Statistical Analysis Plan for Study M15-722

A Multicenter, Single Arm, Open-label Study to Investigate the Efficacy and Safety of Ravagalimab (ABBV-323) in Subjects with Moderate to Severe Ulcerative Colitis Who Failed Prior Therapy

Date: 10 February 2021

Version 3.0

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 ravagalimab (ABBV-323) Study M15-722 Ulcerative Colitis (UC): Ravagalimab in Subjects with Moderately to Severely Active UC Who Failed Prior Therapy.

Study M15-722 examines the efficacy and safety of ravagalimab in subjects with moderately to severely active UC who failed prior therapy.

The analyses of pharmacokinetic and biomarker analyses will not be covered 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 the protocol. Details are outlined in Section 13.0.

2.0 Study Design and Objectives

2.1 Objectives and Hypotheses

To explore the efficacy, safety, and tolerability of ABBV-323 as treatment in subjects with moderately to severely active UC.

2.2 Study Design Overview

This is a Phase 2a, multicenter, single arm, open-label (OL) study to investigate the efficacy and safety of ABBV-323 in subjects with moderate to severe UC who failed prior therapy. The study contains a 35-day screening period, 2 treatment periods, and an 84-day follow up period from the last dose of study drug. The induction period is an open-label 12-week treatment period to evaluate the efficacy, safety, tolerability, PK and RO of ABBV-323 for inducing endoscopic improvement at Week 8. Subjects who achieve

clinical response per Partial Adapted Mayo score at Week 12 may continue into the maintenance period. The maintenance period is an OL 92-week treatment period to assess efficacy and safety for maintenance of response with ABBV-323. An internal data monitoring committee (DMC) will review safety at regular intervals.

Induction Period

Up to approximately 40 subjects will be enrolled in the induction period to receive the following treatment:

- ravagalimab 600 mg intravenous (IV) Week 0; ravagalimab 300mg subcutaneous (SC) Weeks 2, 4, 6, 8, 10.

Prior to protocol Version 4.0 subjects were to be randomized in a 2:1 ratio to the following treatment groups:

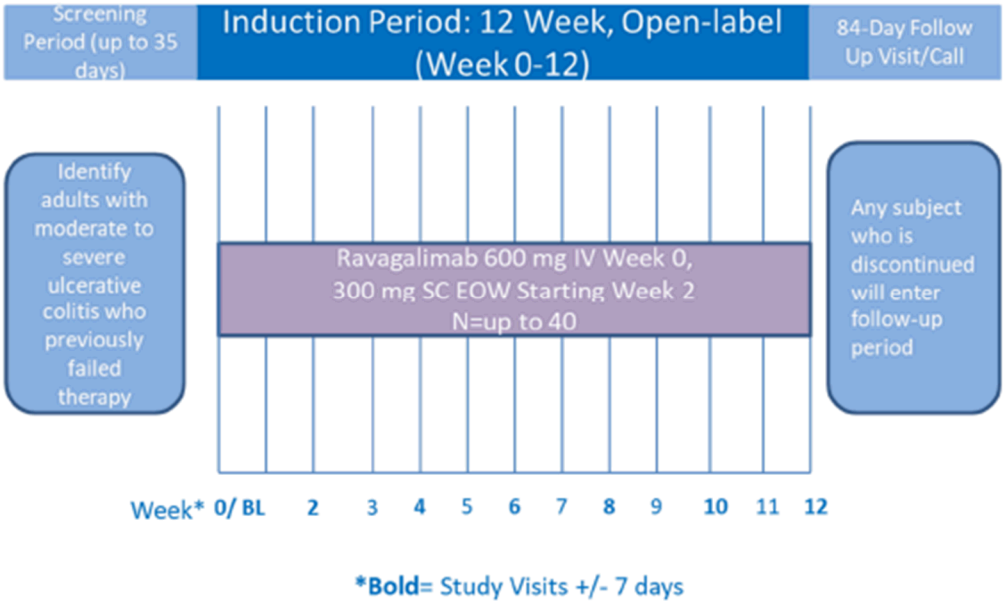
- Group 1: ravagalimab 600mg intravenous (IV) Week 0; ravagalimab 300 mg subcutaneous (SC) Weeks 2, 4, 6, 8, 10
- Group 2: Placebo IV Week 0; Placebo SC Weeks 2, 4, 6, 8, 10

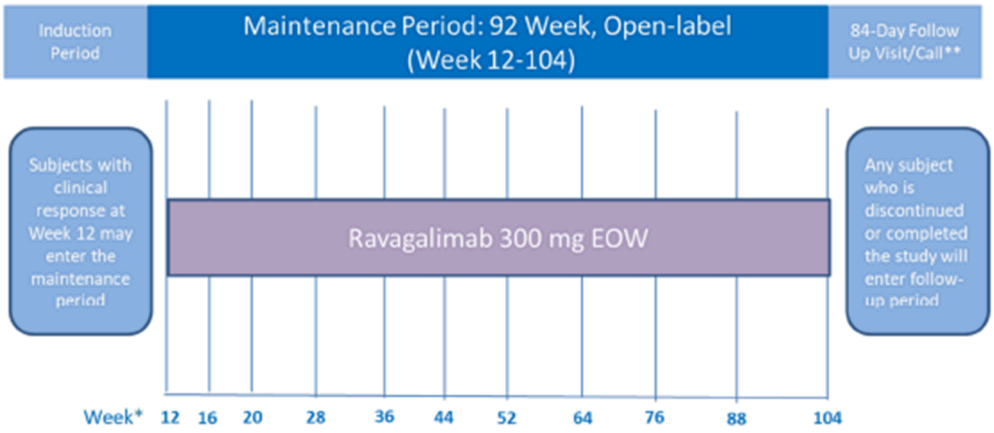
Maintenance Period

Subjects who enter the maintenance period will receive OL ravagalimab 300mg SC every other week (eow) from Week 12 through Week 102.

The schematic of the study is shown in Figure 1.

