Statistical Analysis Plan for Study M14-431

A Multicenter, Randomized, Double-Blind, Placebo-Controlled Induction Study of the Efficacy and Safety of Upadacitinib (ABT-494) in Subjects with Moderately to Severely Active Crohn's Disease Who Have Inadequately Responded to or are Intolerant to Biologic Therapy

Date: 15 November 2021

Version 5.0

Table of Contents

List of Tables

List of Figures

List of Appendices

1.0 Introduction

This Statistical Analysis Plan (SAP) describes the statistical analyses for upadacitinib (ABT-494) Study M14-431, A Multicenter, Randomized, Double-Blind, Placebo Controlled Induction Study of the Efficacy and Safety of Upadacitinib (ABT-494) in Subjects with Moderately to Severely Active Crohn's Disease Who Have Inadequately Responded to or are Intolerant to Biologic Therapy.

The analyses of pharmacokinetics/pharmacodynamic endpoints, pharmacogenetic endpoints, and selected biomarkers will be analyzed separately and are not described in this SAP.

The SAP will not be updated in case of administrative changes or amendments to the protocol unless the changes impact the analysis.

Unless noted otherwise, all analyses will be performed using SAS Version 9.4 (SAS Institute Inc., Cary, NC 27513) or later under the UNIX operating system.

This SAP includes changes to analyses described in previous SAP versions. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Objectives, Hypotheses and Estimands

The objective of Study M14-431 is to evaluate the efficacy and safety of upadacitinib compared to placebo as induction therapy in subjects with moderately and severely active Crohn's Disease (CD).

Primary Efficacy Objective

For EU/EMA regulatory purposes, the primary efficacy objective of Study M14-431 is to evaluate the efficacy based on higher rates of clinical remission per patient reported outcomes (PROs) of very soft or liquid stool frequency and abdominal pain score, and

endoscopic response after 12 weeks of treatment with upadacitinib 45 mg once daily (QD) compared to placebo as induction therapy in subjects with moderately and severely active CD who have inadequately responded to or are intolerant to biologic therapy during the 12-week double-blind induction period (Part 1).

Hypotheses corresponding to the primary efficacy objectives and endpoints are:

- The proportion of subjects achieving clinical remission per PROs at Week 12 treated with upadacitinib 45 mg QD is greater than that treated with placebo, and
- The proportion of subjects achieving endoscopic response at Week 12 treated with upadacitinib 45 mg QD is greater than that treated with placebo.

For US/FDA regulatory purposes, the primary efficacy objective of Study M14-431 is to evaluate the efficacy based on higher rates of clinical remission per Crohn's Disease Activity Index (CDAI) and endoscopic response after 12 weeks of treatment with upadacitinib 45 mg QD compared to placebo as induction therapy in subjects with moderately and severely active CD who have inadequately responded to or are intolerant to biologic therapy during the 12-week double-blind induction period (Part 1).

Hypotheses corresponding to the primary efficacy objectives and endpoints are:

- The proportion of subjects achieving clinical remission per CDAI at Week 12 treated with upadacitinib 45 mg QD is greater than that treated with placebo, and
- The proportion of subjects achieving endoscopic response at Week 12 treated with upadacitinib 45 mg QD is greater than that treated with placebo.

The primary efficacy objectives will be assessed based on Intent-to-Treat (ITT) population, which consists of all randomized subjects who have received at least one dose of double-blinded study drug.

The estimands corresponding to the primary efficacy objectives are defined as follows:

- The difference in the proportion of subjects achieving clinical remission per PROs at Week 12 regardless of premature discontinuation of study drug but before initiation of CD-related confounding medications (See Section 8.1), and without initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population (EU/EMA regulatory purposes).
- The difference in the proportion of subjects achieving clinical remission per CDAI at Week 12 regardless of premature discontinuation of study drug but before initiation of CD-related confounding medications (See Section 8.1), and without initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population (US/FDA regulatory purposes).
- The difference in the proportion of subjects achieving endoscopic response at Week 12 regardless of premature discontinuation of study drug but before initiation of CD-related confounding medications (See Section 8.1), and without initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

Secondary Efficacy Objectives

The secondary efficacy objectives of the Study M14-431 are to demonstrate higher efficacy of treatment with upadacitinib 45 mg QD when compared to placebo during the 12-week double-blind induction period (Part 1) with respect to the ranked secondary endpoints specified in Section 3.2 in subjects with moderately and severely active CD who have inadequately responded to or are intolerant to biologic therapy. The secondary efficacy objectives will be assessed based on ITT population.

Hypotheses corresponding to the ranked secondary efficacy objectives and endpoints are: for each of the ranked secondary endpoints except occurrence of hospitalizations due to CD during Part 1 (Section 3.2), greater proportion of subjects with improvement (or greater mean change from Baseline for continuous endpoints) is achieved with upadacitinib 45 mg QD group when compared to that of placebo. Hypothesis

corresponding to occurrence of hospitalizations due to CD during Part 1 is: smaller proportion of subjects experience at least one occurrence of hospitalizations due to CD during Part 1 with upadacitinib 45 mg QD group when compared to that of placebo.

The estimands corresponding to the secondary efficacy objectives are defined for each of the binary ranked secondary endpoints except occurrence of hospitalizations due to CD during Part 1 as follows: The difference in the proportion of subjects achieving binary endpoints regardless of premature discontinuation of study drug but before initiation of CD-related confounding medications (See Section 8.1), and without initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

The estimand corresponding to the secondary objective for occurrence of hospitalizations due to CD during Part 1 is as follows: The difference in the proportion of subjects with at least one occurrence of hospitalization due to CD during Part 1 regardless of premature discontinuation of study drug (See Section 8.1), and regardless of initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

The estimands corresponding to the secondary efficacy objectives are defined for each of the continuous ranked secondary endpoints as follows: the difference in the mean change from baseline regardless of premature discontinuation of study drug but before initiation of CD-related confounding medications (See Section 8.1), and without initiation or dose escalation of CD-related corticosteroids (See Section 8.2) in the upadacitinib 45 mg QD and placebo groups in the ITT population.

2.2 Study Design Overview

This is a Phase 3, randomized, double-blind, placebo-controlled induction study to evaluate the efficacy and safety of upadacitinib, an orally administered JAK1 inhibitor in adult subjects with moderately to severely active CD who have inadequately responded to

or are intolerant to biologic therapy. This study will enroll approximately 645 subjects at approximately 400 study centers worldwide.

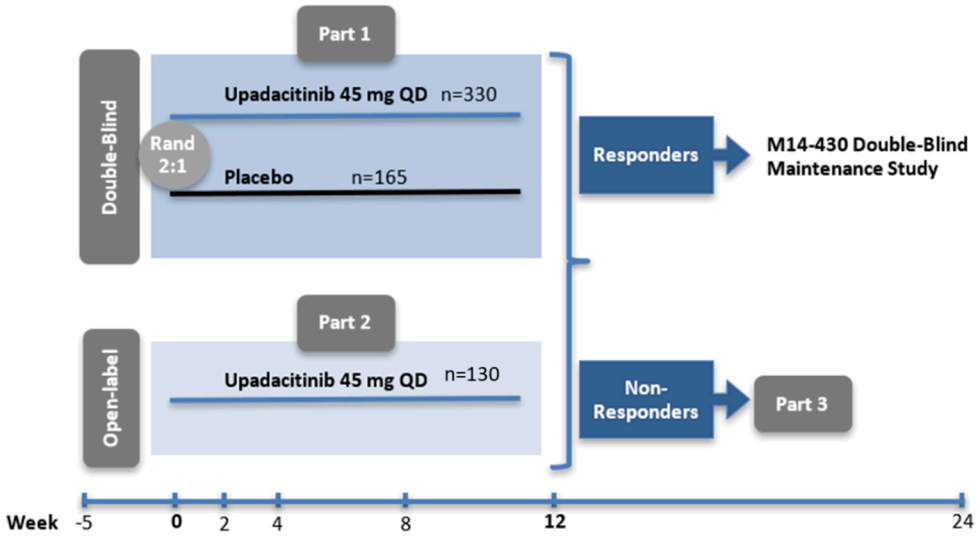
The biologics approved for the treatment of CD include adalimumab, certolizumab, infliximab, ustekinumab and vedolizumab. The study will allow enrollment of approximately 30% subjects who have had inadequate response or intolerance to 3 or more biologics. The study will allow the enrollment of up to 10% of subjects who received a JAK inhibitor prior to study entry if they have not had inadequate response or loss of response.

The duration of the study could be up to 33 weeks and consists of the following parts:

1. Screening period of up to a maximum of 35 days
2. **Part 1:** a randomized, double-blind, placebo-controlled 12-week induction
3. **Part 2:** an open-label, single-arm active 12-week induction
4. **Part 3:** a 12-week Extended Treatment Period for subjects who do not achieve clinical response at Week 12 of Part 1 or Part 2
5. 30-day follow up period for subjects who do not enroll into Study M14-430

The schematics of the overall study design are shown in Figure 1 (Part 1 and Part 2) and Figure 2 (Part 3).

Figure 1. Study M14-431 Study Design – Part 1 and Part 2



QD = once daily; Rand = randomization

Part 1

In Part 1, subjects (n = 495) will be randomized in a 2:1 ratio to upadacitinib 45 mg QD or matching placebo for 12 weeks (Figure 1). Centralized randomization will be implemented in the study. The randomization will be stratified by baseline corticosteroid use (yes or no), endoscopic disease severity (Simplified Endoscopic Score for Crohn's Disease [SES-CD] < 15 or ≥ 15), and number of prior biologic treatments (> 1 or ≤ 1). The data collected from subjects from Part 1 will be used for the primary efficacy analysis.

Visits during the study will occur at Baseline and Weeks 2, 4, 8, and 12/Premature Discontinuation (PD) to collect clinical, endoscopic and laboratory assessments of disease activity.

At Week 12, subjects achieving clinical response, defined as ≥ 30% decrease in average daily very soft or liquid stool frequency (SF) and/or ≥ 30% decrease in average daily

abdominal pain (AP) score (both not worse than Baseline) may be eligible to enter the 52-week, double-blind, maintenance portion of Study M14-430.

All subjects who do not achieve clinical response at Week 12 will be able to enroll in Part 3 (Extended Treatment Period) to receive a double-blind regimen with upadacitinib until Week 24/PD.

Subjects are not eligible to enter Study M14-430 until the ileocolonoscopy procedure at Week 12 has been completed. If the Coronavirus 2019 disease (COVID-19) pandemic precludes a subject from undergoing an endoscopy, a subject can enroll in Study M14-430 if clinical response was achieved at Week 12.

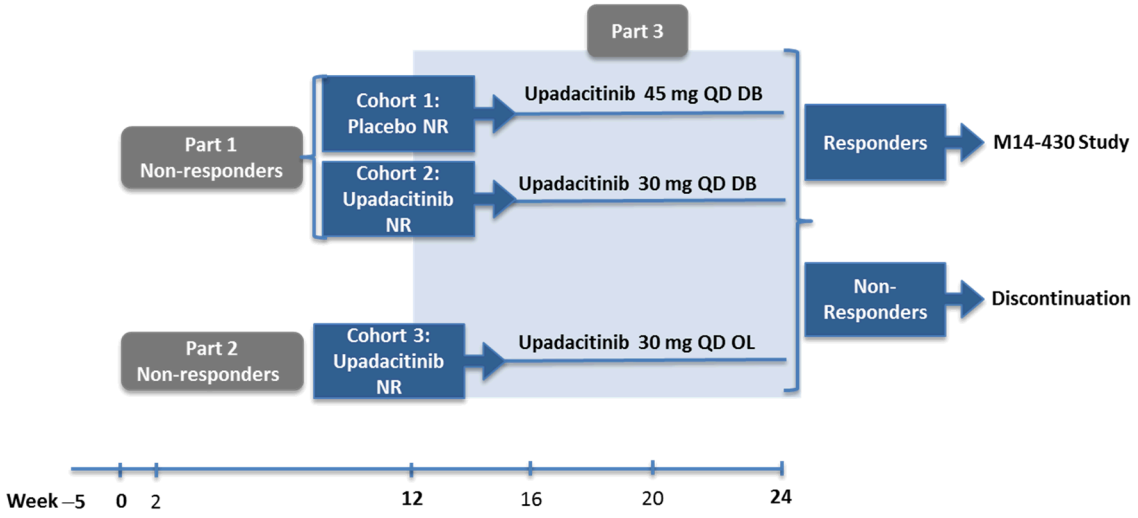
Part 2

Part 2 is an open-label portion (open-label cohort) of this study. Once enrollment in Part 1 is complete, approximately 130 subjects will be enrolled to receive upadacitinib 45 mg QD for 12 weeks (Figure 1). The objective of Part 2 is to have a sufficient number of subjects with clinical response to be re-randomized in the double-blind maintenance portion of Study M14-430, while minimizing unnecessary exposure to placebo. The data collected from subjects from Part 2 will not be part of the primary efficacy analysis for this study and will be reported separately in the clinical study report.

Visits during the study will occur at Baseline and Weeks 2, 4, 8, and 12/PD to collect clinical, endoscopic and laboratory assessments of disease activity.

At Week 12, subjects achieving clinical response may be eligible to enter Study M14-430 (Figure 1). Subjects who do not achieve clinical response at Week 12 will be able to participate in Part 3 (Extended Treatment Period) to receive open-label upadacitinib 30 mg QD until Week 24/PD.

Figure 2. Study M14-431 Study Design – Part 3



DB = double-blind; OL = open-label; NR = non-responders; QD = once-daily

Part 3

Part 3 is a 12-week Extended Treatment Period consisting of 3 cohorts for subjects who do not achieve clinical response at Week 12 in Part 1 or Part 2 (Figure 2). The objectives of Part 3 are to offer blinded upadacitinib induction treatment to placebo non-responders from Part 1 and to evaluate a delayed clinical response to upadacitinib in subjects who did not initially respond to upadacitinib during Part 1 or Part 2.

- **Cohort 1:** Subjects who received double-blind placebo in Part 1 and did not achieve clinical response at Week 12 are eligible to receive double-blind induction treatment with upadacitinib 45 mg QD for 12 weeks (until Week 24).
- **Cohort 2:** Subjects who received double-blind upadacitinib in Part 1 and did not achieve clinical response at Week 12 are eligible to receive double-blind upadacitinib 30 mg QD for 12 weeks (until Week 24).
- **Cohort 3:** Subjects who received open-label upadacitinib during Part 2 and did not achieve clinical response at Week 12 are eligible to receive open-label upadacitinib 30 mg QD for 12 weeks (until Week 24).

Subjects in Cohort 1 and 2 will remain blinded to treatment to avoid unmasking the treatment received during Part 1. The data collected from subjects from Part 3 will not be part of the primary efficacy analysis and will be reported separately in the clinical study report.

