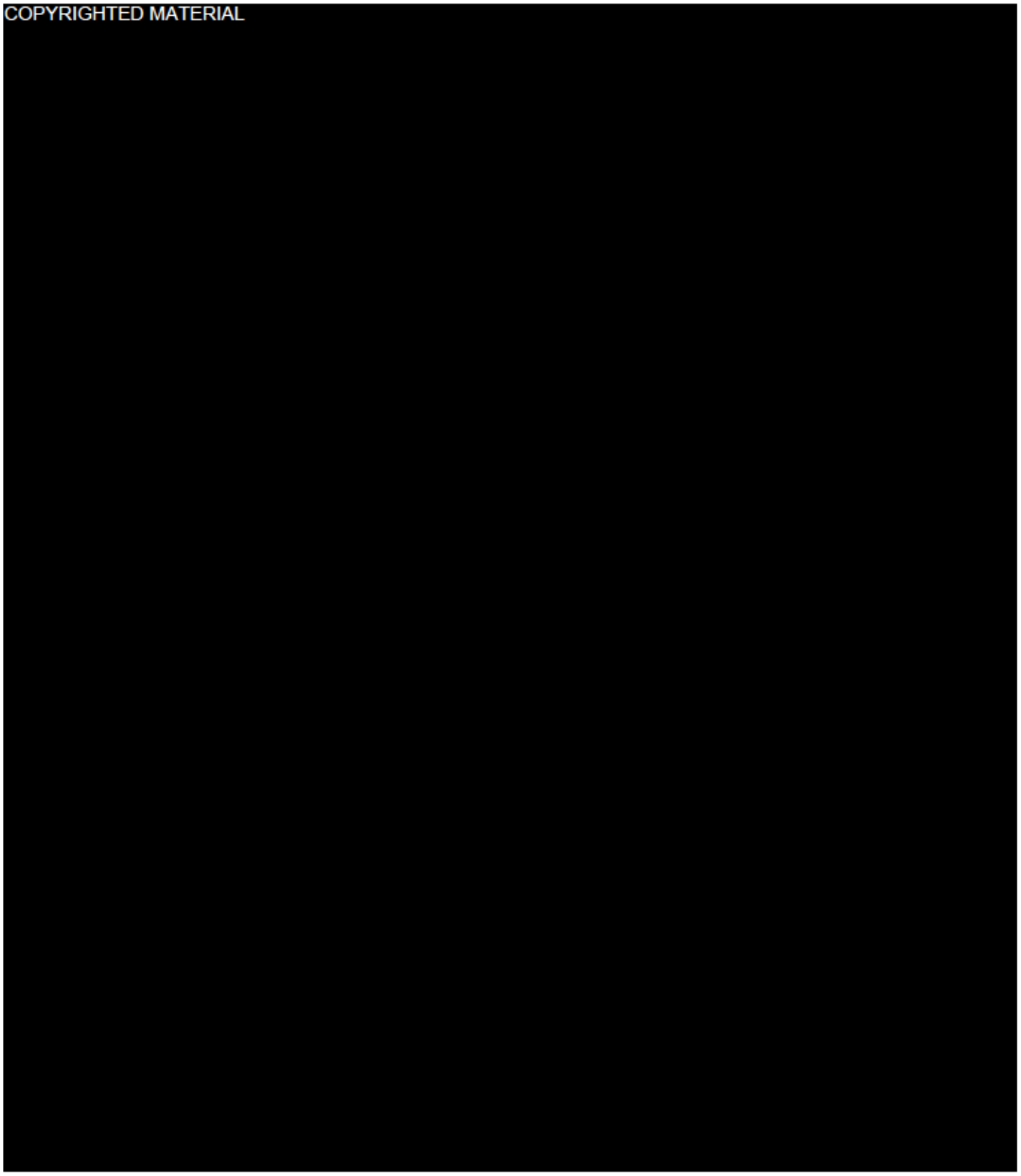
START DATE	ACTION
Partial	Impute start date as earliest possible date (i.e. first day of month if day unknown or 1st January if day and month are unknown)

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:

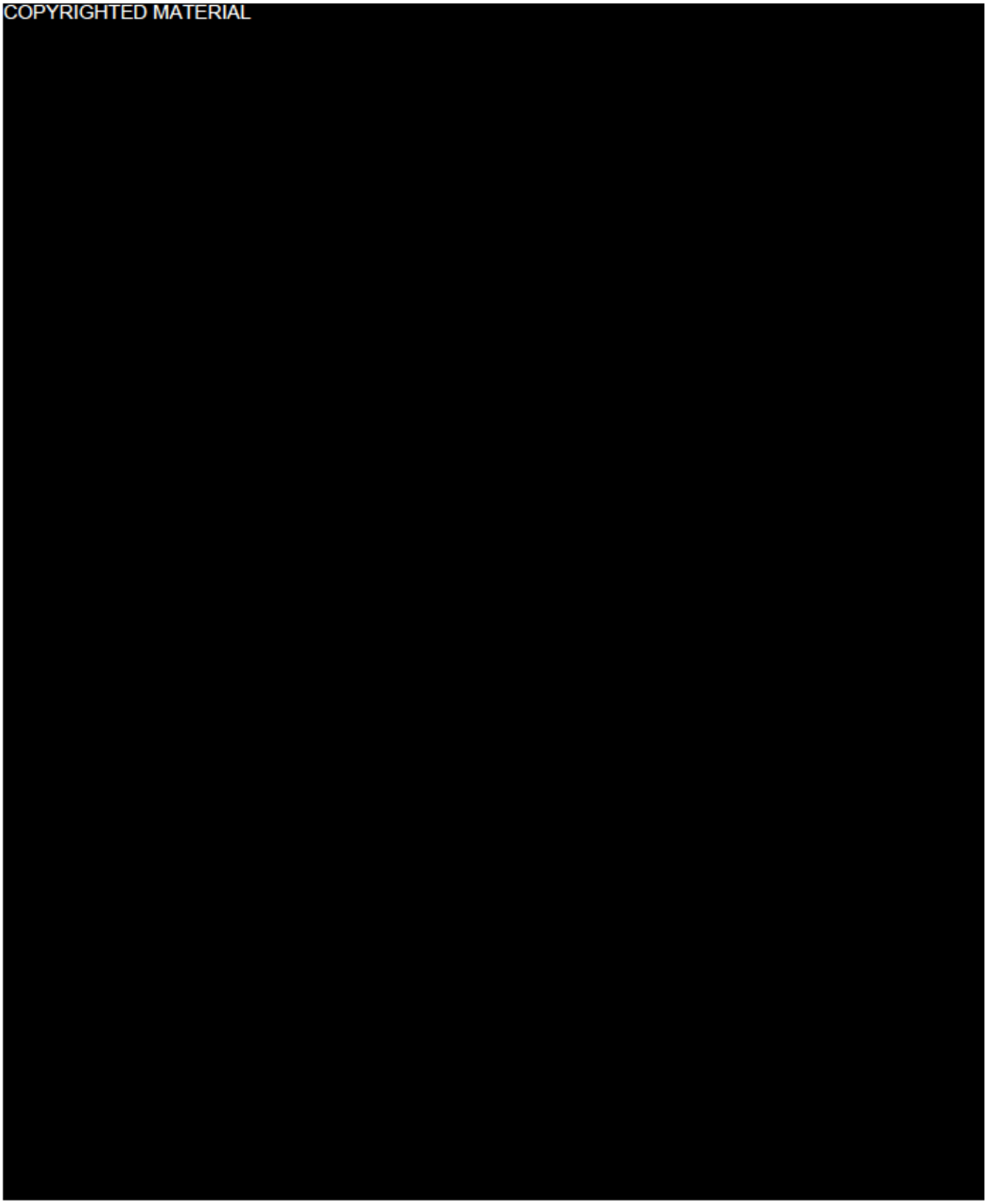
1 = Are you currently employed?
2 = During the past seven days, how many hours did you miss from work due to your problems associated with CD?
3 = During the past seven days, how many hours did you miss from work because of any other reasons?
4 = During the past seven days, how many hours did you actually 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?