Figure 1. Induction and Maintenance Period Schematic





*Bold= Study Visits +/- 7 days
**84-day follow up visit/call from last study drug administration
Study Drug Administration Visits ONLY not listed in the schematic above and will occur at:
Week 14, 18, 22, 24, 26, 30, 32, 34, 38, 40, 42, 46, 48, 50, 54, 56, 58, 60, 62, 66, 68, 70, 72, 74, 78, 80, 82, 84, 86, 90, 92, 94, 96, 98, 100, 102

The primary analysis will be performed after all subjects completed the study induction period (i.e., completed Week 12 assessments). The final analysis will be performed after all subjects completed the study maintenance period. There will be no multiplicity control for this proof of concept study.

2.3 Treatment Assignment and Blinding

This is a single arm, open-label (OL) study.

2.4 Sample Size Determination

The planned sample size is up to approximately 40 subjects.

Meta-analysis for historical placebo data from tofacitinib (OCTAVE1/2)[1], upadacitinib (Phase 2b; data on file) and ustekinumab (Stelara Phase 3 UNIFI)[2] studies gives [redacted] endoscopic improvement rate at Week 8. The matched prior distribution for placebo is [redacted] The prior for ravagalimab treatment group will be set to Jeffreys prior Beta (0.5, 0.5) as non-informative. The probability that the posterior probability of rate difference > 0 is larger than 90% is more than 90% if the endoscopic improvement rate difference is 25% between ravagalimab and historical placebo, given a sample size of 40 subjects for ravagalimab. If ravagalimab has the same endoscopic improvement rate of [redacted] as historical placebo, that probability is approximately [redacted]

3.0 Endpoints

3.1 Primary Endpoint

The primary efficacy endpoint is endoscopic improvement (Mayo endoscopic subscore of 0 or 1) at Week 8. Note that evidence of friability during endoscopy in subjects with a Mayo endoscopic subscore of 0 or 1 will confer an endoscopic subscore of 2.

3.2 Secondary Endpoints

- Clinical remission per Adapted Mayo score at Week 8

- Clinical response per Adapted Mayo score at Week 8
- Clinical response per Partial Adapted Mayo score over time (Week 0 to 12)
- Clinical remission per Full Mayo score at Week 8 in subjects with a Full Mayo score of 6 to 12 at Baseline
- Endoscopic remission at Week 8

3.3 Additional Efficacy Endpoints

Additional efficacy endpoints are:

- Change from Baseline in modified Baron score over time (at visits with endoscopy)
- Achievement with stool frequency subscore (SFS) = 0, rectal bleeding subscore (RBS) = 0, and Mayo endoscopic subscore = 0 over time (at visits with endoscopy)
- Time to clinical response per Partial Adapted Mayo score
- Endoscopic improvement over time (at visits with endoscopy)
- Clinical remission per Adapted Mayo score over time (at visits with endoscopy)
- Clinical response per Adapted Mayo score at over time (at visits with endoscopy)
- Clinical remission per Partial Adapted Mayo score over time
- Clinical response per Partial Adapted Mayo score over time
- Clinical remission per Adapted Mayo score over time among subjects who were taking corticosteroids at Baseline and subsequently discontinued corticosteroids for ≥ 90 days prior to the respective visit
- Discontinuation of corticosteroid use for subjects taking corticosteroids at Baseline at all visits over time
- Endoscopic remission over time (at visits with endoscopy)
- Clinical remission per Full Mayo score over time in subjects with a Full Mayo score of 6 to 12 at Baseline (at visits with endoscopy)
- Clinical response per Full Mayo score over time in subjects with a Full Mayo score of 6 to 12 at Baseline (at visits with endoscopy)

- Achievement of SFS ≤ 1 over time
- Achievement of RBS = 0 over time
- Change from Baseline in RBS over time
- Change from Baseline in SFS over time
- Change from Baseline in fecal calprotectin (FCP) over time
- Change from Baseline in high sensitivity C-reactive protein (hs-CRP) over time
- Change from Baseline in Inflammatory Bowel Disease Questionnaire (IBDQ) over time
- UC-related Hospitalizations (yes/no) over time
- UC-related surgeries (yes/no) over time
- No bowel urgency over time
- Change from Baseline in weekly abdominal pain score over time
- Achievement of ≥ 30% reduction from Baseline in weekly abdominal pain score over time
- Achievement of Weekly abdominal pain score = 0 over time
- Change from Baseline in number of extraintestinal manifestations (EIM) over time (as captured in the EIM electronic case report form [eCRF])
- Histologic remission over time (at visits with endoscopy)
- Mucosal healing (endoscopic and histologic remission) over time (at visits with endoscopy)
- Change from Baseline in Ulcerative Colitis Endoscopic Index of Severity (UCEIS) over time (at visits with endoscopy)
- Change from Baseline in UC-100 score over time (at visits with endoscopy)

3.4 Safety Endpoint(s)

Safety evaluations include AE monitoring, physical examinations, vital sign measurements, and clinical laboratory testing (hematology, chemistry, and urinalysis) as measures of safety and tolerability. An internal DMC is monitoring safety at regular intervals.

4.0 Analysis Populations

The following population sets will be used for the analyses.

The Full Analysis Set (FAS) includes all enrolled subjects who received at least 1 dose of ravagalimab (including subjects randomized to ravagalimab prior to protocol Version 4.0). The FAS will be used for all efficacy and baseline analyses.

Subjects enrolled prior to protocol version 4.0 will be unblinded after they finish Week 8 assessments. Subjects who were randomized to and received ravagalimab will be included in the FAS.

The Safety Analysis Set consists of all subjects who received at least 1 dose of study drug. Subjects randomized to the study prior to Version 4.0 will be analyzed along with subjects enrolled after protocol Version 4.0 for safety analysis.

5.0 Subject Disposition

The total number of subjects who were enrolled (randomized), and treated will be summarized.

A summary of subject accountability will be provided where the number of subjects in each of the following categories will be summarized:

- Subjects enrolled (randomized) in the study;
- Subjects included in key analysis populations (FAS, Safety Analysis Set);
- Subjects who took at least one dose of study drug;
- Subjects who completed protocol-specified treatment;
- Subjects who prematurely discontinued study drug (all reasons and primary reason);

For end of study participation, the number and percentage of subjects who completed the protocol defined follow-up period (or did not with associated reasons) will be summarized.