Subjects are not eligible to enter in Part 3 until the Week 12 endoscopy has been completed. If the COVID-19 pandemic precludes a subject from undergoing an endoscopy, a subject can continue into Part 3. Visits will occur at Weeks 16, 20 and 24/PD to collect clinical, endoscopic and laboratory assessments of disease activity.

During Part 3, subjects with persistent symptoms or worsening of CD may be discontinued at any time.

At Week 24, subjects who achieve clinical response may be eligible to enter Study M14-430. Subjects are not eligible to enter Study M14-430 until the ileo-colonoscopy procedure at Week 24 for evaluation of mucosal inflammation has been completed. If the COVID-19 pandemic precludes a subject from undergoing an endoscopy, a subject can enroll if clinical response was achieved at Week 24.

Subjects who do not achieve clinical response at Week 24 will be discontinued from Study M14-431 and will receive standard of care treatment at investigator's discretion. Subjects who do not achieve clinical response at Week 24 and all subjects who prematurely discontinue the study will have a follow-up visit 30 days from the last dose of study drug to collect information on new or ongoing AEs and laboratory assessments. Subjects will be discontinued from the study if they withdraw consent or if they are deemed unsuitable to continue for any reason by the investigator.

2.3 Treatment Assignment and Blinding

At Baseline, 495 subjects will be randomized in a 2:1 ratio to upadacitinib 45 mg QD or matching placebo. The randomization will be stratified by baseline corticosteroid use (yes or no), endoscopic disease severity (SES-CD < 15 or ≥ 15) and number of prior biologics with prior inadequate response or intolerance (> 1 or ≤ 1).

Subjects who received placebo during Part 1 and did not achieve clinical response at Week 12 are eligible to enroll in Cohort 1 of Part 3 to receive double-blind induction treatment with upadacitinib 45 mg QD for 12 weeks (until Week 24).

Subjects who received upadacitinib during Part 1 and did not achieve clinical response at Week 12 are eligible to enroll in Cohort 2 of Part 3 to receive double-blind treatment with upadacitinib 30 mg QD for 12 weeks (until Week 24).

Subjects who received upadacitinib during Part 2 and did not achieve clinical response at Week 12 are eligible to enroll in Cohort 3 of Part 3 to receive open-label treatment with upadacitinib 30 mg QD for 12 weeks (until Week 24).

Subjects in Cohort 1 and Cohort 2 of Part 3 will remain blinded to treatment to avoid unmasking the treatment received during Part 1.

2.4 Sample Size Determination

The co-primary endpoints are achievement of clinical remission per PROs (EU/EMA) or per CDAI (US/FDA) at Week 12 and achievement of endoscopic response at Week 12. Sample size calculation is based on the maximum sample size needed to detect treatment difference for each of the co-primary endpoints. The proposed assumptions were informed by the results of the 16-week clinical and endoscopic data of upadacitinib Phase 2 CD Study M13-740.

For EU/EMA regulatory purposes: Assuming a rate of 12% for clinical remission per PROs in the placebo group and 29% in the upadacitinib group at Week 12, a total sample size of 495 subjects randomized in a 2:1 ratio (330 subjects in the upadacitinib group and 165 subjects in the placebo group) will be adequate to detect at least a 17% treatment difference in clinical remission per PROs rates at Week 12 between the treatment groups using Fisher's exact test with at least 95% power at a 0.05 two-sided significant level.

For US/FDA regulatory purposes: Assuming a rate of 20% for clinical remission per CDAI in the placebo group and 40% in the upadacitinib group at Week 12, the same

sample size of 495 subjects will be adequate to detect at least a 20% treatment difference in clinical remission per CDAI rates at Week 12 between the treatment groups using Fisher's exact test with at least 95% power at a 0.05 two-sided significant level.

Assuming an endoscopic response rate of 10% in the placebo group and 25% in the upadacitinib group at Week 12, this sample size will be adequate to detect at least a 15% treatment difference in endoscopic response rates at Week 12 between the treatment groups using Fisher's exact test with at least 95% power at a 0.05 two-sided significant level.

In Part 2, assuming a Week 12 clinical response rate of approximately 52% among the subjects treated with upadacitinib 45 mg QD, an additional 150 subjects enrolled in Part 2 would provide adequate number of subjects achieving clinical response to be re-randomized to the maintenance portion of Study M14-430.

3.0 Endpoints

The following endpoint definitions apply to the efficacy variables described below:

- **Clinical remission per PROs:** average daily very soft or liquid SF ≤ 2.8 AND average daily AP score ≤ 1.0 and both not greater than baseline
- **Clinical remission per CDAI:** CDAI < 150
- **Enhanced Clinical Response:** ≥ 60% decrease in average daily very soft or liquid SF and/or ≥ 35% decrease in average daily AP score and both not greater than baseline, or clinical remission per PROs
- **Clinical response 100 (CR-100):** decrease of at least 100 points in CDAI from Baseline
- **Clinical response per PROs:** ≥ 30% decrease in average daily very soft or liquid SF and/or ≥ 30% decrease in average daily AP score and both not greater than Baseline
- **Endoscopic remission:** SES-CD ≤ 4 and at least 2 point reduction from Baseline and no subscore > 1 in any individual variable, as scored by central reviewer

- **Endoscopic response:** decrease in SES-CD > 50% from Baseline of the induction study (or for subjects with an SES-CD of 4 at Baseline of the induction study, at least a 2 point reduction from Baseline), as scored by central reviewer

Details of CDAI, SES-CD and PRO questionnaires are provided in Appendix D, Appendix E and Appendix F respectively.

3.1 Co-Primary Endpoints

Co-primary endpoints for EU/EMA regulatory purposes:

1. Achievement of clinical remission per PROs at Week 12, and
2. Achievement of endoscopic response at Week 12.

Co-primary endpoints for US/FDA regulatory purposes:

1. Achievement of clinical remission per CDAI at Week 12, and
2. Achievement of endoscopic response at Week 12.

3.2 Secondary Endpoints

The ranked secondary endpoints for EU/EMA regulatory purposes are as follows:

Group 1A_E:

1. Achievement of clinical remission per CDAI at Week 12
2. Achievement of clinical remission per PROs at Week 4
3. Achievement of endoscopic remission at Week 12
4. Discontinuation of corticosteroid use for CD and achievement of clinical remission per PROs at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)

5. Change from Baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) total score at Week 12
6. Change from Baseline in Inflammatory Bowel Disease Questionnaire (IBDQ) total score at Week 12

Group 1B_E:

7. Achievement of CR-100 at Week 2
8. Achievement of CR-100 at Week 12
9. Occurrence of hospitalizations due to CD during Part 1
10. Achievement of resolution of extra-intestinal manifestations (EIMs) at Week 12 (population: ITT subjects with EIMs at Baseline) (see Appendix G)

The ranked secondary endpoints for US/FDA regulatory purposes are as follows:

Group 1A_F:

1. Achievement of clinical remission per PROs at Week 12
2. Achievement of endoscopic remission at Week 12
3. Discontinuation of corticosteroid use for CD and achievement of clinical remission per CDAI at Week 12, in subjects taking corticosteroids for CD at Baseline
4. Change from Baseline in FACIT-F total score at Week 12
5. Change from Baseline in IBDQ total score at Week 12

Group 1B_F:

6. Achievement of CR-100 at Week 2
7. Achievement of CR-100 at Week 12
8. Achievement of clinical remission per CDAI at Week 4

9. Occurrence of hospitalizations due to CD during Part 1
10. Achievement of with resolution of EIMs at Week 12 (population: ITT subjects with EIMs at Baseline) (see Appendix G)

3.3 Other Efficacy Endpoints

3.3.1 Other Endpoints for Part 1

- Achievement of:
  - clinical remission per PROs at Week 2, Week 4, Week 8 and Week 12
  - clinical remission per CDAI at Week 2, Week 4, Week 8 and Week 12
  - enhanced clinical response at Week 2, Week 4, Week 8 and Week 12
  - clinical response at Week 2, Week 4, Week 8 and Week 12
  - CR-100 at Week 2, Week 4, Week 8 and Week 12
  - clinical remission per PROs and endoscopic remission at Week 12
  - clinical remission per CDAI and endoscopic remission at Week 12
  - enhanced clinical response and endoscopic response at Week 12
  - corticosteroid use discontinuation for CD and endoscopic remission at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)
  - corticosteroid use discontinuation for CD and enhanced clinical response at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)
  - corticosteroid use discontinuation for CD and endoscopic response at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)
  - corticosteroid use discontinuation for CD at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)
  - ≥ 50% reduction in the corticosteroid dose for CD from Baseline at Week 12 (population: ITT subjects taking corticosteroids for CD at Baseline)

- decrease in SES-CD > 50% from Baseline or endoscopic remission at Week 12, as scored by central reviewer
- SES-CD ≤ 2 at Week 12
- SES-CD ulcerated surface subscore of 0 at Week 12 in subjects with SES-CD ulcerated surface subscore ≥ 1 at Baseline, as scored by a central reviewer
- SES-CD ulcerated surface subscore ≤ 1 in each segment at Week 12 in subjects with a SES-CD ulcerated surface subscore ≥ 2 at Baseline, as scored by a central reviewer
- endoscopic response at Week 12 among subjects with SES-CD subscore of 3 in the narrowing component at Week 12, which was not present at Baseline
- endoscopic remission at Week 12 among subjects with SES-CD narrowing component subscore between 0 and 2 in all intestinal segments at Baseline
- endoscopic response at Week 12 among subjects with SES-CD narrowing component subscore between 0 and 2 in all intestinal segments at Baseline
- endoscopic response or endoscopic remission at Week 12 among subjects with SES-CD narrowing component subscore between 0 and 2 in all intestinal segments at Baseline
- Colonic and Ileal Global Histologic Disease Activity Score (CGHAS/IGHAS) histologic remission (defined as CGHAS/IGHAS score ≤ 2) at Week 12 in subjects with abnormal histology at Baseline
- ≥ 50% reduction in draining fistulas at Week 12 (population: ITT subjects with draining fistulas at Baseline)
- resolution of draining fistulas at Week 12 (population: ITT subjects with draining fistulas at Baseline)
- at least 50% decrease from Baseline in the number of days with very soft or liquid stools (Type 6 or 7, per Bristol Stool Chart) during the last 7 days at Week 12
- IBDQ remission (IBDQ ≥ 170 points) at Week 4, Week 12
- IBDQ response (increase in IBDQ ≥ 16 points from Baseline) at Week 4, Week 12

- response in IBDQ fatigue item (increase of IBDQ fatigue item score ≥ 1) at Week 4, Week 12
- response in IBDQ Bowel Symptom domain (increase of IBDQ bowel symptom domain score ≥ 8) at Week 12
- increase in FACIT-Fatigue total score from baseline ≥ 7 at Week 12
- increase in FACIT-Fatigue total score from baseline ≥ 9 at Week 12
- increase in FACIT-Fatigue total score from baseline ≥ 11 at Week 12
- resolution of EIMs at Week 2, Week 4, Week 8 and Week 12 (population: ITT subjects with EIMs at Baseline)

- Occurrence of:
  - hospitalizations during the 12-Week Double-Blind Induction Period (Part 1)
  - CD-related surgeries during the 12-Week Double-Blind Induction Period (Part 1)
- Change from Baseline in:
  - IBDQ at Week 4, Week 12
  - individual IBDQ domain scores (bowel, emotional, social, systemic) at Week 4, Week 12
  - individual IBDQ item under Bowel Symptom domain (for Q1, Q5, Q9, Q13, Q17, Q20, Q22, Q24, Q26, and Q29) at Week 4, Week 12
  - Work Productivity and Activity Impairment Questionnaire-Crohn's Disease (WPAI-CD) at Week 4, Week 12
  - European Quality of Life 5 Dimensions 5 level version (EQ-5D-5L) at Week 4, Week 12
  - Short Form-36 (SF-36) at Week 4, Week 12
  - Crohn's Symptoms Severity Questionnaire (CSS) at Week 4, Week 12
  - fecal calprotectin (FCP) at Week 4, Week 12
  - high sensitivity C-reactive protein (hs-CRP) at Week 2, Week 4, Week 8 and Week 12
  - average daily AP score at Week 2, Week 4, Week 8 and Week 12
  - average daily very soft/liquid SF at Week 2, Week 4, Week 8 and Week 12

- average daily total SF at Week 2, Week 4, Week 8 and Week 12
- CDAI at Week 2, Week 4, Week 8 and Week 12
- SES-CD at Week 12
- CGHAS/IGHAS at Week 12 among subjects with abnormal histology at Baseline
- FACIT-F at Week 4 and Week 12

3.3.2 Other Endpoints for Part 2

Same efficacy endpoints as those listed in Part 1 (Section 3.1 and Section 3.2) are considered for Part 2 of the study. All the Part 2 efficacy endpoints are considered other endpoints and ranking of the endpoints is not applicable for Part 2.

3.3.3 Other Endpoints for Part 3

- Achievement of:
  - clinical remission per PROs at Week 16, Week 20 and Week 24
  - clinical remission per CDAI at Week 16, Week 20 and Week 24
  - enhanced clinical response at Week 16, Week 20 and Week 24
  - clinical response at Week 16, Week 20 and Week 24
  - CR-100 at Week 16, Week 20 and Week 24
  - Endoscopic response at Week 24
  - endoscopic remission at Week 24
  - clinical remission per PROs and endoscopic remission at Week 24
  - clinical remission per CDAI and endoscopic remission at Week 24
  - enhanced clinical response and endoscopic response at Week 24
  - corticosteroid use discontinuation for CD and clinical remission per PROs at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
  - corticosteroid use discontinuation for CD and clinical remission per CDAI at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)

- corticosteroid use discontinuation for CD and endoscopic remission at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
- corticosteroid use discontinuation for CD and enhanced clinical response at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
- corticosteroid use discontinuation for CD at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
- corticosteroid use discontinuation for CD and endoscopic response at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
- ≥ 50% reduction in the corticosteroid dose for CD from Baseline at Week 24 (population: ITT subjects taking corticosteroids for CD at Week 12)
- resolution of EIMs at Week 24 (population: ITT subjects with EIMs at Baseline or Week 12)
- decrease in SES-CD > 50% from Baseline or endoscopic remission at Week 24, as scored by central reviewer
- SES-CD ≤ 2 at Week 24
- SES-CD ulcerated surface subscore of 0 at Week 24 in subjects with SES-CD ulcerated surface subscore ≥ 1 at Baseline, as scored by a central reviewer
- SES-CD ulcerated surface subscore ≤ 1 in each segment at Week 24 in subjects with a SES-CD ulcerated surface subscore ≥ 2 at Baseline, as scored by a central reviewer
- CGHAS/IGHAS histologic remission (defined as CGHAS/IGHAS score ≤ 2) at Week 24 in subjects with abnormal histology at Baseline
- ≥ 50% reduction in draining fistulas at Week 24 (population: ITT subjects with draining fistulas at Baseline)
- resolution of draining fistulas at Week 24 (population: ITT subjects with draining fistulas at Baseline or Week 12)

  - at least 50% decrease in the number of days with very soft or liquid stools (Type 6 or 7, per Bristol Stool Chart) during the last 7 days at Week 24.
- Occurrence of:
  - hospitalizations due to CD during Part 3
  - hospitalizations during Part 3
  - CD-related surgeries during Part 3
- Change from Baseline in:
  - IBDQ at Week 24
  - WPAI-CD at Week 24
  - EQ-5D-5L at Week 24
  - CSS at Week 24
  - FACIT-F at Week 24
  - SF-36 at Week 24
  - FCP at Week 16, Week 24
  - hs-CRP at Week 16, Week 20 and Week 24
  - average daily AP score at Week 16, Week 20 and Week 24
  - average daily very soft/liquid SF at Week 16, Week 20 and Week 24
  - average daily total SF at Week 16, Week 20 and Week 24
  - CDAI at Week 16, Week 20 and Week 24
  - SES-CD at Week 24
  - CGHAS/IGHAS at Week 24 among subjects with abnormal histology at Baseline
- Time to:
  - clinical response
  - enhanced clinical response
  - clinical remission per PROs
  - clinical remission per CDAI
  - CR-100

3.4 Safety Endpoints

The following endpoints will be included in the safety analyses:

- Treatment emergent adverse events;
- Serious adverse events;
- Adverse events of special interest;
- Adverse events leading to discontinuation of study drug;
- Vital signs and laboratory tests.