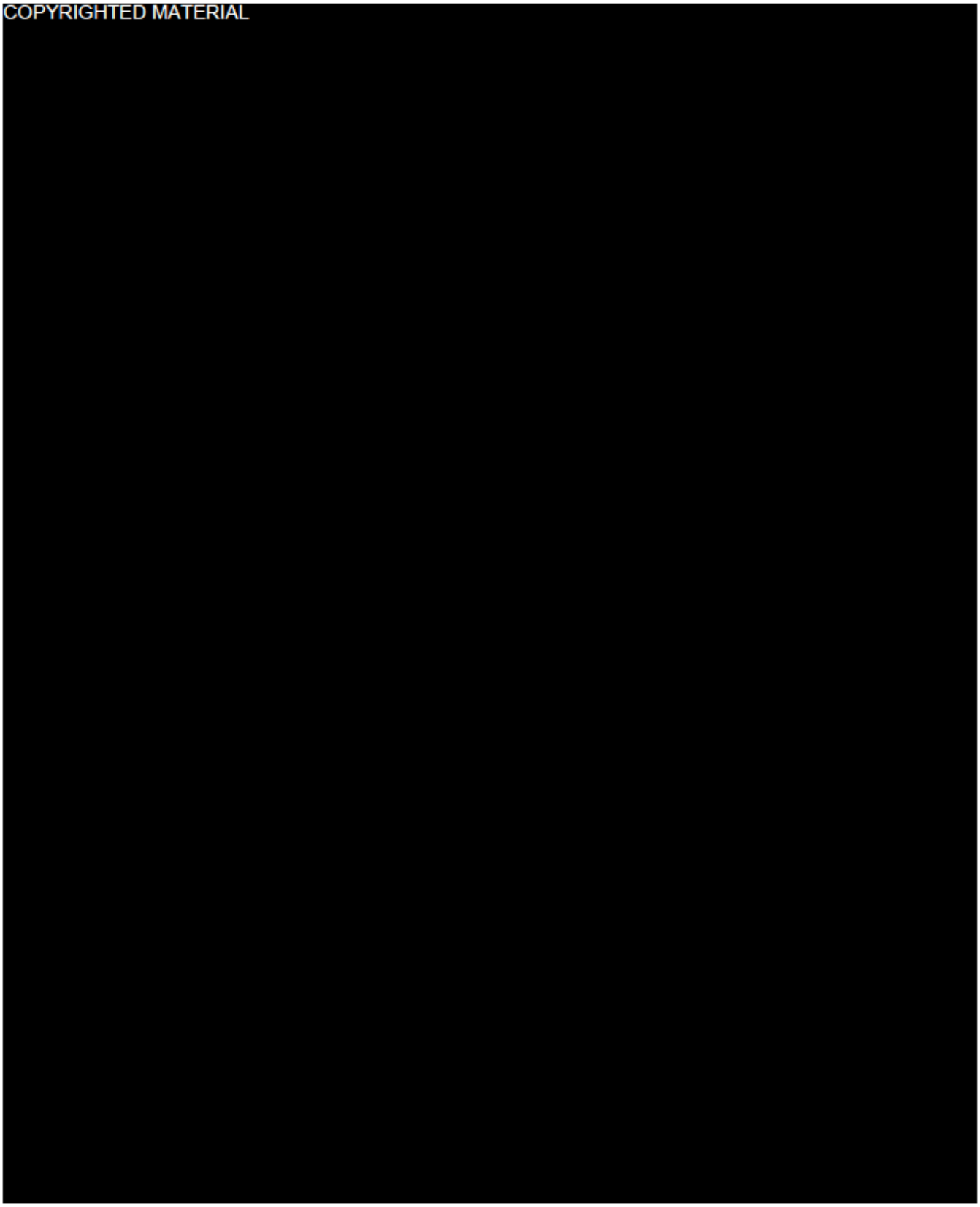
Derivation:

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.