6.0 Study Drug Duration and Compliance

For the FAS, duration of treatment will be summarized. Duration of treatment is defined for each subject as last dose date minus first dose date + 14 days. Duration of treatment will be summarized 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 (1 to 14, 15 to 28, 29 to 42, 43 to 56, 57 to 84, and > 84 days) will be summarized.

Treatment compliance will be summarized for the entire treatment period for the FAS. Treatment compliance is defined as the number of injection actually taken divided by the number of injection 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 FAS. 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 and standard deviation, median, minimum and maximum).

7.1 Demographics and Baseline Characteristics

Continuous demographic variables include

- Age (years),
- weight (kg),
- body weight – female (kg)

- body weight – male (kg)
- height (cm),
- height – female (cm)
- height – male (cm)
- body mass index (bmi) (kg/ m^2).

Categorical demographic variables include

- sex (male, female),
- age (< 40, 40 - < 65, ≥ 65 years),
- ethnicity (Hispanic or Latino, not Hispanic or Latino)
- race (white, non-white),
- region (Europe, North America, or rest of world),
- tobacco user (current, former, never, unknown),
- alcohol user (current, former, never, unknown).

Continuous disease characteristics at baseline include

- full mayo score and its components (stool frequency, rectal bleeding, PGA and endoscopy subscores),
- partial mayo score,
- adapted mayo score,
- hs-CRP (mg/L),
- fecal calprotectin (mg/kg),
- albumin (g/L),
- UC disease duration in years,
- modified Baron score,
- UCEIS,
- Geboes score,
- number of extraintestinal manifestations.

Categorical disease characteristics include

- baseline steroid use (yes, no),
- baseline immunosuppressant use (yes, no),
- baseline aminosalicylates use (yes, no),
- prior exposure to biologic/JAK therapy for UC (1, >1)
- prior inadequate response/intolerance to biologic/JAK therapy for UC (1, > 1)
- prior exposure to biologic therapy for UC (1, >1),
- prior exposure to anti-TNF use for UC (yes, no),
- baseline hs-CRP (<=5mg/L, > 5mg/L),
- mayo endoscopic subscore (2, 3),
- adapted mayo score (<= 7, > 7),
- full mayo score (<= 9, > 9),
- disease duration (≤ 3 years, > 3 years),
- disease extent (left-sided, extensive/pancolitis, limited to rectum)

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 and preferred term) will be summarized. The system organ class (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 preferred term).

7.3 Prior and Concomitant Medications

Prior and concomitant medications will be summarized by generic name. A prior medication is defined as any medication taken prior to the date of the first dose of study drug. A concomitant medication is defined as any medication that started prior to the date

of the first dose of study drug and continued to be taken after the first dose of study drug or any medication that started on or after the date of the first dose of study drug, but not after the date of the last dose of study drug plus 84 days. 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 Efficacy Analyses

8.1 General Considerations

UC-Related corticosteroids intercurrent events are defined as follows:

- subjects not on UC-related corticosteroids (systemic or locally acting corticosteroids for UC) at Baseline who initiated UC-related corticosteroids during the Induction Period;
- subjects on UC-related systemic corticosteroids at Baseline who had dosages increased to greater than the prednisone equivalent dose of corticosteroid at Baseline, or initiation of any rectal corticosteroids during the Induction Period regardless of rectal corticosteroid dose.

The time point of the UC-related corticosteroids intercurrent event is defined as the date when one of the scenarios above occurs for a subject. As such, subjects will be considered as "non-responder" for binary endpoints at or after the occurrence of the UC-related corticosteroids intercurrent event through the end of the Induction Period. For continuous endpoints, all measurements at or after the occurrence of the UC-related corticosteroids intercurrent event through the end of the Induction Period will not be used in the analysis.

8.2 Handling of Missing Data

Missing data imputation will be done only for efficacy variables (categorical and continuous variables). For the analysis other than efficacy, no imputation will be done for missing values.

Censoring

For the time-to-event endpoint, i.e., time to clinical response per Partial Adapted Mayo score, missing data will be considered as censored at the last available time for the information.

Non-Responder Imputation incorporating Multiple Imputation to handle missing data due to COVID-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. In addition, at or after the UC-related corticosteroids intercurrent event (see Section 8.1), subjects will be counted as non-responders. The only exception is that missing data due to COVID-19 infection or logistical restrictions related to the COVID-19 pandemic will be handled by Multiple Imputation (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.

Mixed-Effect Model Repeated Measurements (MMRM)

MMRM analysis will be conducted using mixed-effect model including observed measurements at all visits. For the MMRM analysis, data collected at or after UC-Related corticosteroids intercurrent event will be excluded. The mixed model includes the categorical fixed effects of visit, baseline corticosteroid use (yes vs no), baseline Mayo endoscopic sub-score (2 vs 3), the continuous fixed covariate of baseline measurement and treatment and treatment-by-visit interaction, if applicable. An unstructured variance covariance matrix will be used. If there is a convergence issue, a compound symmetry structure and a variance components structure will be considered in sequence. The parameter estimations are based on the assumption of data being missing at random and using the method of restricted maximum likelihood (REML).

As Observed (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. All observed data will be used in the analysis.

8.3 Primary Efficacy Endpoint and Analyses

8.3.1 Primary Efficacy Endpoint

The primary efficacy endpoint is endoscopic improvement (Mayo endoscopic subscore of 0 or 1) at Week 8. Note that evidence of friability during endoscopy in subjects with a Mayo endoscopic subscore of 0 or 1 will confer an endoscopic subscore of 2.

8.3.2 Handling of Missing Data for the Primary Efficacy Endpoint

For the primary analysis, NRI-C will be used for missing data imputation.

8.3.3 Primary Analysis of the Primary Efficacy Endpoint

Analysis of the primary efficacy endpoint will be conducted on the FAS. Comparison of the primary efficacy endpoint will be performed between the ravagalimab group and the historical placebo group using Bayesian approach with placebo distribution [redacted] from historical data meta-analysis and prior for ravagalimab Beta (0.5, 0.5), the Jeffrey's prior for binomial distribution. The posterior probability of rate difference > 0 given data, i.e., $P(\theta_{ABBV-323} - \theta_{Placebo} > 0|data)$ will be calculated. In addition, the mean difference and 95% credible interval for the mean difference will be provided.