4.0 Analysis Populations

The following population sets will be used for the analyses.

Intent-to-Treat (ITT) Populations

The ITT population for the 12-week double blind induction period (Part 1) (denoted by **ITT1**) includes all randomized subjects who received at least one dose of double-blind study drug during Part 1. The ITT1 population will be used for all efficacy and baseline analyses for Part 1.

For ITT1 population, subjects will be included in the analysis according to the treatment groups that they were randomized to.

The ITT population for the 12-week open-label induction period (Part 2) (denoted by **ITT2**) includes all subjects who have received at least one dose of study drug in Part 2. The ITT2 population will be used for all efficacy and baseline analyses for Part 2.

The ITT population 3 (**ITT3**) includes all subjects who have received at least one dose of study drug in the Part 3.

Per-Protocol Population

The Per-Protocol (PP) population represents a subset of the ITT population, consisting of subjects that did not have a major protocol deviation. Major protocol deviations leading to exclusion from the PP population will be identified prior to database lock.

For **PP** population, subjects will be included in the analysis according to the treatment groups that they were randomized to.

Safety Populations

The safety population for Part 1 (denoted by **SA1**) includes all subjects who received at least one dose of the study drug in Part 1.

The safety population for Part 2 (denoted by **SA2**) includes all subjects who received at least one dose of the study drug in Part 2.

The safety population for Part 3 (denoted by **SA3**) includes all subjects who received at least one dose of the study drug in Part 3.

The safety population for subjects who receive upadacitinib in Part 1 or Part2 and enrolled in Part 3 (denoted by **SA4**) is a subset of SA3 which includes Cohort 2/Cohort 3 of SA3, defined as subjects who received upadacitinib in Part 1 or Part 2 and received at least one dose of study drug in Part 3.

The all upadacitinib safety population (**SA-UPA**) includes all subjects who received at least one dose of upadacitinib in Part 1 or Part 2 or Part 3.

For the safety populations, subjects are assigned to a treatment group based on the "as treated" treatment group, regardless of the treatment randomized. The "as treated" will be determined by the most frequent dose regimen received in the analysis period.

5.0 Subject Disposition

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized for each treatment group for Part 1, for overall for Part 2 and for each cohort for Part 3:

- Subjects randomized in Part 1;
- Subjects who took at least one dose of study drug (Part 1, Part 2 and Part 3);
- Subjects who completed protocol-specified treatment (Part 1, Part 2 and Part 3);
- Subjects who prematurely discontinued study drug (Part 1, Part 2 and Part 3);
- Subjects who prematurely discontinued from study (Part 1, Part 2 and Part 3).

Number and percentage of subjects who discontinued study drug and who withdrew from the study will be summarized by reason (primary reason and all reasons) for each treatment group within Part 1, for overall for Part 2 and for each cohort within Part 3. Subjects with multiple reasons for premature discontinuation will be counted once in the calculation of the number and percentage of total discontinuations.

In addition, subjects who reported at least one of the protocol deviation categories (Appendix A) will be summarized for each treatment group.

6.0 Study Drug Duration and Compliance

For the safety populations (SA1, SA2 and SA3), duration of treatment will be summarized for each treatment group for Part 1 (SA1), for overall for Part 2 (SA2) and for each cohort for Part 3 (SA3). Duration of treatment is defined for each subject as last dose date minus first dose date + 1. Duration of treatment will be summarized for each treatment group using the number of subjects treated, mean, standard deviation, median, minimum, and maximum. In addition, the number and percentage of subjects in each treatment duration interval (≥ 2 weeks, ≥ 4 weeks, ≥ 6 weeks, ≥ 8 weeks, ≥ 10 weeks, ≥ 12 weeks) will be summarized.

Treatment compliance will be summarized by treatment group in Part 1, for overall in Part 2 and by cohort in Part 3. Treatment compliance is defined as the number of tablets actually taken divided by the number of tablets that should have been taken. Percent compliance will be summarized.

7.0 Demographics, Baseline Characteristics, Medical History, and Prior/Concomitant Medications

Demographics, baseline or disease characteristics, medical history, and prior and concomitant medications will be summarized for the ITT1 population overall and by treatment group, and overall for ITT2. Categorical variables will be summarized with the number and percentage of subjects; percentages will be calculated based on the number of non-missing observations. Continuous variables will be summarized with descriptive statistics (number of non-missing observations, mean, standard deviation, median, minimum, and maximum).

7.1 Demographics and Baseline Characteristics

Demographic/vital Signs

- Sex (Male, Female)
- Age (Years)
- Age Category (≥ 18 years - < 40 years, ≥ 40 years - < 65 years, ≥ 65 years)
- Ethnicity (Hispanic/Latino, Non-Hispanic/Latino)
- Race category (White, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Multiple)
- Geographic Region category 1 (US, ex-US)
- Geographic Region category 2 (as defined in Appendix J)
- Tobacco/Nicotine Use (currently smokes, ex-smoker, never smoked, unknown)
- Alcohol use (current drinker, former drinker, non-drinker, unknown)
- Body weight (kg) – overall and by sex
- Body Mass Index (BMI, kg/m^2)

- BMI category
  - Underweight [< 18.5 kg/m^2]
  - Normal [≥ 18.5 and < 25 kg/m^2]
  - Overweight [≥ 25 and < 30 kg/m^2]
  - Obese [≥ 30 kg/m^2]
- Body temperature (°C)
- Height (cm)
- Pulse (bpm)
- Systolic and diastolic blood pressure (mmHg)

Patient Reported Outcome Questionnaires at Baseline

- IBDQ total and domain scores
- WPAI-CD - 4 sub-scores
- CSS
- FACIT-F and its components
- EQ-5D-5L - Index score and VAS
- SF-36 - 8 domain scores, PCS and MCS

Other Baseline Characteristics

- CDAI
- Average Daily very soft or liquid SF
- Average Daily AP score
- SES-CD
- SES-CD category (< 15, ≥ 15)
- Crohn's Disease Duration (years)
- Crohn's Disease Duration category (≤ 5 years, > 5 years)
- CGHAS/IGHAS Score
- Baseline CD-related corticosteroid use (yes, no)

- Baseline CD-related immunosuppressant use (yes, no)
- Baseline CD-related aminosalicylates use (yes, no)
- Prior biologic use/failure history
  - Prior biologic failure history (≤ 1, 2, ≥ 3)
  - Prior failure to anti-TNF agent (yes, no)
  - Prior vedolizumab/natalizumab failure (yes, no)
  - Prior ustekinumab failure (yes, no)
- hs-CRP (mg/L)
- hs-CRP category (< 5, ≥ 5 mg/L)
- FCP (µg/g)
- FCP category (≤ 250 µg/g, > 250 µg/g)
- Draining fistulas (yes, no)
- Non-draining fistulas (yes, no)
- Anal fissures (yes, no)
- CD Location Per SES-CD (ileal, ileocolonic, colonic; Appendix H)
- EIM (yes, no)

In addition, clinical tests at baseline such as TB test, pregnancy test and electrocardiogram assessment will be summarized. Crohn's Disease history (Montreal classification) and Crohn's Disease surgical history will be summarized separately as well.

7.2 Medical History

Medical history data will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). The actual version of the MedDRA coding dictionary will be noted in the statistical tables and clinical study report. The number and percentage of subjects in each medical history category [by MedDRA system organ class (SOC) and preferred term (PT)] will be summarized overall and by treatment group. The SOC will be presented in alphabetical order, and the preferred terms will be presented in alphabetical order within each SOC. Subjects reporting more than one condition/diagnosis will be counted only once in each row (SOC or PT).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name (ITT1 and ITT2 population for prior medications; ITT1, ITT2 and ITT3 populations for concomitant medications). A prior medication is defined as any medication taken prior to the date of the first dose of study drug in Part1 or Part 2.

A concomitant medication for Part 1/Part 2 is defined as any medication other than study drug that (1) was started prior to the first dose study drug in Part 1/Part 2 and continued to be taken after the first dose of study drug in Part 1/Part 2, or (2) was started after the first dose of study drug in Part 1/Part 2, but (i) prior to the first dose of study drug in Part 3 (if applicable), or (ii) prior to the first dose of study drug of M14-430 (if applicable), or (iii) within the last dose + 30 days, whichever is earlier.

A concomitant medication for Part 3 is defined as any medication other than study drug that (1) was started prior to the first dose of study drug in Part 3 and continued to be taken after the first dose of study drug in Part 3 or (2) was started after the first dose of study drug in Part 3, but (i) prior to first dose of M14-430 (if applicable), or (ii) last dose + 30 days, whichever is earlier.

The number and percentage of subjects taking medications will be summarized by generic drug name based on the World Health Organization (WHO) Drug Dictionary for both prior and concomitant medications.

8.0 Handling of Potential Intercurrent Events for the Primary and Key Secondary Endpoints

Potential intercurrent events considered in the Study M14-431 include 1) premature discontinuation of study drug and 2) initiation or dose escalation of CD-related corticosteroids defined in Section 8.2. As suggested by Qu et al. (2021), a mix of strategies will be used to handle different intercurrent events for the efficacy analysis.

8.1 Intercurrent Event 1 (IE1): Premature Discontinuation of Study Drug

Following the intent-to-treaty (ITT) principle, a treatment policy strategy will be used to handle the IE1. Data collected on/after premature discontinuation of study drug will be used for all efficacy endpoints.

The only exception is when data are collected on/after the date subjects switching to other treatment for CD, which will be handled as follows:

- **Binary endpoints other than the occurrence of hospitalization and/or the occurrence of CD-related surgery endpoints:** A composite strategy will be used in the efficacy analysis: the subject will be considered as non-responders at all visits on/after the initiation of other CD-related confounding medications.
- **Binary endpoints regarding the occurrence of hospitalization and/or the occurrence of CD-related surgery endpoints**: A treatment policy strategy will be used regardless.
- **Continuous endpoints:** A hypothetical strategy will be used in the efficacy analysis. Data collected on/after the initiation of other CD-related confounding medications will not be used in the analysis. Efficacy analysis will be handled by using corresponding statistical models (e.g., the Mixed-Effect Model Repeated Measurement [MMRM] model).

The CD-related confounding medications include:

- Biologics approved for CD (Adalimumab, Certolizumab, Infliximab, Ustekinumab, Vedolizumab, Natalizumab)
- Any investigational compound
- Systemic corticosteroids or locally acting steroids, rectal corticosteroid
- Immunosuppressants (azathioprine, mercaptopurine, methotrexate, tacrolimus, cyclosporine)

8.2 Intercurrent Event 2 (IE2): CD-Related Corticosteroids

The CD-related corticosteroids intercurrent event is defined as follows.

In Part 1/Part 2:

1. subjects not on CD-related corticosteroids (systemic or oral locally acting corticosteroids for CD) at baseline which are then initiated during Part 1/Part 2
2. a) subjects on CD-related corticosteroids at baseline who have dosages increased to greater than the prednisone equivalent dose taken at baseline until Week 10 or

   b) subjects with any dose increase in CD-related corticosteroids between Week 10 and Week 12
3. subjects not on rectal corticosteroid at baseline who initiated rectal corticosteroids during Part 1/Part 2, regardless of rectal corticosteroid dose
4. subjects on rectal corticosteroids at baseline who have the dosage increased to greater than the dose taken at baseline during Part 1/Part 2;
5. subjects on rectal corticosteroids at baseline and initiated different rectal corticosteroids during Part 1/Part 2, regardless of rectal corticosteroid dose

In Part 3:

1. subjects not on CD-related corticosteroids (systemic or oral locally acting corticosteroids for CD) at baseline which are then initiated during Part 3
2. a) subjects on CD-related corticosteroids at baseline who have dosages increased to greater than the prednisone equivalent dose taken at baseline and Week 12 until Week 22 or

   b) subjects with any dose increase in CD-related corticosteroids between Week 22 and Week 24

3. subjects not on any rectal corticosteroid at baseline and during Part 1/Part 2 who initiated rectal corticosteroids during Part 3, regardless of rectal corticosteroid dose
4. subjects on rectal corticosteroids at baseline or during Part 1/Part 2 who have dosage increased to greater than the dose taken at baseline during Part 3;
5. subjects on rectal corticosteroids at baseline or during Part 1/Part 2, and initiated different rectal corticosteroids during Part 3, regardless of rectal corticosteroid dose

The time point of the CD-related corticosteroids intercurrent event is defined as the date when one of the scenarios above occurs for a subject.

Binary endpoints other than the occurrence of hospitalization and/or the occurrence of CD-related surgery endpoints: A composite strategy will be used in the efficacy analysis: the subject will be considered as non-responders at all visits on/after the initiation of CD-related corticosteroids.

Binary endpoints regarding the occurrence of hospitalization and/or the occurrence of CD-related surgery endpoints: A treatment policy strategy will be used regardless.

Continuous endpoints: A hypothetical strategy will be used in the efficacy analysis. Data collected on/after the initiation of CD-related corticosteroids will not be used in the analysis. Efficacy analysis will be handled by using corresponding statistical models (e.g., the Mixed-Effect Model Repeated Measurement [MMRM] model).