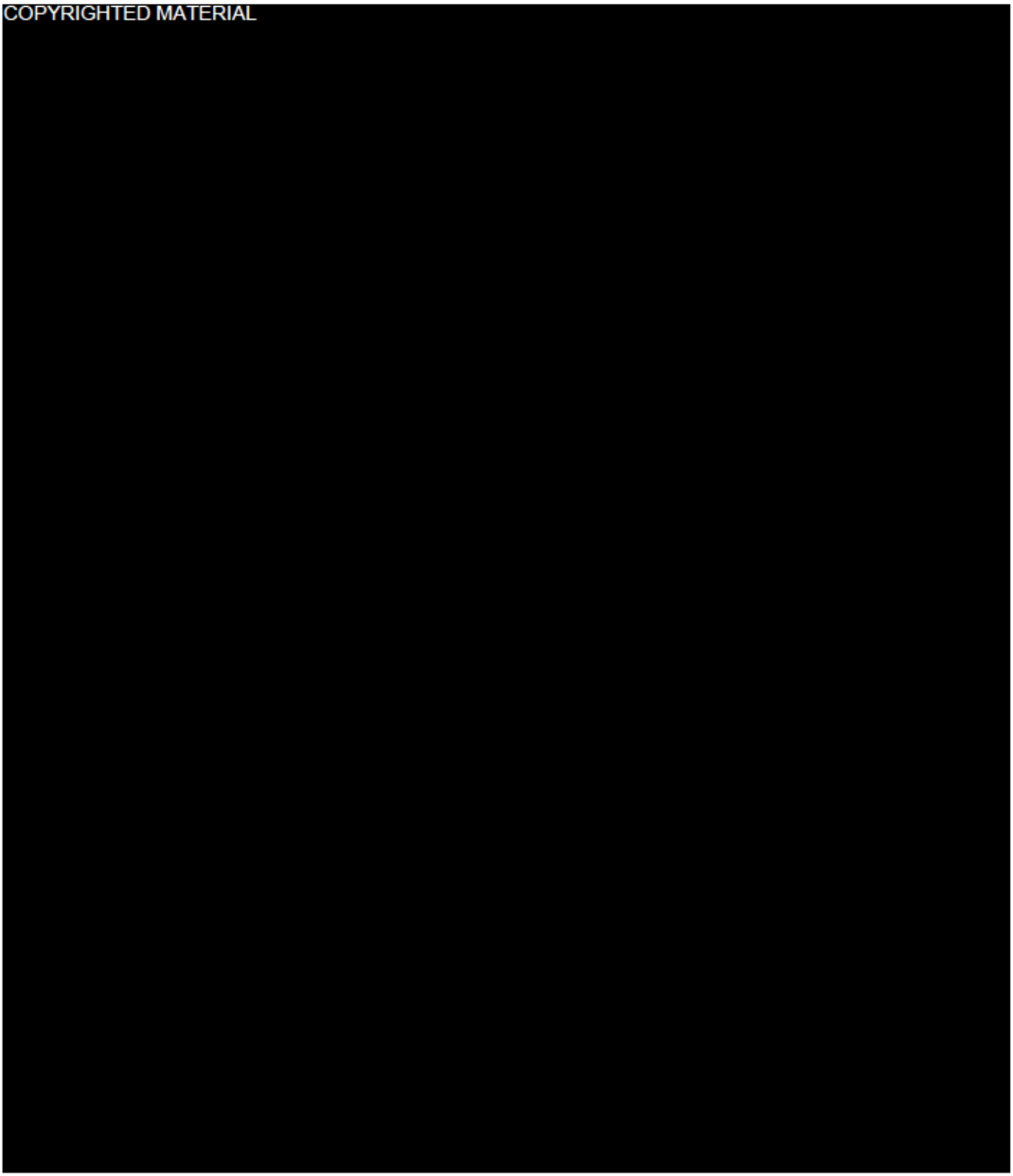
APPENDIX 8. CD-PRO/SS SCORING RULES





APPENDIX 9. FUNCTIONAL ASSESSMENT OF CHRONIC ILLNESS THERAPY - FATIGUE (FACIT-F) SCALE (VERSION 4)

APPENDIX 10. FACIT-F SUBSCALE SCORING



CCI

[redacted]

CCI

[redacted]

CCI

[redacted]