8.3.4 Sensitivity and Supplementary Analysis of the Primary Efficacy Endpoint

The same Bayesian approach will be also conducted based on the FAS using AO data as a sensitivity analysis.

As a supplementary analysis for the primary efficacy endpoint, the following approach will be used. For more details, see Appendix D.

1. Bayesian approach using [redacted] prior for placebo group and Beta(0.5, 0.5) prior for treatment group. Note [redacted]

In addition, point estimate and 95% CI using normal approximation will be provided for the response rate for the ravagalimab group using NRI-C and AO data.

8.3.5 Additional Analyses of the Primary Efficacy Endpoint

In addition, a set of synthetic placebo subjects will be created from Bio-IR placebo subjects of Upadacitinib UC Study M14-234 (Substudy 1, ph2b and Substudy 2, ph3) based on the similarity of inclusion/exclusion criteria and definitions of efficacy endpoints of the studies.

A logistic regression with potential firth adjustment model with being treated as the outcome will be performed using treatment subjects in Study M15-722 and placebo subjects from the two historical Study M14-234 Substudy 1 and Study M14-234 Substudy 2. Selected common baseline demographics and disease characteristics variables among all the three studies will be included as predictive covariates in the model. See Appendix E for the full list. Propensity score for each subject will be the estimated probability of being treated conditional on the covariates.

A 1:1 matched placebo group will be generated from the above set of synthetic placebo subjects using "greedy matching" i.e., nearest neighbor matching without replacement within a pre-specified caliper distance 0.2 of the standard deviation of the logit of the propensity scores, to match the treatment subjects in Study M15-722. Other matching ratios or calipers may also be considered to preserve matched number of patients and considering the similarity with historical data. If multiple placebo subjects have propensity scores that are equally close to that of the treated subject, one of these placebo subjects will be selected at random. Subjects who are not selected in the matching process will be excluded from data analysis. Pairwise comparison between the matched treatment and placebo groups will be performed for primary endpoints using the Cochran-

Mantel-Haenszel (CMH) test stratified by baseline corticosteroid use (yes, no) and Mayo endoscopic subscore (2 vs 3). NRI-C will be used to handle missing data.

As a sensitivity analysis, 1:1 matched placebo group will be generated from the above set of synthetic placebo subjects using optimal matching to match the treated subjects in Study M15-722.

Additional analysis based on stratification on the propensity score may also be performed. All the subjects including synthetic placebo subjects in the logistic regression model will be ranked according to their estimated propensity score. Subjects are then stratified into five equal-size groups based on the treatment arm using the quintiles of the estimated propensity score. Pairwise comparison between the treatment and synthetic placebo groups will then be performed using the Cochran- Mantel-Haenszel (CMH) test stratified by the quintiles of the estimated propensity score, baseline corticosteroid use (yes, no) and Mayo endoscopic subscore (2 vs 3). NRI-C will be used to handle missing data.

8.4 Secondary Efficacy Endpoint and Analyses

For binary endpoints, point estimate and 95% CI using normal approximation will be provided for the response rate for the ravagalimab using NRI-C and AO data. The same Bayesian approach specified in Section 8.3.3 will be also conducted based on the FAS using NRI-C and AO data, if applicable.

The same additional analyses specified in Section 8.3.5 will be also conducted for secondary endpoints. Point estimate, 95% CI and p-value will be provided for the treatment comparison between ravagalimab and the placebo group based on the CMH test adjusting for baseline corticosteroid use (yes vs no) and Mayo endoscopic sub-score (2 vs 3).

8.5 Additional Efficacy Analyses

Additional endpoints defined in Section 3.3 will be summarized at scheduled time points.

Additional Efficacy Analyses for Induction Period

For additional binary endpoints, point estimate and 95% CI using normal approximation will be provided for the response rate of the ravagalimab group using NRI-C and AO data.

For additional continuous endpoints, the LS-mean and 95% CI will be reported for the ravagalimab group using MMRM model and ANCOVA model baed on AO data with baseline corticosteroid use (yes vs no) and baseline Mayo endoscopic sub-score (2 vs 3) as the fixed factors and the corresponding baseline value as the covariates.

For selected additional binary endpoints, point estimate and 95% CI using normal approximation will be provided for the response rate of ravagalimab and matched placebo group. Point estimate, 95% CI and p-value will be provided for the treatment comparison between ravagalimab and the matched placebo group based on the CMH test adjusting for baseline corticosteroid use (yes vs no) and Mayo endoscopic sub-score (2 vs 3).

For selected additional continuous endpoints, the LS mean and 95% CI will be reported for ravagalimab and matched placebo group. The LS mean treatment difference and associated 95% CI and p-values between ravagalimab and matched placebo group will be provided using MMRM model. For the MMRM analysis, data collected at or after UC-Related corticosteroids intercurrent event will be excluded. Supplementary analysis will be conducted based on ANCOVA model using AO data, with treatment, baseline corticosteroid use (yes vs no) and baseline Mayo endoscopic sub-score (2 vs 3) as the fixed factors and the corresponding baseline value as the covariates.

Additional Efficacy Analyses for Maintenance Period

Descriptive statistics will be provided for the ravagalimab group. These include the number of observations, mean, standard deviation, 95% CI, median, minimum, and maximum for continuous endpoints using AO data; and frequencies and percentages with 95% CI using normal approximation for binary endpoints using AO data.

9.0 Safety Analyses

9.1 General Considerations

Safety data will be summarized for the Safety Analysis Set. Safety summaries will be presented.

There are two sets of planned safety analysis: safety analysis for the Induction Period, and long-term safety analysis.

Safety Analysis for the Induction Period

Safety analysis for the Induction Period will include reporting of adverse events (AEs), laboratory, and vital signs measurements. Frequency 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. All continuous laboratory parameters and vital signs variables at each visit will also be summarized. Frequency tables of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided.

Long-Term Safety Analysis

Long-term safety analysis will be performed on all cumulative safety data in Induction Period and Maintenance Period up to data cutoff date for any interim lock as deemed appropriate and final lock.