9.0 Efficacy Analyses

9.1 General Considerations

The primary and secondary efficacy endpoints as specified in Section 3.0 will be analyzed based on ITT1 for Part 1. The difference between treatment groups for the primary and secondary efficacy endpoints will be tested with graphical multiplicity adjustment[1] to

ensure a strong control of family-wise type I error rate at significance level alpha = 0.05 (2-sided) (see Section 12.0). Analyses of other efficacy endpoints for Part 1 will be performed using ITT1 at significance level alpha = 0.05 (2-sided) without multiplicity adjustment.

No statistical comparisons will be performed for Part 2 and Part 3. For all efficacy endpoints, the descriptive statistics will be provided by treatment group for Part 1, for overall for Part 2 and by cohort for Part 3. The statistics include number of observations, mean, standard deviation, 95% CI, minimum, median, and maximum for continuous variables, and number, percent and two-sided 95% CI for discrete variables. And number of subjects with missing data due to COVID-19 pandemic will be summarized for primary and secondary efficacy endpoints.

Unless otherwise specified, any subject who is randomized based on a wrong stratum will be analyzed according to the actual stratum the subject belongs to.

The Primary Analysis will be performed after all enrolled subjects have completed Part 1, Part 2 and Part 3, and the database has been locked. This will be the only and final efficacy analysis.

"Baseline" refers to the last non-missing observation prior to the first administration of study drug in Part 1 or Part 2 or prior to the randomization if no study drug is given.

Analysis of Categorical Variables:

For categorical variables, frequencies and percentages will be reported for each treatment group in Part 1, for overall for Part 2 and for each cohort in Part 3. Comparison of upadacitinib group and placebo group will be performed using the Cochran Mantel-Haenszel (CMH) test adjusting for stratification factors (baseline steroid use (Yes, No), endoscopic disease severity (SES-CD < 15, ≥ 15) and number of prior biologics with prior inadequate response or intolerance (≤ 1, > 1)). If there is a stratum for a treatment group that has no subject in it, a value of 0.1 will be added to all cells in the corresponding table in order to prevent dividing by 0, as suggested in Greenland and Robins (1985).[2] Point

estimates and 95% confidence intervals (CI) for the difference in proportions between upadacitinib group and placebo group will be provided. Construction of CI for the common risk difference will be based on the Mantel-Haenszel estimate adjusting for stratification factors.

For endpoints with small sample size (such as resolution draining fistulas at Week 12 in subjects with draining fistulas at Baseline) or small number of responders (e.g., occurrence of CD-related hospitalizations during Part 1), comparison of upadacitinib group and placebo group will be performed using the Chi-Square test (or Fisher's exact test if more than 20% of the cells have expected counts of less than 5).

Analysis of Continuous Variables:

For continuous variables, the model based mean and standard error will be provided. The mean value of the variable at baseline and each scheduled visit will also be presented for each treatment group. Continuous variables collected longitudinally will be using a Mixed-Effect Model Repeated Measurement (MMRM) model. Continuous efficacy variables which are collected at only one post-baseline visit (such as SES-CD) will be analyzed using an Analysis of Covariance (ANCOVA) model. Point estimates and 95% CIs of mean change from baseline within each treatment group, and the difference between upadacitinib group and placebo group will be provided.

9.2 Handling of Missing Data

Missing data could occur due to various reasons, including missing visits/assessments, early withdrawal from the clinical trials, or missing due to COVID-19 infection or logistic restrictions during the pandemic. Assessments on or after any intercurrent event will be handled as specified in Section 8.0.

The COVID-19 pandemic is interfering with the conduct of many ongoing trials, with potential impacts on treatment duration and the collection, analysis, and the interpretation of clinical trial data. For example, some scheduled visits may be missed due to self-quarantine or local government restrictions on travel; some visits may also be delayed

or canceled due to healthcare resource constraints during the pandemic. Impacted visits due to COVID-19 will be recorded in the database. The probability of having missed visits and missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic can be reasonably assumed to be unrelated to the unobserved values. Therefore, for the purpose of statistical analysis, it is reasonable to assume that these missing data are missing at random (MAR) and the statistical models that require MAR assumption are appropriate. The intent is to provide reliable estimates of the treatment effects under the scenario without the impact of COVID-19 pandemic.

Missing data for the efficacy analyses will be handled using the methods described below.

9.2.1 Categorical Endpoints

For binary efficacy endpoints other than the occurrence of hospitalization and the occurrence of CD-related surgery, missing data will be handled using the following approaches:

- The primary approach for handling missing data in the analysis of binary endpoints will use **N**on-**R**esponder **I**mputation (NRI) while incorporating Multiple Imputation (MI) to handle missing data due to **C**OVID-19 (NRI-C).
  - The NRI-C will categorize any subject who does not have an evaluation during a pre-specified visit window (either due to missing assessment or due to early withdrawal from the study) as a non-responder for the visit. The only exception is that missing data due to COVID-19 infection or logistical restriction will be handled by MI. At each visit, subjects will be characterized as responders or non-responders based on MI imputed values if missing due to COVID-19; otherwise, subjects will be considered as non-responders for missing due to other reasons in the NRI-C approach. In addition, at and after the CD-related corticosteroids intercurrent event (see Section 8.2) and on/after the date of initiation of CD-related confounding medications after premature discontinuation of study drug (see Section 8.1), subjects will be considered as non-responders and will not be imputed by MI.

- A sensitivity analysis for binary endpoints will use **NRI** with **N**o special data handling for missing due to **C**OVID-19 (NRI-NC).
  - NRI-NC will be performed in the same way as NRI-C without the exception above. Missing due to COVID-19 infection or logistical restriction will also be counted as non-responders. Subjects on or after the occurrence of the CD-related corticosteroids intercurrent event and on/after the date of initiation of CD-related confounding medications after premature discontinuation of study drug will still be considered as non-responders. This is the same method as "NRI" as defined in the protocol.
- Details on MI method in NRI-C are described below:
  - Markov Chain Monte Carlo (MCMC) will be first applied to augment data into monotonic missing pattern (where applicable) and PROC MI will be used to generate 30 datasets using the regression method. If the binary endpoints are derived from the continuous variables, PROC MI will be applied to the continuous variables. The variables to be included in the imputation model are: treatment group, randomization stratification factors (baseline steroid use (Yes, No), endoscopic disease severity (SES-CD < 15, ≥ 15) and number of prior biologics with inadequate response or intolerance (≤1, > 1)), age, gender, weight, baseline measurement (if applicable) and post-baseline measurements at each visit up to the end of the analysis period. The random seed for MCMC and the random seed for PROC MI are specified in Appendix I. The imputed post-baseline measurements will be rounded to the same precision as the observed data before the determination of responder status. Subjects will be characterized as responders or non-responders based on MI imputed data. Using the CMH model adjusted by randomization stratification factors, the endpoints will be analyzed using each of the 30 imputed datasets. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between upadacitinib group and placebo group, using Rubin's rule[3]. Note that measurements will be set to missing for 1) at or after the occurrence of the CD-related corticosteroids intercurrent event 2) on/after the date of initiation of CD-related confounding medications after premature discontinuation of study drug

before applying MI. After the MI imputation, an NRI override will be implemented for missing values 1) missing values due to reasons other than COVID-19 infection or logistic reasons, or 2) at or after the occurrence of the CD-related corticosteroids intercurrent event or 3) on/after the date of initiation of CD-related confounding medications after premature discontinuation of study drug, that is, regardless of MI imputed values, subjects satisfying 1) or 2) or 3) will be considered as "non-responder" for binary efficacy endpoints.

For binary efficacy endpoints regarding the occurrence of hospitalization and the occurrence of CD-related surgeries, missing data will be handled using the following approaches.

- As observed analysis (AO): The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study.

9.2.2 Continuous Endpoints

For continuous efficacy endpoints where Mixed-Effect Model Repeat Measurement (MMRM) analysis or Analysis of Covariance (ANCOVA) is performed, missing data will be handled using the following approaches.

- MMRM: The repeat measurement analysis will be conducted using a mixed model including observed measurements at all post-baseline visits up to the end of the analysis period. The mixed model includes the categorical fixed effects of treatment, visit and treatment-by-visit interaction, stratification factors at randomization, and the continuous fixed covariates of baseline measurement. An unstructured variance covariance matrix will be used. If the model cannot converge, an appropriate variance-covariance structure matrix (e.g., autoregressive or compound symmetry) will be used. Parameter estimation is based on the method of restrictive maximum likelihood (REML). MMRM will

be the primary approach in the analysis of continuous variables with repeated measurements.

- ANCOVA: For continuous efficacy variables that are collected at only one post-baseline visit (such as SES-CD), ANCOVA will be used. The model includes the categorical fixed effects of treatment, stratification factors at randomization, and the continuous fixed covariates of baseline and measurement.

9.2.3 Additional Analyses

In addition, As Observed (AO) and tipping point analysis will be performed as supplementary and sensitivity analyses, if appropriate, for categorical or continuous endpoints.

- AO: The AO analysis will not impute values for missing evaluations, and thus a subject who does not have an evaluation on a scheduled visit will be excluded from the AO analysis for that visit. AO will include all values collected in the study.
- Tipping Point Analysis for Binary Endpoints

  The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on each of the upadacitinib group and the placebo group can vary independently. The response rate among those subjects with missing response is assumed to be p0 for placebo group and p1 for upadacitinib group, and the response rate p0 and p1 systematically vary from 0% to 100% by every 10% respectively. Given a set of (p0, p1), the subjects with missing response will be randomly assigned as responders or non-responders using binomial distribution to generate 30 imputed datasets, and the same CMH method used for the primary analysis will be performed on each of the imputed datasets to obtain the results for each comparison between the upadacitinib group versus the placebo group. SAS PROC MIANALYZE will be used to generate the final inferences of the risk difference between upadacitinib group and placebo group, using Rubin's rule.[3]

  If one pair of (p0, p1) is found to reverse the study result interpretation in terms of p-value larger than 0.05, then the (p0, p1) is identified as the tipping

point. The results for a grid of (p0, p1) combinations are provided in tabular format. Note that missing data 1) on or after the occurrence of the CD-related corticosteroids intercurrent event or 2) on/after the date of initiation of CD-related confounding medications after premature discontinuation of study drug will be considered as non-responders and will not be imputed in the tipping point analysis.

- Tipping Point Analysis for Continuous Endpoints

  The tipping point analyses are two-dimensional, i.e., assumptions about the missing outcomes on each of the upadacitinib group and the placebo group can vary independently. In addition, the focus is on scenarios where missing outcomes on upadacitinib are worse than the imputed values on upadacitinib, while missing outcomes on placebo are better than the imputed values on placebo. Missing values are first imputed via MI under MAR assumption, and then a shift parameter is applied to the imputed values (a different shift parameter may be specified for each treatment group). This is implemented by PROC MI using the MNAR statement. The imputation uses a two-step approach, augmentation step using MCMC and imputation step using Monotone Regression. The MNAR statement is applied in the imputation step. The number of imputed datasets is 30.

  In cases where the shifted values are smaller than the minimum or larger than the maximum value of the endpoint, (i.e., out of range), the minimum or maximum value of the endpoint is used in further analysis steps. For each pair of shift parameters, the same MMRM/ANCOVA method used for the primary analysis will be performed on each of the imputed datasets to obtain the results for each upadacitinib treatment group versus the placebo group comparison. These results will be aggregated using Rubin's rule.[3]

- If a pair of shift parameters is found to reverse the study result interpretation in terms of p-value larger than 0.05, then the pair of shift parameters is identified as the tipping point.

9.3 Co-Primary Efficacy Endpoints and Analyses

9.3.1 Co-Primary Efficacy Endpoints

The co-primary endpoints for the primary analysis of efficacy are:

For EU/EMA regulatory purposes:

- Achievement of clinical remission per PROs at Week 12, and
- Achievement of endoscopic response at Week 12

For US/FDA regulatory purposes:

- Achievement of clinical remission per CDAI at Week 12, and
- Achievement of endoscopic response at Week 12

9.3.2 Primary Analysis of Co-Primary Efficacy Endpoints

The attributes of the estimands corresponding to the co-primary efficacy endpoints are summarized in Table 1.

The co-primary endpoints will be analyzed between upadacitinib group and placebo group using the CMH test, stratified by the randomization factors (baseline steroid use (Yes, No), endoscopic disease severity (SES-CD < 15, ≥ 15) and number of prior biologics with prior inadequate response or intolerance (≤ 1, > 1)) based on ITT1.

Table 1. Summary of the Estimand Attributes of the Co-Primary Efficacy Endpoints

Estimand Label	Attributes of the Estimand				
	Treatment	**Endpoint**	**Population**	**Intercurrent Events**	**Statistical Summary**
Co-Primary 1a	Upadacitinib 45 mg QD vs. placebo	Achievement of clinical remission per PROs at Week 12 (for EU/EMA)	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data after IE1 will be used until initiation of CD-related confounding medications All subjects will be considered as non-responders at or after IE2	Difference in the proportion of subjects achieving clinical remission per PROs
Co-Primary 1b	Upadacitinib 45 mg QD vs. placebo	Achievement of clinical remission per CDAI at Week 12 (for US/FDA)	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data after IE1 will be used until initiation of CD-related confounding medications All subjects will be considered as non-responders at or after IE2	Difference in the proportion of subjects achieving clinical remission per CDAI

Table 1. Summary of the Estimand Attributes of the Co-Primary Efficacy Endpoints (Continued)

Estimand Label	Attributes of the Estimand				
	Treatment	Endpoint	Population	Intercurrent Events	Statistical Summary
Co-Primary2	Upadacitinib 45 mg QD vs. placebo	Achievement of endoscopic response at Week 12	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data after IE1 will be used until initiation of CD-related confounding medications All subjects will be considered as non-responders at or after IE2	Difference in the proportion of subjects achieving endoscopic response

Both of the co-primary efficacy endpoints (EU/EMA: 1a and 2, US/FDA: 1b and 2 as specified in Table 1 above) will be tested at two-sided significance level of 0.05.
A CMH based two-sided 95% CI for the difference between treatment groups will be constructed.

The NRI-C approach for handling missing data will be used for main analysis for the co-primary endpoints.

9.3.3 Sensitivity and Supplementary Analyses of the Co-Primary Efficacy Endpoints

Sensitivity analyses of the primary analysis of the co-primary efficacy endpoints are NRI-NC and tipping point analysis.

In addition, sensitivity analyses which are based on "9-2 approach" (as detailed in Appendix D) to calculate CDAI score and average daily SF/AP will be performed on primary endpoints:

- Achievement of clinical remission per PROs at Week 12 (for EU/EMA regulatory purposes)
- Achievement of clinical remission per CDAI at Week 12 (for US/FDA regulatory purposes)

A supplementary analysis of the co-primary endpoints corresponded to the AO analysis specified in Section 9.2 will be performed in which all data after IE1 and IE2 will be used. CMH analysis as detailed in Section 9.1 will be repeated using AO data handling without any imputation as an additional analysis. The analysis will be conducted on the ITT1 population who have the efficacy measurement at Week 12 visit.

9.3.4 Additional Analyses of the Co-Primary Efficacy Endpoints

NRI-C analysis for the co-primary endpoints will also be conducted on the PP population. In this analysis, subjects with deviations that could potentially impact the analysis of primary endpoint will be excluded. The criteria will be fully defined in the classification plan. Exclusion of subjects will be adjudicated by the therapeutic area medical director (TAMD) and reasons for the subjects to be excluded will be documented and finalized before the Induction Study database lock for the primary analysis. Treatment difference between upadacitinib 45 mg QD and placebo with point estimate and 95% CI will be presented using NRI-C approach with the CMH method as detailed in Section 9.1 and Section 9.2.

9.4 Secondary Efficacy Analyses

9.4.1 Main Analyses of Ranked Secondary Efficacy Endpoints

The attributes of the estimands corresponding to the ranked secondary efficacy endpoints are summarized in Table 2.

In general, continuous secondary efficacy variables with repeated measurements will be analyzed using a MMRM model including factors for treatment group, visit, visit by treatment interaction, and stratification variables. The MMRM analysis is considered primary for inferential purposes. Continuous secondary efficacy variables which are collected at only one post-baseline visit (such as SES-CD) will be analyzed using ANCOVA approach.

Categorical secondary efficacy variables will be analyzed using the CMH test controlling for stratification factors (baseline steroid use (Yes, No), endoscopic disease severity (SES-CD $< 15, \geq 15$) and number of prior biologics with prior inadequate response or intolerance ($\leq 1, > 1$)). A CMH based two-sided 95% CI for the difference between treatment groups will be constructed. NRI-C method for missing data will be used for categorical secondary endpoints except occurrence of hospitalizations due to CD during Part 1. The endpoints of "Occurrence of hospitalization due to CD during Part 1" will be analyzed using Chi-square test (or Fisher's exact test if $\geq 20\%$ of the cells have expected cell count < 5).

The NRI-C method will be the primary approach for missing data handling for categorical secondary endpoints except hospitalization due to CD during Part 1. The endpoint of "Occurrence of hospitalization due to CD during Part 1" will be analyzed based on the observed data using AO approach only.

Table 2. Summary of the Estimand Attributes of the Key Secondary Efficacy Endpoints

	Attributes of the Estimand				
	Treatment	**Endpoint**	**Population**	**Handling Intercurrent Events**	**Statistical Summary**
Categorical Key Secondary except occurrence of hospitalizations due to CD during Part 1	Upadacitinib 45 mg QD vs. placebo	Achievement of clinical remission per CDAI (EU/EMA) or per PROs (US/FDA) at Week 12, endoscopic remission at Week 12, discontinuation of corticosteroid use for CD and achievement of clinical remission per PROs (EU/EMA) or per CDAI (US/FDA) at Week 12, CR-100 at Week 12, resolution of EIM at Week 12, clinical remission per PROs (EU/EMA) or per CDAI (US/FDA) at Week 4, CR-100 at Week 2	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data after IE1 will be used until initiation of CD-related confounding medications All subjects will be considered as non-responders at or after IE2	Difference in the proportion of subjects achieving each binary secondary endpoint

Table 2. Summary of the Estimand Attributes of the Key Secondary Efficacy Endpoints (Continued)

	Attributes of the Estimand				
	Treatment	**Endpoint**	**Population**	**Handling Intercurrent Events**	**Statistical Summary**
Occurrence of hospitalizations due to CD during Part 1	Upadacitinib 45 mg QD vs. placebo	Occurrence of hospitalizations due to CD during Part 1	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part 1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data will be used regardless of IE	Difference in the proportion of subjects with occurrence of hospitalizations due to CD during Part 1
Continuous Key Secondary	Upadacitinib 45 mg QD vs. placebo	Change from Baseline in FACIT-F, IBDQ at Week 12	ITT1 population (Subjects who were randomized and received at least one dose of study drug in Part 1)	IE1: premature discontinuation of study drug IE2: Initiation or dose escalation of CD-related corticosteroids All data after IE1 will be used until initiation of CD-related confounding medications All data after IE2 will not be used for continuous endpoints	Difference in the mean change from Baseline in IBDQ total score and FACIT-F score

9.4.2 Sensitivity and Supplementary Analyses for Key Secondary Efficacy Endpoints

For categorical key secondary endpoints except occurrence of hospitalizations due to CD during Part 1, sensitivity analyses of the of the main analyses of are NRI-NC, and tipping point analysis. NRI-NC and tipping point analysis will be performed for all categorical key secondary endpoints except occurrence of hospitalizations due to CD during Part 1.

For continuous key secondary endpoints, sensitivity analysis of the of the main analysis is tipping point analysis.

A supplementary analysis of categorical key secondary endpoints except occurrence of hospitalizations due to CD during Part 1 corresponded to the AO analysis specified in Section 9.2 will be performed in which all data after IE1 and IE2 will be used. CMH analysis as detailed in Section 9.1 will be repeated using AO data handling without any imputation as an additional analysis. The analyses will be conducted on the ITT1 population who have the efficacy measurement at corresponding visit for the endpoint.

9.5 Additional Efficacy Analyses

Part 1

Other efficacy endpoints will be analyzed similarly as the secondary efficacy endpoints as described in Section 9.4. No multiplicity adjustment will be considered.

Part 2 and Part 3

Only descriptive statistics will be provided for the Part 2 and Part 3 endpoints specified in Section 3.3. The analyses for Part 2 will be performed using ITT2, and the analyses for Part 3 will be performed using the ITT3 by cohort. Categorical endpoints will be summarized by number, percent and two-sided 95% CI; and continuous endpoints will be summarized by n, mean, standard deviation, 95% CI, minimum value, median, and

maximum value; and time to response endpoints will be summarized by n, number of events, and the 25%, 50%, 75% percentiles of time (in days).

A programming error from the ediary vendor was identified in the total number of daily stools question on patient ediaries that does not allow the transaction to progress if the answer 0 to question 1, "How many stools in total have you had in the last 24 hours from last record until now?" The vendor has confirmed the issue was only present in languages that are translated from right to left. For this program, that includes Hebrew and Arabic in Israel and Egypt, respectively.

Since data from total stool frequency is only applicable to non-ranked additional efficacy endpoints, a sensitivity analysis will be performed for the additional efficacy endpoint of change from Baseline in average daily total stool frequency at all scheduled visits, by excluding subjects from Egypt and Israel.

9.6 Efficacy Subgroup Analyses

To evaluate the consistency of the efficacy over demographic and other baseline characteristics, subgroup analysis will be performed for the co-primary endpoints for the subgroups listed in Table 3 below using ITT1.

For subgroup analysis, point estimate and 95% CI for each treatment group as well as point estimate and 95% CI for treatment differences between upadacitinib group and placebo group will be presented. No p-value will be provided. If, for a subgroup factor except age, sex and race (Table 3), any of the resulting categories has fewer than 10% of the planned study size, subgroup analyses will not be presented for that category.

Table 3. Subgroups for Efficacy Analysis

Subgroup Factor	Categories
Sex	Male, Female
Age	≤ median, > median
Age	≥ 18 years - < 40 years, ≥ 40 years - < 65 years, ≥ 65 years
Race	White, non-White
Race	White, Black or African American, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Multiple
Baseline fecal calprotectin	≤ 250 µg/g, > 250 µg/g
Baseline CD-related Steroids Use	Yes, No
Baseline CD-related immunosuppressant use	Yes, No
Baseline CD-related aminosalicylate use	Yes, No
Baseline CD-related antibiotic use	Yes, No
Baseline hs-CRP	≤ 5 mg/L, > 5 mg/L
Baseline CDAI	≤ 300, > 300
Baseline very soft/liquid SF	≤ median, > median
Baseline AP score	≤ median, > median
Baseline SES-CD	< 15, ≥ 15
Crohn's Disease Location at Baseline	Colonic only, Ileal only, Ileal-colonic
Body Weight	≤ median, > median
BMI	Underweight [< 18.5 kg/m^2], Normal [≥ 18.5 and < 25 kg/m^2], Overweight [≥ 25 and < 30 kg/m^2], Obese [≥ 30 kg/m^2]
Disease Duration	≤ 5 years, > 5 years
Geographic Region	North America, Western Europe, South/Central America, Eastern Europe, Asia, Other
Number of Prior Biologics Failed	≤ 1, > 1
Prior TNF Failure	1, > 1
Prior Ustekinumab Failure	Yes or No
Prior Vedolizumab/Natalizumab Failure	Yes or No

10.0 Safety Analyses

10.1 General Considerations

Safety analyses will be performed on the safety population for Part 1 (SA1) and SA-UPA population as defined in Section 4.0. In addition, safety summaries will be performed on the safety analysis set for Part 2 (SA2) and Part 3 (SA3).

The standard safety analyses will include reporting of adverse events (AEs), adverse events of special interest (AESIs), laboratory, and vital signs measurements. Frequency tables and exposure adjusted event rate per 100 patient-years tables of subjects with treatment-emergent adverse events (TEAEs) by system organ class (SOC) and by preferred term (PT) as in the Medical Dictionary for Regulatory Activities (MedDRA) dictionary will be provided by treatment group. All continuous laboratory parameters and vital signs variables at each visit will also be summarized by treatment group. Frequency tables of subjects meeting criteria for potentially clinically important vital sign values and for potentially clinically important laboratory values will be provided by treatment group.

The baseline for safety analysis will be treatment dependent. For the Part 1/Part 2 laboratory and vital signs measurements, the baseline value is defined as the last available measurement before study drug administration for all subjects. Starting from the Part 3, the baseline value for subjects initially randomized to upadacitinib will remain the same. However, the baseline value for subjects initially randomized to placebo will be redefined as the last available measurement before first dose of upadacitinib administration in Part 3.

For the safety analysis, subjects are assigned to a treatment group based on the treatment actually received, regardless of the treatment randomized.

Missing safety data will not be imputed.

10.2 Adverse Events

AEs will be summarized and presented using primary MedDRA SOCs and PTs according to the version of the MedDRA coding dictionary used for the study at the time of database lock. The actual version of the MedDRA coding dictionary used will be noted in the AE tables and in the clinical study report. Specific AEs will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same adverse event occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported.

10.2.1 Treatment-Emergent Adverse Events

Part 1 (SA1 population) /Part 2 (SA2 population): TEAEs for Part 1/Part 2 are defined as events that begin either on or after the first dose of the study drug in Part 1/Part 2 and until (i) the first dose of study drug in the Study M14-430 (if applicable), or (ii) until first dose of study drug in Part 3 (if applicable), or (iii) within 30 days after the last dose administration of the study drug in Part 1/Part 2, whichever is earlier.

Part 3 (SA3 population): TEAEs for Part 3 are defined as events that begin either on or after the first dose of the study medication in the Part 3 and until (i) first dose of study drug in the Study M14-430 (if applicable), or (ii) within 30 days after the last dose of the study drug in the Part 3, whichever is earlier.

All Upadacitinib (SA-UPA population): TEAEs for "All Upadacitinib" are defined as events that begin either on or after the first dose of upadacitinib and (i) until first dose of study drug in the Study M14-430 (if applicable), or (ii) within 30 days after the last dose of upadacitinib in the study, whichever is earlier.

24-Week Upadacitinib (SA4 population): TEAEs for "24-Week Upadacitinib" (SA4) are defined as events that begin either on or after the first dose of upadacitinib and until (i) first dose of study drug in the Study M14-430 (if applicable), or (ii) within 30 days after the last dose of upadacitinib in the study, whichever is earlier.

If an incomplete onset date was collected for an adverse event, the event will be assumed to be treatment-emergent unless there is other evidence that confirms that the event was not treatment-emergent (e.g., the event end date was prior to the study drug start date). The number and percentage of subjects experiencing treatment-emergent AEs will be summarized. For Part 1, the risk difference and 95% CI between upadacitinib group and placebo group will be presented.

10.2.2 Adverse Event Overview

An overview of AEs will be presented consisting of the number and percentage of subjects experiencing at least one event for each of the following AE categories:

- Any TEAE
- Any TEAE related to COVID-19
- Any TEAE that was rated as related to study drug by the investigator.
- Any severe TEAE
- Any serious TEAE.
- Any TEAE leading to discontinuation of study drug.
- Any TEAE leading to death.
- Treatment-emergent AESI (Appendix B).
- All deaths
    - Deaths occurring ≤ 30 days after last dose of study drug
    - Deaths occurring > 30 days after last dose of study drug
    - Deaths related to COVID-19

In addition, an overview of AEs per 100 patient-years of study exposure will be presented for the AE categories defined above. The number of TEAEs reported, the total number of years of study drug exposure, and the TEAE rate per 100 patient-years will be presented.

10.2.3 Treatment-Emergent Adverse Events by SOC and/or PT

TEAEs will be summarized by SOC and PT; by maximum relationship to study drug as assessed by the investigator (e.g., reasonable possibility or no reasonable possibility) and SOC and PT; by maximum severity and SOC and PT; and by subject number and SOC and PT. Specific AEs will be counted once for each subject for calculating percentages, unless stated otherwise. In addition, if the same AE occurs multiple times within a subject, the highest severity and level of relationship to investigational product will be reported. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

For summary by maximum severity, if a subject has an AE with an unknown severity, then the subject will be counted in the severity category of unknown, even if the subject has another occurrence of the same event with a severity present. The only exception is if the subject has another occurrence of the same AE with the most extreme severity – severe. In this case, the subject will be counted under the severe category.

For summary by maximum relationship to study drug, if a subject has an AE with an unknown relationship, then the subject will be counted in the relationship category of "unknown," even if the subject has another occurrence of the same event with a relationship present. The only exception is if the subject has another occurrence of the same AE with a relationship assessment of "Reasonable Possibility." In this case, the subject will be counted under the "Reasonable Possibility" category.

In addition, TEAEs will be summarized by PT and sorted by decreasing frequency for the upadacitinib 45 mg QD treatment group.

10.2.4 Treatment-Emergent Adverse Events per Patient-Years of Exposure

TEAEs will be summarized by event rate per 100 subject years, defined as

$$100 \times \frac{Number\ of\ TEAEs}{Total\ Patient\ Years}$$

where total patient years is defined as the sum of the study drug exposure of all subjects, normalized by 365.25, and rounded to 1 decimal place. Note that one event per preferred term per day per subject will be counted in the calculation of the number of AEs (i.e., a preferred term will not be counted twice on the same day for the same subject).