Long-term safety analyses include reporting of AE rate adjusted by cumulative exposure, descriptive summary in laboratory parameters and vital sign variables, and rate of potentially clinically significant laboratory and vital signs values.

Exposure-adjusted event rate (EAER) will be provided for long term safety analysis. For EAER calculation, the numerator will be the total number of AEs reported for the event (i.e., a subject can contribute more than one event to the numerator) and the denominator

will be the total exposure time among subjects under the treatment. The number of AEs reported (numerator), the total number of years of study drug exposure (denominator calculated as total number of days exposed to study drug (i.e., last dose date – first dose date + 84 days) for all treated subjects divided by 365.25), and the exposure-adjusted AE event rate per 100 patient-years calculated as ([numerator (number of AEs)/denominator]) • 100 will be presented.

All continuous laboratory parameters and vital signs variables at each visit will be summarized.

Frequency tables and listings of subjects meeting criteria for potentially clinically significant vital sign values and for potentially clinically significant laboratory values will be provided. Missing safety data will not be imputed.

9.2 Adverse Events

Adverse events (AEs) will be summarized and presented using primary MedDRA System Organ Classes (SOCs) and preferred terms (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 adverse events 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.

9.2.1 Treatment-Emergent Adverse Events

Treatment-emergent AEs are defined as any AE with the onset that is on or after the first dose of study drug to the last dose date prior to cut-off date + 84 days or the cut-off date for database lock, whichever occurs first. Events where the onset date is the same as the study drug start date are assumed to be treatment emergent. All treatment-emergent AEs will be summarized overall, as well as by primary MedDRA SOC and Preferred

Term. The SOCs will be presented in alphabetical order, and the PTs will be presented in alphabetical order within each SOC.

The number and percentage of subjects experiencing treatment-emergent AEs will be summarized.

9.2.2 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 treatment-emergent AE
- Any COVID-19 related treatment-emergent AE
- Any treatment-emergent AE related to study drug according to the investigator
- Any severe treatment-emergent AE
- Any serious treatment-emergent AE
- Any treatment-emergent AE leading to discontinuation of study drug
- Any treatment-emergent AE leading to death
- AEs of Special Interest
- All deaths
    - COVID-19 related deaths
    - Deaths occurring ≤ 84 days after last dose of study drug
    - Deaths occurring > 84 days after last dose of study drug.

COVID-19 cases will be identified using relevant terms for COVID-19 search available in MedDRA version 23.1 or higher. All COVID-19 related AEs and COVID-19 related deaths will be summarized in the TEAE Overview table.

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

Treatment-emergent adverse events 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 adverse events 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.

In addition, treatment-emergent adverse events will be summarized by PT and sorted by decreasing frequency for the active group.

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

Exposure-adjusted AEs per 100 patient-years will be provided, where AEs per 100 patient-years of exposure are defined as the number of AEs divided by the total exposure in 100 patient-years.

9.2.5 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.

9.2.6 Adverse Events of Special Interest

Adverse events of special interest (AESI) will be summarized by SOC and PT and will be based on standardized or company MedDRA queries (SMQs or CMQs) or based on adjudication results. Adverse events of special interest are categorized as follows:

AESI Grouping	Categories (AESI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Serious Infections and Opportunistic Infections (excluding Tuberculosis and Herpes Zoster)	Serious Infections	Serious PTs of the SOC Infections and Infestations	PTs	Y
	Opportunistic Infections excluding Tuberculosis and Herpes Zoster	Opportunistic infections CMQ excluding Tuberculosis and Herpes Zoster (code 800000189)	PTs	Y
Active Tuberculosis	Active Tuberculosis	Active Tuberculosis CMQ (code 80000188)	PTs	Y
Malignancies	All Possible Malignancies	Narrow Malignancies (SMQ 20000090)	All Possible Malignancies	N
	Malignant Tumours	Narrow Malignant tumours (SMQ 20000194)	Malignant Tumours	Y
	Non-Melanoma Skin Cancer (NMSC)	Broad Skin malignant tumours (SMQ 20000204) excluding terms identified by the Melanoma CMQ (code 80000119)	Non-Melanoma Skin Cancer (NMSC)	Y
	Malignancies excluding NMSC	'Malignancies excluding NMSC' is identified by the 'Malignant Tumours' search excluding terms identified by the 'Non-melanoma skin cancer' (NMSC) search.	PTs	Y

AESI Grouping	Categories (AESI)	Search Criteria	Terms to Display	Include in AE Overview (Y/N)
Hypersensitivity Reaction	Hypersensitivity	Narrow Hypersensitivity (SMQ 20000214)	PTs	Y
Anaphylactic Reaction	Anaphylactic Reaction	Narrow Anaphylactic reaction (SMQ 20000021)	PTs	Y
Hepatic Events	Hepatic Events	Broad Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions (SMQ 20000013) Broad Hepatitis, non-infectious (SMQ 20000010) Broad Cholestasis and jaundice of hepatic origin (SMQ 20000009) Broad Liver related investigations, signs and symptoms (SMQ 20000008) Narrow Liver-related coagulation and bleeding disturbances (SMQ 20000015)	PTs	Y
Hematologic disorder	Hematologic Events	Broad "Haematopoietic Cytopenias" (SMQ 20000027)	PTs	Y

9.3 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 operations manual (e.g., hematology and clinical chemistry) will be summarized.

Each laboratory 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 selected laboratory variables, with the number of observations, baseline mean, and visit mean.

The baseline and post-baseline laboratory observations will be categorized as Grade 0, Grade 1, Grade 2, Grade 3, and Grade 4 and shifts from baseline grade to worst on-therapy grade will be summarized. A similar shift table will be provided to summarize shifts from baseline to the final post-baseline value. Toxicity grading scale is based on CTCAE version 4.03 (Appendix B).

Laboratory abnormalities will be evaluated based on Potentially Clinically Significant (PCS) criteria (Appendix B). For each laboratory PCS criterion, the number and percentage of subjects who have a laboratory value meeting the criteria will be summarized. Listings will be provided to summarize subject-level laboratory data for subjects meeting PCS criteria.