The TEAE rates per 100 patient-years of exposure will be provided for each AE category in the AE overview summary (defined in Section 10.2.2) and for TEAE summary by SOC and PT.

The study drug exposure for each safety population is defined in the table below.

Safety Population	Scenario	Study Drug Exposure Definition
SA1, SA2	Enrolls in Study M14-430 or Part 3 at the end of Part 1/Part 2	Last dose date in Part 1/Part 2 - 1st dose date in Part 1/Part 2 + 1
	Does not participate in the Study M14-430 or Part 3	
SA3	Enrolls in Study M14-430	Last dose date in Part 3 - 1st dose date in Part 3 + 1
	Does not participate in the Study M14-430	
SA4	Enrolls in Study M14-430	Last dose date in Study M14-431 - 1st dose date in Part 1/Part 2 + 1
	Does not participate in the Study M14-430	
SA-UPA	Enrolls in Study M14-430	Last dose date of upadacitinib in Study M14-431 - 1st dose date in upadacitinib in M14-431 + 1
	Does not participate in the Study M14-430	

10.2.5 SAEs (Including Deaths) and Adverse Events Leading to Study Drug Discontinuation

SAEs (including deaths) and AEs leading to study drug discontinuation will be summarized by SOC and PT and in listing format.

10.2.6 Adverse Events of Special Interest

The AESI categories will be identified by the search criteria per Standard MedDRA Queries (SMQs)/Company MedDRA Queries (CMQs) specified in Appendix B.

Treatment-emergent AESIs will be summarized by SOC and PT and listing format. Additionally, AESI rates per 100 patient years of exposure will be provided for each AESI category (Appendix B) in the AE overview summary.

10.3 Analysis of Laboratory Data

Data collected from central and local laboratories, including additional laboratory testing due to an SAE, will be used in all analyses, except for baseline where SAE-related laboratory assessments on or before the first dose of study drug will be excluded. The clinical laboratory tests defined in the protocol (e.g., hematology and clinical chemistry) will be summarized.

Analysis of Quantitative Laboratory Parameters

Each laboratory variable will be summarized for all time points (starting at baseline) with the number of non-missing observations, mean, standard deviation, median, minimum and maximum. Mean change from Baseline to each applicable post-baseline visit will be summarized for selected laboratory variables, with the number of observations, baseline mean, and visit mean. The change from baseline mean, standard error, and 95% CI will be presented for the mean change from Baseline within each treatment group.

For SA1 population, treatment group differences between upadacitinib group and placebo group for changes from baseline will be analyzed using a one-way Analysis of Variance (ANOVA) model with treatment as a fixed factor and 95% CI for treatment difference will be presented for selected laboratory parameters.

Shift Table Analyses

Changes in laboratory parameters will be tabulated using shift tables categorized as low, normal, or high based on the normal ranges of the laboratory used for each sample. A shift table from baseline to minimum and maximum value (based on normal range) will be created. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value.

Potentially Clinically Significant Laboratory Values

The criteria for potentially clinically significant laboratory values will be determined by CTCAE criteria of Grade 3, Grade 4 and ≥ Grade 3, with a grade worsening compared to baseline. Toxicity grading scale is based on National Cancer Institute Common Toxicity Criteria (NCI CTC) AE version 4.03. The number and percentage of subjects meeting the criteria for potentially clinically significant laboratory values will be summarized by treatment group. Listings of subject-level laboratory data will be provided for subjects meeting the criteria.

Assessment of Liver Elevations

The liver-specific laboratory tests include the serum glutamic pyruvic transaminase (SGPT/ALT), serum glutamic-oxaloacetic transaminase (SGOT/AST), alkaline phosphatase (ALP), and total bilirubin (TBL). The frequencies and percentages of subjects with post baseline liver specific function test values that meet the following criteria of potential clinical interest will be summarized by treatment group:

- ALT > 3 × ULN
- ALT > 5 × ULN
- ALT > 10 × ULN
- ALT > 20 × ULN
- AST > 3 × ULN
- AST > 5 × ULN

- AST > 10 × ULN
- AST > 20 × ULN
- TBL > 2 × ULN
- ALP > 1.5 × ULN
- ALT and/or AST > 3 × ULN and TBL > 1.5 × ULN
- ALT and/or AST > 3 × ULN and TBL > 2 × ULN

where ULN is the upper normal limit. The maximum ratio relative to the ULN will be used to determine if subjects meet any of the criteria listed above.

A listing of possible Hy's Law cases, defined as those meeting the following criteria will also be provided:

- ALT and/or AST > 3 × ULN and TBL > 1.5 × ULN

In addition, eDISH plots will be created displaying post-baseline total bilirubin versus post-baseline ALT, in terms of the maximum ratio relative to the ULN (not necessarily concurrent).

10.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse rate, body weight, respiratory rate and body temperature will be summarized.

Each vital sign variable will be summarized for all time points (starting with Baseline) with the number of non-missing observations, mean and standard deviation, median, minimum and maximum. Mean change from Baseline to each applicable post-baseline visit will be summarized for each vital sign variable, with the number of observations, Baseline mean, and visit mean. The change from Baseline mean, standard error, and 95% CI will be presented for the mean change from Baseline within each treatment group.

For SA1 population, treatment group differences between upadacitinib group and placebo group for changes from Baseline will be analyzed using a one-way Analysis of Variance

(ANOVA) with treatment as a fixed factor and 95% CI for treatment difference will be presented for each vital sign variable.

Vital sign variables will be evaluated based on potentially clinically important (PCI) criteria (Appendix C). For each vital sign PCI criterion, the number and percentage of subjects who have a vital sign value meeting the criteria will be summarized. Listings will be provided to summarize subject-level vital sign data for subjects meeting PCI criteria.

11.0 Interim Analyses

There will be no efficacy interim analyses planned for this study.

11.1 Data Monitoring Committee

An external data monitoring committee (DMC) composed of persons independent of AbbVie and with relevant expertise in their field will review unblinded safety data from the ongoing study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

A separate DMC charter describes the roles and responsibilities of the DMC members, frequency of data reviews, relevant data to be assessed, and general operations.

Since there are no efficacy analyses for early stopping, no alpha adjustment is needed.

12.0 Overall Type-I Error Control

The overall type I error rate of the co-primary and ranked secondary endpoints will be strongly controlled using a fixed sequence multiple-testing procedure[1] as well as Holm procedure[5] as described below. Specifically, the testing will utilize the sequence of hypothesis testing for the co-primary endpoints followed by a set of ranked key secondary endpoints (Group 1A_E for EU/EMA, Group 1A_F for US/FDA) in the order as specified in Section 3.2, and will begin with testing each of the co-primary endpoints using two-sided α of 0.05. If both of the co-primary endpoints are statistically significant at

two-sided 0.05 level of significance, the Group 1A (Group 1A_E for EU/EMA, Group 1A_F for US/FDA) ranked secondary endpoints will be tested at two-sided α of 0.05 in the order specified in Section 3.2. If the co-primary endpoints and the Group 1A (Group 1A_E for EU/EMA, Group 1A_F for US/FDA) ranked secondary endpoints are statistically significant at two-sided 0.05 level of significance, the Group 1B (Group 1B_E for EU/EMA, Group 1B_F for US/FDA) ranked secondary endpoints will be tested using the Holm procedure at two-sided α of 0.05. The graphs for the testing procedures are provided in Figure 3 (for EU/EMA regulatory purpose) and Figure 4 (for US/FDA regulatory purpose).

Figure 3. Graphical Multiple Testing Procedure for Co-Primary and Ranked Secondary Endpoints for EU/EMA

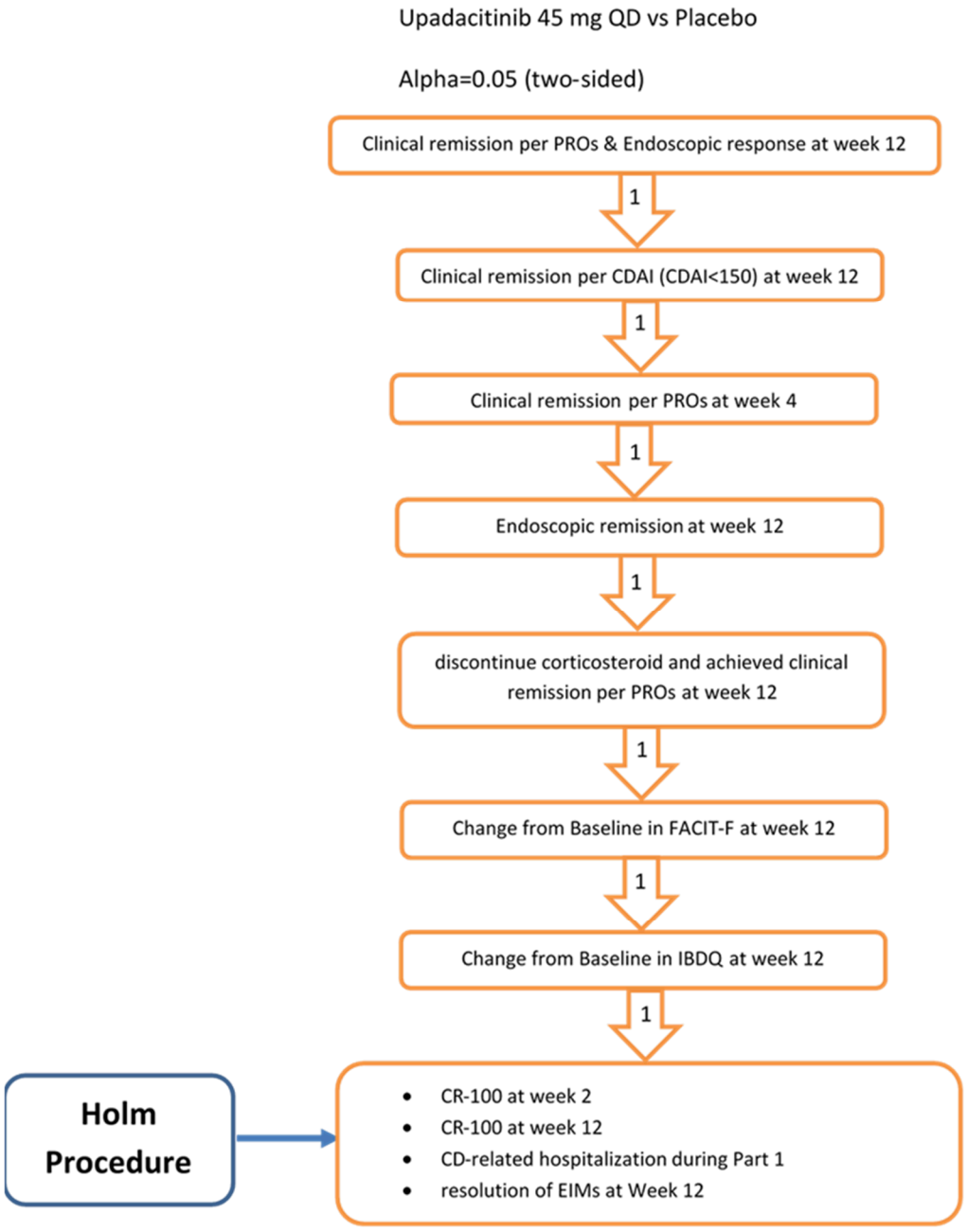
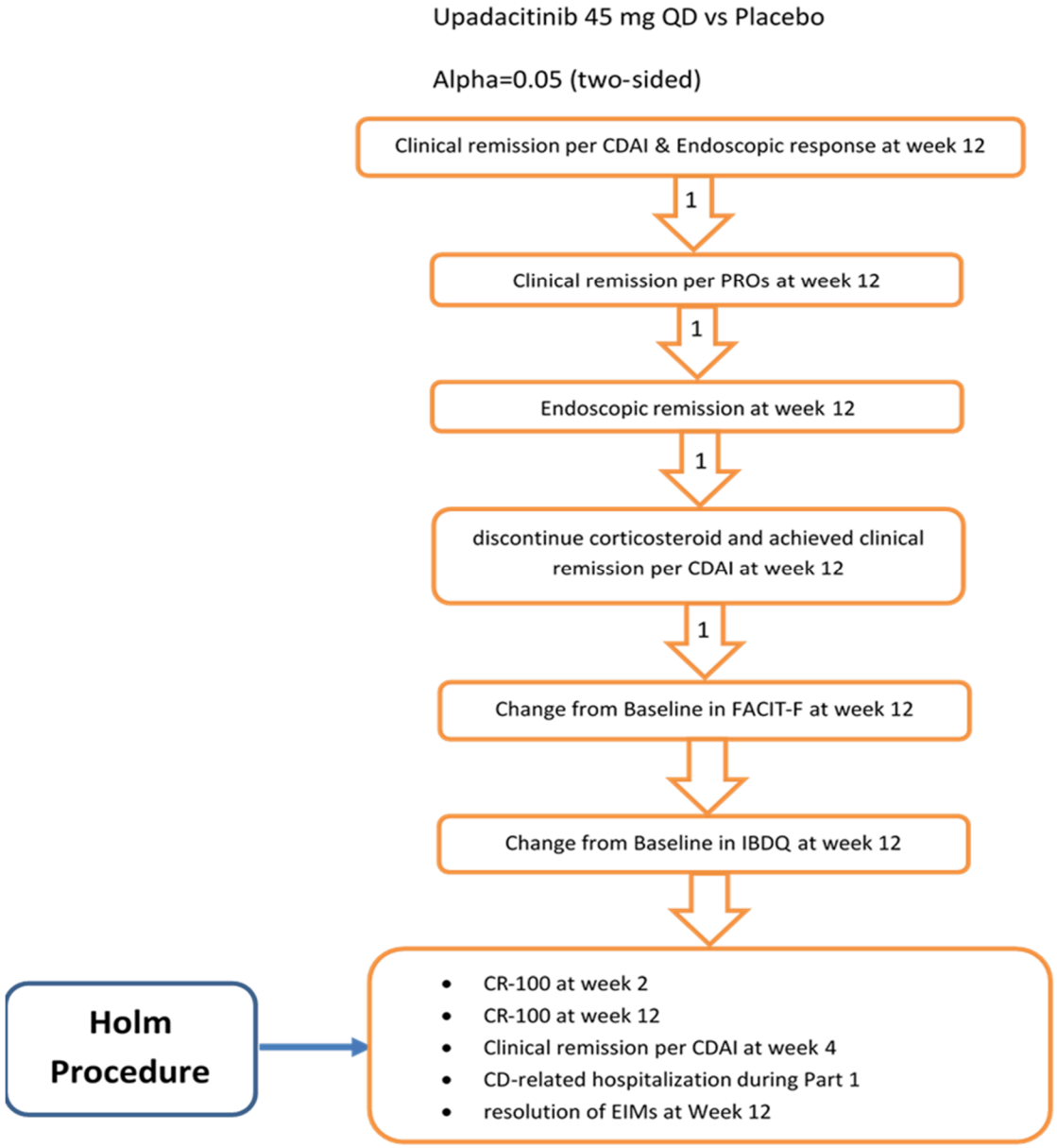


Figure 4. **Graphical Multiple Testing Procedure for Co-Primary and Ranked Secondary Endpoints for US/FDA Graphical Multiple Testing Procedure for Co-Primary and Ranked Secondary Endpoints for US/FDA**



13.0 Version History

Table 4. SAP Version History Summary

Version	Date	Summary
1.0	05 Oct 2017	Original version
2.0	29 Jan 2018	• Clarification provided on Part 1 efficacy endpoints; specified Part 2 and Part 3 efficacy endpoints (Section 4.3). • Modified the graphical testing procedure to incorporate Holm procedure for the Group 1B endpoints in place of Hochberg procedure (Section 4.6). • Race and Region subgroup categories added. Details and clarification added for subgroup analyses (Section 5.2). • Specified the primary analysis for inference purposes for the continuous secondary efficacy endpoints (Section 6.3).
3.0	28 Jan 2021	• Revised the SAP per the latest SAP template. • The following changes have occurred in order to reflect changes in the protocol amendment and regulatory guideline. o Updated study design based on protocol amendments to reduce the number of subjects planned for open-label induction Part 2 (from 360 to 150) o Updated co-primary endpoint and ranked secondary endpoints to address agency requests and reflect updates in protocol amendments o Added NRI-C method for handling missing data due to COVID-19. NRI-C will be used for the primary efficacy analysis and NRI-NC will be considered as sensitivity analysis o Removed the Last Observation Carried Forward (LOCF) approach from the missing data imputation as LOCF potentially can result in a biased estimation of treatment effect and underestimate the variability • Added definitions of estimand for primary and key secondary endpoints.

Table 4. SAP Version History Summary (Continued)

Version	Date	Summary
4.0	23 Jul 2021	• Add premature withdrawal from the study as an intercurrent event. • Replace Pattern Mixture Model (PMM) with tipping point analysis as the sensitivity analyses for all primary and key secondary endpoints. • Update the cutoff of CGHAS/IGHAS histology remission endpoint (defined as CGHAS/IGHAS score ≤ 2) where cutoff is updated from "≤ 3" to "≤ 2" to align with published literature (Section 3.3.1/ Section 3.3.2). • Update criteria of liver function assessment from "≥" to ">" (Section 10.3). • Add and define SA4 population (Section 4.0) and corresponding TEAE for the safety analysis (Section 10.2.1) in subjects who took upadacitinib in both 12-Week Induction Period and Extended Treatment Period • Add three Other Efficacy Endpoints (Section 3.3) on clinically meaningful improvements in FACIT-F • Add details of calculation of total patient year in the exposure adjusted AE analysis (Section 10.2.4) • Delete "endoscopic remission at Week 12 among subjects with SES-CD subscore of 3 in the narrowing component at Week 12, which was not present at Baseline" (result in 0% remission rate) and "endoscopic remission or endoscopic response at Week 12 among subjects with SES-CD subscore of 3 in the narrowing component at Week 12, which was not present at Baseline" as they are not applicable by definition • Replace ANCOVA-C with ANCOVA as ANCOVA is valid under MAR missingness (Section 9.2.2) • Corrected typo on the TEAE definition for SA-UPA population (update "40 days" to "30 days," Section 10.2.1)

Table 4. SAP Version History Summary (Continued)

Version	Date	Summary
5.0	15 Nov 2021	• Based on FDA guidance (E9(R1) addendum Revision 1, May 2021), removed study withdrawal from the scope of intercurrent events (Section 2.1, Section 8.0, Section 9.3.2 Table 1, Section 9.4.1 Table 2) • Added a sensitivity analysis to the primary endpoints in Section 9.3.3 per agency request, and provide details on derivation of total CDAI score, average daily SF/AP (Section 3.0, Appendix D) • Add five Other Efficacy Endpoints (Section 3.3.1, Section 3.3.3) on in (1) change in FACIT-F at Week 4 and Week 12; (2) resolution of EIM at Week 2, Week 4, Week 8 and Week 12; (3) achievement of CR-100 at Week 16, Week 20 and Week 24; (4) time to CR-100; (5) achievement of endoscopic response at Week 24. • Provided more clarity that baseline characteristics will also be summarized for ITT2 (Section 7.0) • Clarified intercurrent event handling strategies (Section 8.1 and Section 8.2) • Updated the category for Race to adhere with CDISC controlled terminology, and corrected a typo in the Baseline SES-CD variable (update "≤ 15, > 15" to "< 15, ≥ 15") in subgroup analysis (Section 7.1, Section 9.6 Table 3) • Provided more clarity on the concomitant medication definition (Section 7.3) • Provided more clarity on the scope of summary statistics for the Extended Treatment Period (Section 9.1) • Provided more clarity on handling premature discontinuation of study drug intercurrent event in NRI-C analysis (Section 9.2.1) • Provided more clarity on the analyses for time to response endpoints (Section 9.5) • Add sensitivity analyses on additional efficacy endpoint of change from Baseline in average daily total stool frequency (Section 9.5). • Provided more clarity on the TEAE definition, and also the scope of AE overview (Section 10.2.1, Section 10.2.2)

14.0 References

1. Bretz F, Maurer W, Brannath W, et al. A graphical approach to sequentially rejective multiple test procedures. Stat Med. 2009;28(4):586-604.

2. Greenland S, Robins JM. Estimation of a common effect parameter from sparse follow-up data. Biometrics. 1985;41(1):55-68.
3. Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley & Sons, Inc. 1987.
4. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. Psychol Methods. 1997;2(1):64-78.
5. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat. 1979;6:65-70.
6. Qu Y, Shurzinske L, Sethuraman S. Defining estimands using a mix of strategies to handle intercurrent events in clinical trials. Pharm Stat. 2021;20(2):314-23.

Appendix A. Protocol Deviations

The number and percentage of subjects who reported at least one of the following protocol deviation categories will be provided.

- Subject entered into the study even though s/he did not satisfy entry criteria.
- Subject developed withdrawal criteria during the study and was not withdrawn.
- Subject received wrong treatment or incorrect dose of study.
- Subject took prohibited concomitant medication.

Appendix B. Definition of Adverse Events of Special Interest

AESI will be identified by the following CMQ, SMQ, and other search criteria:

Table B-1. AESI for Upadacitinib with SMQs/CMQs/PTs Searches

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Serious Infections	CMQ		"Infections" – Subset for SAEs
Opportunistic Infections excluding Tuberculosis and Herpes Zoster	CMQ		"Opportunistic Infection excluding tuberculosis and herpes zoster"
Active Tuberculosis	CMQ		"Active Tuberculosis"
Herpes Zoster	CMQ		"Herpes Zoster"
Adjudicated Gastrointestinal Perforations	Based on adjudicated results (per GI Perforation charter) of events identified by the "Gastrointestinal Perforation" SMQ Narrow search		
Anemia	CMQ		"Non-Hemolytic and Non-Aplastic Anemias"
Neutropenia	CMQ		"Hematological Toxicity – Neutropenia"
Lymphopenia	CMQ		"Hematological Toxicity – Lymphopenia"
Creatine Phosphokinase (CPK) Elevations	PT		Search for the PT of "Blood creatine phosphokinase increased."
Hepatic Disorders	SMQ	Narrow	"Drug Related Hepatic Disorders – Comprehensive Search"
Renal Dysfunction	SMQ	Narrow	"Acute Renal Failure"
Malignancies (all types)	SMQ		"Malignant tumors"

AESI	Type of MedDRA Query	Broad or Narrow Search	SMQ/CMQ Search Criteria
Malignancies excluding non-melanoma skin cancer (NMSC)	SMQ	Narrow	Malignant Tumors SMQ (Narrow) removing NMSC output
NMSC	SMQ	Narrow	Skin Malignant tumors (Narrow SMQ) removing Melanoma CMQ
Lymphoma	SMQ		"Malignant Lymphomas"
Adjudicated Cardiovascular Events	Output from CAC[a]		
MACE*			
Cardiovascular Death			
Non-fatal Myocardial Infarction			
Non-fatal Stroke			
Other Cardiovascular Events			
Undetermined/Unknown Cause of Death			
Adjudicated Thrombotic Events	Output from CAC[a]		
VTE**			
Deep Vein Thrombosis			
Pulmonary Embolism			
Other Venous Thrombosis			
Arterial Thromboembolic Events (non-cardiac, non-neurologic)			

CAC = Cardiovascular Adjudication Committee; CMQ = company MedDRA query; PT = preferred term; SMQ = standard MedDRA query

* MACE: Major Adverse Cardiovascular Events, defined as cardiovascular death, non-fatal myocardial infarction and non-fatal stroke.

** VTE: Venous thromboembolic events, defined as deep vein thrombosis (DVT) and pulmonary embolism (PE) (fatal and non-fatal).

a. Reviewed and adjudicated by an independent Cardiovascular Adjudication Committee in a blinded manner.

Table B-2. Exposure-Adjusted Rate Analyses for AESI

Analysis	EAER*	EAIR**	Cumulative Inc. Prop.
AESI (Section 10.2.6)			
Serious Infections	X	X	X
Opportunistic Infections excluding Tuberculosis and Herpes Zoster	X	X	X
Active Tuberculosis	X	X	X
Herpes Zoster	X	X	X
Adjudicated Gastrointestinal Perforations	X	X	
Anemia	X	X	X
Neutropenia	X	X	X
Lymphopenia	X	X	X
Creatine Phosphokinase (CPK) Elevations	X	X	
Hepatic Disorders	X	X	
Renal Dysfunction	X	X	
Malignancies		X	X
Malignancies excluding Non-Melanoma Skin Cancer (NMSC)		X	X
NMSC		X	X
Lymphoma		X	X
Adjudicated Cardiovascular Events		X	X
Adjudicated Thrombotic Events		X	X

* EAER: Exposure-adjusted event rate.

** EAIR: Exposure-adjusted incidence rate.

Appendix C. Potentially Clinically Important Criteria for Safety Endpoints

The criteria for Potentially Clinically Important (PCI) criteria for vital sign findings are described in Table C-1.

Table C-1. Criteria for Potentially Clinically Important Vital Sign Values

Vital Sign	Category	Criteria for Potential Clinically Significant Vital Signs
Systolic blood pressure	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline
Pulse	Low	Value ≤ 50 bpm and decrease ≥ 15 bpm from Baseline
	High	Value ≥ 120 bpm and increase ≥ 15 bpm from Baseline
Weight	High	> 7% increase from baseline
	Low	> 7% decrease from baseline

Appendix D. Crohn's Disease Activity Index (CDAI)

The Crohn's Disease Activity Index (CDAI) is a composite instrument that includes patient symptoms evaluated over 7 days (abdominal pain, stool frequency and general well-being), as well as physical and laboratory findings. These items are scored individually, weighted, and do not contribute equally to the overall score. The CDAI is derived from summing up the weighted individual scores of eight items as outlined below:

			Factor	Subtotal
1. Number of liquid or very soft stools (Record the frequency per day)	__+___+___+___+___+___+___+___=___ Days: 1 2 3 4 5 6 7 Sum	×	2	
2. Abdominal pain rating: 0 = none, 1 = mild, 2 = moderate, 3 = severe	__+___+___+___+___+___+___+___=___ Days: 1 2 3 4 5 6 7 Sum	×	5	
3. General well-being: 0 = generally well, 1 = slightly underpar, 2 = poor, 3 = very poor, 4 = terrible	__+___+___+___+___+___+___+___=___ Days: 1 2 3 4 5 6 7 Sum	×	7	
4. Number of 6 listed categories the subject now has **Check all items that apply:** ☐ Arthritis/arthralgia ☐ Iritis/uveitis ☐ Erythema nodosum/pyoderma gangrenosum/aphthous stomatitis ☐ Fissure, abscess and/or anal fistula (draining/non-draining) ☐ Other cutaneous fistula (draining/non-draining) Fistula ☐ Fever over 100°F (37.8°C) during past week	_____ _____ Record "0" if no categories checked	×	20	
5. Taking Lomotil/Imodium/ Loperamide/opiates for diarrhea 0 = no, 1 = yes	_____	×	30	
6. Abdominal mass 0 = none, 2 = questionable, 5 = defined	_____	×	10	
7. Hematocrit: ___.__	Male: (47 – hematocrit) = Female: (42 – hematocrit) = Subtotal If hematocrit > normal, enter "0"	×	6	
8. Body weight: ___.__(kg) Standard weight: ___.__(kg)	100 × [1 – (Body wt/Standard wt)] = Percent below standard weight: _____ If body wt > std. wt, enter "0"	×	1	
			Total	

Higher CDAI scores indicate more severe disease. For the calculation of CDAI and components (daily stool frequency, abdominal pain score and general well-being) at each visit, the following rules will be applied:

1. Identify the visit date with CDAI components that are collected, and set it as the CDAI calculation date.
2. Calculate the subtotal scores from the component 1-3 (liquid or very soft stool frequency, abdominal pain score and general well-being) as follows:
   a. "14-4 approach"
      i. Select the diary data from 14 days prior to the CDAI calculation date, and set the data from the 4 days (the day prior to, on the day of and two days after the endoscopy date) to missing;
      ii. Take non-missing diary data from 7 most recent days from step i. If there are multiple entries on the same day for the eDiary component, use the worst result (larger number);
      iii. If there are non-missing diary data from 7 days, the subtotal score is calculated as the sum of the 7-day scores, and multiplying the factor for the corresponding component as listed in the table above;
      iv. If there are non-missing diary data from less than 7 days but greater than 3 days (4, 5, 6 days), the subtotal score is calculated as average of the dairy data and times 7, and multiplying the factor for the corresponding component as listed in the table above;
      v. If there are only less than 4 days of non-missing diary data available, the subtotal score will be set to missing.
   b. "9-2 approach"
      i. Select the diary data from 9 days prior to the CDAI calculation date, and set the data from the 2 days (the day prior to, and on the day of the endoscopy date) to missing;
      ii. Take non-missing diary data from 7 most recent days from step i. If there are multiple entries on the same day for the eDiary component, use the worst result (larger number);

iii. If there are non-missing diary data from 7 days, the subtotal score is calculated as the sum of the 7-day scores, and multiplying the factor for the corresponding component as listed in the table above;

iv. If there are non-missing diary data from less than 7 days but greater than 3 days (4, 5, 6 days) OR from 3 consecutive days, the subtotal score is calculated as average of the dairy data and times 7, and multiplying the factor for the corresponding component as listed in the table above;

v. If there are non-missing diary data from 3 non-consecutive days OR less than 3 days of non-missing diary data available, the subtotal score will be set to missing.

c. Average daily SF/AP will be determined in the same way as described above to derive PROs-related endpoints (for example, clinical remission per PROs);

d. The "14-4 approach" will be used to calculated CDAI, average daily SF/AP at each CDAI date. Sensitivity analyses which are based on "9-2 approach" to calculate CDAI and SF/AP will be performed on primary endpoints (Clinical Remission per CDAI at Week 12, Clinical Remission per PROs at Week 12).