In addition, ALT/SGPT, AST/SGOT, alkaline phosphatase, and total bilirubin will be categorized as follows:

- ALT $\geq 3 \times$ ULN
- ALT $\geq 5 \times$ ULN
- ALT $\geq 10 \times$ ULN
- ALT $\geq 20 \times$ ULN

- AST ≥ 3 × ULN
- AST ≥ 5 × ULN
- AST ≥ 10 × ULN
- AST ≥ 20 × ULN
- TBL ≥ 2 × ULN
- Alkaline phosphatase ≥ 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 limit of normal.

A listing of potentially clinically significant liver function laboratory values will be provided. The listing will include all subjects who met any of the following 4 criteria:

- ALT ≥ 3 × ULN, or
- AST ≥ 3 × ULN, or
- Alkaline phosphatase ≥ 1.5 × ULN, or
- Total bilirubin ≥ 1.5 × ULN.

A listing of possible Hy's Law cases, defined as those who meet all of the following conditions at any post-baseline visit will be provided:

- ALT of > 3 × ULN or AST of > 3 × ULN,
- Total bilirubin ≥ 2 × ULN.

9.4 Analysis of Vital Signs

Vital sign measurements of systolic and diastolic blood pressure, pulse 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.

Vital sign variables will be evaluated based on potentially clinically significant (PCS) criteria (Appendix B). For each vital sign PCS 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 PCS criteria.

10.0 Other Analyses

Not applicable.

11.0 Interim Analyses

After 30 ongoing subjects complete Week 8 assessments, an interim analysis will be performed in order to explore if proof of concept has already been demonstrated. This interim analysis will inform the planning of future Phase 2b and 3 studies. Enrollment into the induction period will continue until approximately 40 subjects have been enrolled. For this proof of concept study, no alpha adjustment will be conducted for this interim analysis.

11.1 Data Monitoring Committee

An internal data monitoring committee (DMC) will review unblinded safety data from the study. The primary responsibility of the DMC will be to protect the safety of the subjects participating in this study.

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

12.0 Overall Type-I Error Control

No multiplicity adjustment will be performed for this single-arm open-label proof of concept study.

13.0 Version History

Table 1. SAP Version History Summary

Version	Date	Summary
1.0	11 June 2018	Original version
2.0	15 Aug 2018	Section 6.0 Secondary Endpoints: Highlighted key secondary endpoints and moved the rest to exploratory endpoints. Rationale: to align with protocol amendment per regulatory feedback. Section 7.0 Statistical Analyses for Efficacy: Updated sample size estimation language. Rationale: To align with protocol amendment per regulatory feedback. Section 11.0 Efficacy Analyses: Added priors for sensitivity analysis. Rationale: To align with regulatory feedback.
3.0	10 Feb 2021	Section 7.0 Demographics, baseline characteristics added Section 8.0 Analysis method updated per Protocol updates Appendix C: PCS criteria updated Section 8.3.3, Section 8.3.4 and Appendix D: updated historical rate [redacted] as primary analysis due to new UPA UC phase 3 data available after protocol finalization and included historical rate [redacted] which was specified in protocol as sensitivity analysis.

14.0 References

1. Sandborn WJ, Su C, Sands BE, et al. Tofacitinib as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2017;376(18):1723-36.
2. Sands BE, Sandborn WJ, Panaccione R, et al. Ustekinumab as Induction and Maintenance Therapy for Ulcerative Colitis. N Engl J Med. 2019;381(13):1201-14.

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. Common Toxicity Criteria (CTC) Grade for Laboratory Data

Test	Grade 1	Grade 2	Grade 3	Grade 4
		Chemistry Variables		
SGPT/ALT increased	> ULN – 3.0 × ULN	> 3.0 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
SGOT/AST increased	> ULN – 3.0 × ULN	> 3.0 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
GGT increased	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
ALP increased	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 20.0 × ULN	> 20.0 × ULN
TBL increased	> ULN – 1.5 × ULN	> 1.5 – 3.0 × ULN	> 3.0 – 10.0 × ULN	> 10.0 × ULN
Creatinine increased	> ULN – 1.5 × ULN	> 1.5 – 3.0 × ULN	> 3.0 – 6.0 × ULN	> 6.0 × ULN
CPK increased	> ULN – 2.5 × ULN	> 2.5 – 5.0 × ULN	> 5.0 – 10.0 × ULN	> 10.0 × ULN
Total Cholesterol increased	> ULN - 7.75	> 7.75 - 10.34	> 10.34 - 12.92	> 12.92
Albumin decreased	< LLN – 30 g/L	< 30 – 20 g/L	< 20 g/L	-
Triglycerides increased	1.71 - 3.42 mmol/L	> 3.42 - 5.7 mmol/L	> 5.7 - 11.4 mmol/L	> 11.4 mmol/L
Glucose	< LLN - 3.0 mmol/L or > ULN - 8.9 mmol/L	< 3.0 - 2.2 mmol/L or > 8.9 - 13.9 mmol/L	< 2.2 - 1.7 mmol/L or > 13.9 - 27.8 mmol/L	< 1.7 mmol/L or > 27.8 mmol/L
Sodium	< LLN – 130 mmol/L or > ULN – 150 mmol/L	> 150 – 155 mmol/L	< 130 – 120 mmol/L or > 155 – 160 mmol/L	< 120 mmol/L or > 160 mmol/L
Potassium	< LLN - 3.0 mmol/L or > ULN - 5.5 mmol/L	< LLN - 3.0 mmol/L or > 5.5 - 6.0 mmol/L	< 3.0 - 2.5 mmol/L or > 6.0 - 7.0 mmol/L	< 2.5 mmol/L or > 7.0 mmol/L
Calcium	< LLN - 2.0 mmol/L or > ULN - 2.9 mmol/L	< 2.0 - 1.75 mmol/L or > 2.9 - 3.1 mmol/L	< 1.75 - 1.5 mmol/L or > 3.1 - 3.4 mmol/L	< 1.5 mmol/L or > 3.4 mmol/L