3. Subtotal scores for components 4-6 (Number of 6 listed categories, usage of Lomotil/Imodium/Loperamide/opiates for diarrhea and abdominal mass) will be determined using the data collected on the CDAI calculation date and the corresponding formula in the table above;

4. Subtotal score for component 7 (Hematocrit) is calculated as follows:

a. Identify the Hematocrit (%) value form the same visit as the CDAI calculation date, use the Hematocrit that is closest to the CDAI calculation date;

b. Hematocrit values obtained from previous visits can be used.

c. The subtotal score for Hematocrit is calculated using the formula in the table above. If Hematocrit value is greater or equal to 47% for male subjects (greater or equal to 42% for female subjects), the subtotal score is set to 0.

5. Subtotal score for component 8 (percentage below standard weight) is calculated as follows:
    a. Standard weight is calculated based on height obtained at screening, sex and study protocol Appendix H;
    b. Body weight is based on the measurement obtained on the CDAI calculation data, or previous visits if the measurement is missing.
    c. The subtotal score for percentage below standard weight is calculated using the formula in the table above.
6. CDAI score is calculated as the sum of subtotal scores of 8 components, and the total CDAI score will be rounded to 1 decimal. If any of the subtotal scores are missing, the CDAI score will be missing.

Appendix E. Simple Endoscopic Score – CD (SES-CD)

SES-CD is calculated based the sum of individual segment values for four endoscopic variables (presence and size of ulcers, ulcerated surface, affected surface and presence of narrowing). Each variable in each segment be scored 0 to 3 resulting in SES-CD values ranging from 0 to 56 with higher scores indicating more severe disease.

Variable	Score			
	0	1	2	3
Size of ulcers (cm)	None	Aphthous ulcers (diameter 0.1–0.5)	Large ulcers (diameter 0.5–2)	Very large ulcers (diameter >2)
Ulcerated surface (%)	None	<10	10–30	>30
Affected surface (%)	Unaffected segment	<50	50–75	>75
Presence of narrowings	None	Single, can be passed	Multiple, can be passed	Cannot be passed

*Total SES-CD: sum of the values of the 4 variables for the 5 bowel segments. Values are given to each variable and for every examined bowel segment (for example, rectum, left colon, transverse colon, right colon and ileum).

Medscape Source: Nat Rev Gastroenterol Hepatol ©2009 Nature Publishing Group

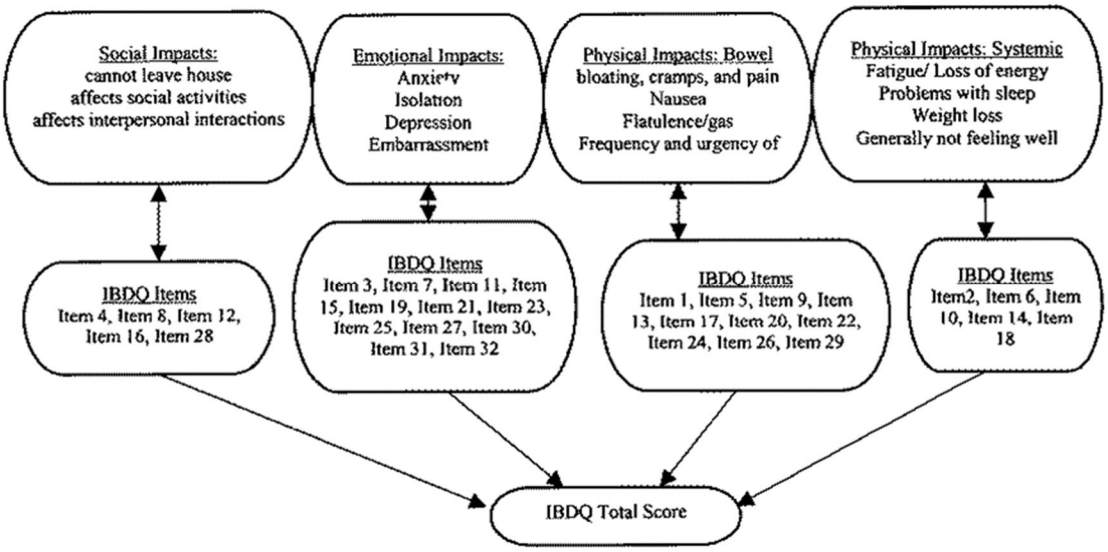
SES-CD Scoring:

	Rectum	Sigmoid and Left Colon	Transverse Colon	Right Colon	Ileum	Total
Size of Ulcers Enter: 0 if none 1 if aphthous ulcers (Ø 0.1 to 0.5 cm) 2 if large ulcers (Ø 0.5 to 2 cm) 3 if very large ulcers (Ø > 2 cm)						
Ulcerated Surface Enter: 0 if none 1 if < 10% 2 if 10% – 30% 3 if > 30%						
Affected Surface Enter: 0 if unaffected segments 1 if < 50% 2 if 50% – 75% 3 if > 75%						
Presence of Narrowing Enter: 0 if none 1 if single, can be passed 2 if multiple, can be passed 3 if cannot be passed						
					TOTAL =	

Appendix F. Patient Reported Outcomes Questionnaires Descriptions

IBDQ – Inflammatory Bowel Disease Questionnaire

The IBDQ is a 32-item (ranges 1 – 7) self-report questionnaire for patients with IBD to evaluate the patient reported outcomes across 4 dimensions: bowel symptoms (loose stools, abdominal pain), systemic symptoms (fatigue, altered sleep pattern), social function (work attendance, need to cancel social events), and emotional function (anger, depression, irritability). The IBDQ total Score ranges from 32 to 224 with a higher score indicating better outcome.



The derivation of the IBDQ total score and the four IBDQ domain scores are as follows:

Domain Scores:

S1. BOWEL SYMPTOM: Q1 + Q5 + Q9 + Q13 + Q17 + Q20 + Q22 + Q24 + Q26 + Q29
[score ranges from 10 - 70],

S2. SOCIAL FUNCTION: Q4 + Q8 + Q12 + Q16 + Q28
[score ranges from 5 - 35],

S3. SYSTEMIC SYMPTOM: Q2 + Q6 + Q10 + Q14 + Q18
[score ranges from 5 - 35],

S4. EMOTIONAL FUNCTION: Q3 + Q7 + Q11 + Q15 + Q19 + Q21 + Q23 + Q25 + Q27 + Q30 + Q31 + Q32
[score ranges from 12 - 84].

Total Score:

IBDQ Total Score = SUM of (bowel symptom domain score, social function domain score, systemic symptom domain score, emotional function domain score).

The following convention applies to IBDQ:

When not more than 20% of items in a domain of IBDQ were missing, it was substituted with the mean values from the items completed in the particular domain; otherwise, they were treated as missing. The 20% threshold in each domain is: bowel symptom domain: 2 items, systemic symptom domain: 1 item, social function domain: 1 item, emotional function domain: 2 items. If any of the 4 domain scores is missing, the total IBDQ score will be set to missing.

WPAI-CD – Work Productivity and Activity Impairment Questionnaire-Crohn's Disease

The WPAI assesses the impact of the condition on work productivity losses and impairment in daily activity. WPAI has six items covering four domains: Absenteeism (work time missed), measured as the number of hours missed from work in the past 7 days due to a condition related problem. Scores are expressed as impairment percentages, adjusting for hours actually worked according to the WPAI scoring algorithm; Presenteeism (impairment at work/reduced on-the-job effectiveness), measured as the impact of the condition on productivity while at work (i.e., reduced amount or kind of work, or not as focused as usual). Responses are recorded on a 0 – 10 Likert scale

(where, 0 = no effect of CD on work and 10 = severe impact of CD while at work); productivity loss (overall work impairment), measured as the sum of hours missed due to condition i.e., absenteeism and number of hours worked with impairment i.e., product of number of hours worked and presenteeism; and activity impairment (i.e., activities other than paid work like work around house, cleaning, shopping, traveling, studying), recorded and scored in same way as presenteeism. Higher numbers indicate greater impairment and less productivity.

CSS – Crohn's Symptoms Severity Questionnaire

The CSS questionnaire is a 14-item assessment of both GI and non-GI CD symptoms using a 7-day recall period. The 14 items assess the frequency or intensity of each individual symptom by utilizing a 5-point Likert-type scale. An assessment of an individual items frequency and intensity is measured as follows:

- Frequency: 1 = Never, 2 = Rarely, 3 = Sometimes; 4 = Often, 5 = Always
- Intensity: 1 = Not at all, 2 = A little bit; 3 = Somewhat; 4 = Quite a bit; 5 = Very much

An overall symptom score is calculated by combining ratings from the 14 individual items (possible scores range from 14 to 70) with higher scores indicating greater symptom frequency or intensity.

SF-36 – Short Form 36

The SF-36 questionnaire is a self-administered multi-domain scale with 36 items. Eight subscales cover a range of functioning: physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health. The scoring yields a physical component score, a mental component summary score, and subscale scores. Higher scores represent better outcomes. The concepts measured by the SF-36 are not specific to any age, disease, or treatment group, allowing comparison of relative burden of different diseases and the benefit of different treatments.

EQ-5D – European Quality of Life 5 Dimensions

The EQ-5D-5L is a standardized non-disease specific instrument for describing and valuing health-related quality of life. The EQ-5D consists of 5 dimensions: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Each dimension has 5 levels: no problem, slight problem, moderate problem, severe problem or unable to do the activity. It also contains a Visual Analogue Scale (VAS). Subjects are asked to indicate the level that describes their current level of function or experience for each dimension. As a measure of health status, it provides a descriptive profile and can be used to generate a single index value for health status, where full health is equal to 1 and death is equal to 0. The VAS records the subject's assessment of his/her own health along a vertical 20 cm line, which has health state scores between 0 and 100.

FACIT-F – Functional Assessment of Chronic Illness Therapy-Fatigue

The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system is a comprehensive compilation of questions that measure health related quality of life in patients with cancer and other chronic diseases (www.facit.org). The FACIT-F scale is a symptom-specific subscale of FACIT and is composed of 13 fatigue-related questions as follows. The response to each question is as follows: 0=Not at all, 1=A little bit, 2=Somewhat, 3=Quite a bit, 4=Very much.

1. I feel fatigued
2. I feel weak all over
3. I feel listless ("washed out")
4. I feel tired
5. I have trouble starting things because I am tired
6. I have trouble finishing things because I am tired
7. I have energy
8. I am able to do my usual activities

9. I need to sleep during the day
10. I am too tired to eat
11. I need help doing my usual activities
12. I am frustrated by being too tired to do the things I want to do
13. I have to limit my social activity because I am tired

In order to have higher values representing lower level of fatigue, all of the items, except for item 7 "I have energy" and item 8 "I am able to do my usual activities," are assigned reversed scores: reversed score = 4 – raw score. The total FACIT-F score will then be calculated as follows:

- FACIT-F score = 13×[Sum of answered-item scores / Number of items answered]

The score ranges from 0 to 52, 52 being the lowest level of fatigue.

When there are missing data, provided that more than 50% of the items (i.e., at least 7 of 13 items) were answered in FACIT-F questionnaire, the total score will not be deemed as missing but be calculated as the mean response for all the non-missing-value items. If less than 50% of the items were answered, the total score will have a missing value.

Appendix G. Extra-Intestinal Manifestations (EIMs)

The following 16 items consist of evaluation of EIM, each of the item will be evaluated based on a binary response (Yes/No). Subjects with any EIM at baseline will be included in the analysis of resolution of EIMs, only if none of the EIMs are presented (all 16 items with outcome of "No") will be considered as "achievement of resolution of EIMs."

1. Peripheral arthropathy	9. Oral aphthous ulcers
2. Axial arthropathy (including sacroiliitis or ankylosing spondylitis)	10. Primary sclerosing cholangitis
3. Episcleritis	11. Autoimmune hepatitis
4. Uveitis	12. Venous thromboembolism
5. Iritis	13. Chronic obstructive pulmonary disease
6. Erythema nodosum	14. Bronchiectasis
7. Pyoderma gangrenosum	15. Nephrolithiasis
8. Sweet's syndrome	16. Anemia

Appendix H. Cronh's Disease Location Per SES-CD

For isolated ileal disease in the analysis using SES-CD: Patients who have SES-CD = 0 in all segments other than ileum are considered to have isolated ileitis.

For ileocolonic disease in the analysis using SES-CD: Patients who have SES-CD different than 0 in the ileum and at least one of the colonic segments.

For colonic disease in the analysis using SES-CD: Patients who have SES-CD = 0 in the ileum and SES-CD different than 0 in at least one of the colonic segments.

Appendix I. Random Seeds

In case of non-convergence, the random seed will be updated by adding 100000 at each attempt until convergence of model happens.

Table I-1. Random Seeds for NRI-C

Variables	Random Seed	
	MCMC Procedure	**PROC MI**
CDAI	14431	24431
SES-CD (for 20 individual variables)	N/A	24432
SF	14441	24441
AP score	14451	24451
EIM	N/A	24452
IBDQ total score	14461	24461
FACIT-F	14471	24471
CGHAS	14491	24491

Appendix J. Geographic Region

Below table lists the countries/regions considered for each geographic region.

Geographic Region	Countries/Regions
Asia	Japan (JPN), China (CHN), Malaysia (MYS), South Korea (KOR), Taiwan (TWN)
Eastern Europe	Bosnia and Herzegovina (BIH), Croatia (HRV), Czech Republic (CZE), Estonia (EST), Hungary (HUN), Lithuania (LTU), Poland (POL), Romania (ROU), Russia (RUS), Serbia (SRB), Slovakia (SVK), Slovenia (SVN), Turkey (TUR)
North America	Canada (CAN), United States (USA), Puerto Rico (PRI)
Other	Australia (AUS), Egypt (EGY), Israel (ISR), South Africa (ZAF)
South/Central America	Argentina (ARG), Brazil (BRA), Chile (CHL), Mexico (MEX)
Western Europe	Austria (AUT), Belgium (BEL), Denmark (DNK), France (FRA), Germany (DEU), Greece (GRC), Ireland (IRL), Italy (ITA), Netherlands (NLD), Norway (NOR), Portugal (PRT), Spain (ESP), Switzerland (CHE), Sweden (SWE), United Kingdom (GBR)