Test	Grade 1	Grade 2	Grade 3	Grade 4
Hematology Variables				
Hemoglobin decreased	< LLN – 100.0 g/L	< 100.0 – 80.0 g/L	< 80.0 g/L	-
Neutrophil count decreased	< LLN – 1.5 × 10^9/L	< 1.5 – 1.0 × 10^9/L	< 1.0 – 0.5 × 10^9/L	< 0.5 × 10^9/L
WBC decreased	< LLN – 3.0 × 10^9/L	< 3.0 – 2.0 × 10^9/L	< 2.0 – 1.0 × 10^9/L	< 1.0 × 10^9/L
Lymphocyte count decreased	< LLN – 0.8 × 10^9/L	< 0.8 – 0.5 × 10^9/L	< 0.5 – 0.2 × 10^9/L	< 0.2 × 10^9/L
Lymphocyte count increased	-	> 4.0 – 20.0 × 10^9/L	20.0 × 10^9/L	-
Platelets count decreased	< LLN - 75.0 × 10^9/L	< 75.0 - 50.0 × 10^9/L	< 50.0 - 25.0 × 10^9/L	< 25.0 × 10^9/L

Appendix C. Potentially Clinically Significant Criteria for Safety Endpoints

The criteria for Potentially Clinically Significant (PCS) laboratory findings are described in Table 2 and Table 3, and the PCS criteria for vital sign findings are described in Table 4.

Table 2. Criteria for Potentially Clinically Significant Hematology Values

Hematology Variables	Units	Definition of Potentially Clinically Significant NCI CTCAE Grade 3 or Greater	
		Very Low	Very High
Hemoglobin	g/dL	< 8.0	
Neutrophil	10^9/L	< 1.0	
Lymphocytes	10^9/L	< 0.5	> 20.0
WBC count	10^9/L	< 2.0	
Platelets count	10^9/L	< 50.0	

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically significant finding.

Table 3. Criteria for Potentially Clinically Significant Chemistry Values

		Definition of Potentially Clinically Significant NCI CTCAE Grade 3 or Greater	
Chemistry Variables	**Units**	**Very Low**	**Very High**
TBL	mcmol/L		> 3.0 × ULN
ALP	U/L		> 5.0 × ULN
SGOT/AST	U/L		> 5.0 × ULN
SGPT/ALT	U/L		> 5.0 × ULN
Albumin	g/L	< 20	
Glucose	mmol/L	< 2.2	> 13.9
Triglycerides	mmol/L		> 5.7
Creatinine	mcmol/L		> 3.0 × ULN
Sodium	mmol/L	< 130	> 155
Potassium	mmol/L	< 3.0	> 6.0
Calcium	mmol/L	< 1.75	> 3.1
Total Cholesterol	mmol/L		> 10.34
GGT			> 5.0 × ULN

Note: A post-baseline value must be more extreme than the baseline value to be considered a potentially clinically significant finding.

Table 4. Criteria for Potentially Clinically Significant Vital Sign Values

Vital Sign	**Category**	**Criteria for Potential Clinically Significant Vital Signs**
Systolic blood pressure (mmHg)	Low	Value ≤ 90 mmHg and decrease ≥ 20 mmHg from Baseline
	High	Value ≥ 160 mmHg and increase ≥ 20 mmHg from Baseline
Diastolic blood pressure (mmHg)	Low	Value ≤ 50 mmHg and decrease ≥ 10 mmHg from Baseline
	High	Value ≥ 100 mmHg and increase ≥ 10 mmHg from Baseline

Appendix D. Details for Historical Data Borrowing

D1. Historical Placebo Rate

The historical placebo rate [redacted] (Figure 2) is based on a robust meta-analysis of four historical trials evaluating centrally-assessed endoscopic improvement rate at Week 8 in a total of 325 placebo subjects with moderate to severe UC who failed prior anti-TNF or biologic therapy. These studies have similar population, inclusion and exclusion criteria as shown in Table 5.

Table 5. Major Inclusion/Exclusion Criteria

Criteria	Historical Studies				Current Study
	OCTAVE-1	**OCTAVE-2**	**Stelara Ph3**	**Upadacitinib M14-234**	**M15-722**
Target Population	Sub population: Anti-TNF Failure	Sub population: Anti-TNF Failure	Sub population: Anti-TNF Failure	Subpopulation: Inadequate Response to Biologic Therapy	Prior Therapy (bio or tofa) Failure
Disease Activity	Baseline Mayo Score 6-12, Endoscopy subscore >= 2	Baseline Mayo Score 6-12, Endoscopy subscore >= 2	Baseline Mayo Score 6-12, Endoscopy subscore >= 2	Baseline Adapted Mayo Score 5-9, Endoscopy subscore >= 2	Baseline Adapted Mayo Score 5-9, Endoscopy subscore >= 2
Age	>= 18 years	>= 18 years	>= 18 years	18-75 years	18-75 years
Prior Medication	Treatment experienced	Treatment experienced	Previously not treated with interleukin-12 or interleukin-23 antagonists	Previously not treated with JAK inhibitor (e.g., tofacitinib, baricitinib, filgotinib)	Previously treated with biologic therapy or tofa
Prohibited Medication	TNF antagonists	TNF antagonists	TNF antagonists	All biologic therapy or other JAK inhibitor	All biologic therapy or small molecule medication with potential impact on UC
Concomitant medication	Oral Aminosalicylates, oral glucocorticoids	Oral Aminosalicylates, oral glucocorticoids	Oral Corticosteroids, Aminosalicylates, immunomodulators	Oral Corticosteroids, Antibiotics, Aminosalicylates, and/or MTX	Oral Corticosteroids, Antibiotics, Aminosalicylates, or immunomodulators

In all the four trials, evidence of friability during endoscopy in subjects with a Mayo endoscopic subscore of 0 or 1 are conferred an endoscopic subscore of 2.

- Tofacitinib Phase 3
  - OCTAVE – Induction 1 (4/65)
  - OCTAVE – Induction 2 (4/65)
- Stelara Phase 3 UNIFI – Induction (11/161).

- [redacted]

Figure 2. Meta-analysis of Historical Placebo Rate

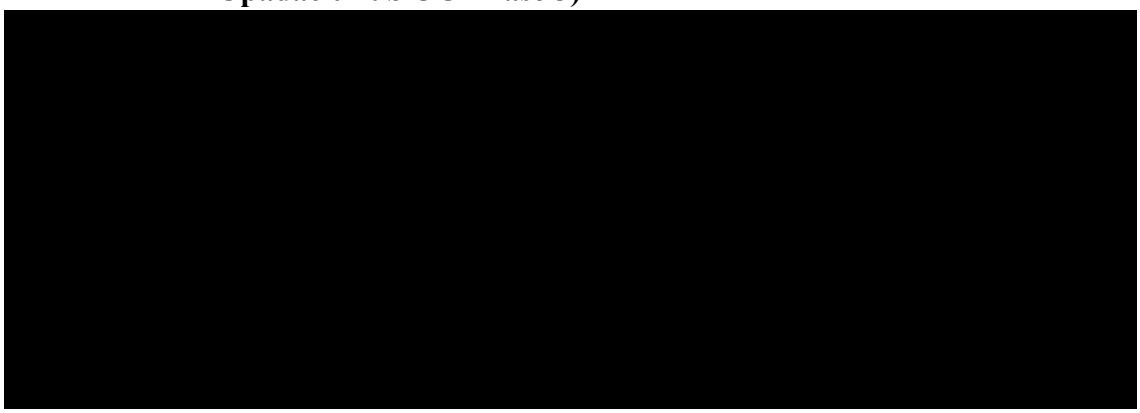
[redacted]

Prior to the database lock for the interim analysis, Upadacitinib UC Phase 3 study data became available. This Phase 3 study has the same population, inclusion and exclusion criteria with Upadacitinib UC Phase 2b study as shown in Table 5.

- Tofacitinib Phase 3
  - OCTAVE – Induction 1 (4/65)
  - OCTAVE – Induction 2 (4/65)
- Stelara Phase 3 UNIFI – Induction (11/161).
- [redacted]
- [redacted]

A historical rate of [redacted] is obtained based on the meta-analysis, see Figure 3 as below. The prior distribution was updated to [redacted] Primary analysis will be conducted using this prior distribution for placebo.

Figure 3. Meta-analysis of Updated Historical Placebo Rate (including Upadacitinub UC Phase 3)



Historical rates for secondary endpoints at Week 8 are summarized in Table 6 below.

Table 6. Historical Placebo Data for Secondary Endpoints

Endpoints	Historical Data					
	OCTAVE-1/2	Stelara Ph3	[redacted]	[redacted]	[redacted]	[redacted]
Clinical remission per Adapted Mayo score	NA	NA	[redacted]	[redacted]	[redacted]	[redacted]
Clinical response per Adapted Mayo score	NA	NA	[redacted]	[redacted]	[redacted]	[redacted]
Clinical response per Partial Adapted Mayo score	NA	NA	[redacted]	[redacted]	[redacted]	[redacted]
Clinical remission per Full Mayo score	NA	2/161	[redacted]	[redacted]	[redacted]	[redacted]
Endoscopic remission	NA	NA	[redacted]	[redacted]	[redacted]	[redacted]

NA = data are not applicable in subpopulation for respective endpoints.

Note that these five trials above enrolled a mixed population of subjects with anti-TNF-naïve or anti-TNF-IR, Bio-IR or non Bio-IR. Only Bio-IR subgroup data was used from these studies with mixture population.

D2. Prior Distribution Determination

Table 7 shows the false positive rate and true positive rate for different assumed true placebo response rates and ravagalimab response rates given the historical placebo response rate is [redacted] respectively. Note the range of this placebo rate is determined by the 90% confidence interval of placebo rate based on meta-analysis. False positive rate is calculated as the probability that the posterior probability of rate difference > 0 is larger than 90% assuming there is no treatment difference between ravagalimab and placebo. True positive rate is calculated as the probability that the posterior probability of rate

difference > 0 is larger than 90% assuming there is a 25% treatment difference between ravagalimab and placebo.

[redacted]

[redacted] prior corresponds to borrowing [redacted] subjects with a historical placebo rate of [redacted] and [redacted] prior corresponds to borrowing [redacted] subjects with a historical placebo rate of [redacted]. Operating characteristics of using [redacted] prior are the similar as using [redacted] prior based on the calculation.

Based on Table 7, within 90% confidence interval of placebo rate [redacted] 20 historical subjects or less could be borrowed while controlling false positive rates < 10% and not exceeding in study sample size.

D3. Comparison between ravagalimab and Placebo.

Posterior probability of rate difference [redacted] given the endoscopic improvement at Week 8 results, i.e., [redacted] will be calculated as follows. [redacted] are the response rates of ravagalimab and placebo, respectively.

Let n and N be the number of responders and number of subjects at Week 8. The conjugate priors used for [redacted] are [redacted] respectively. Therefore the posterior distributions for $\theta_{\text{ravagalimab}}$ and $\theta_{Placebo}$ will be [redacted] respectively. Note that in primary analysis, the

response rate is estimated using NRI-C which incorporates MI to handle missing data due to Covid-19; then the number of responders (n) is calculated based on the total number of subjects and estimated response rate, without rounding to an integer, to calculate the posterior probability.

Then the posterior probability is given by

[redacted]

where B(p, q) is the Beta function with parameters p and q, δ is corresponding to the pre-specified treatment difference under the proposed alternative hypothesis.

The posterior mean difference will be determined as [redacted].

In addition, [redacted] credible interval for $\theta_{\text{ravagalimab}} - \theta_{Placebo}$ will be calculated from

and [redacted]

where CI_U and CI_L are [redacted] credible interval upper limit and lower limit.

Similar comparison will be performed using [redacted] prior for placebo group. Note that if observed response rates in ravagalimab group and historical placebo group are both zero, no comparison will be performed using Bayesian approach.

Appendix E. Details for Analysis with a Synthetic Placebo Arm

E1. Historical Data Selection

The historical data are carefully selected based on the similarity in target population, entry criteria, prior and concomitant medications, central review of endoscopies and years conducted between historical trials and Study M15-722. Subject-level historical placebo data consists of subjects from two recent AbbVie clinical trials with similar populations (Bio-IR), entry criteria and study designs: upadacitinib phase 2b/3 Study M14-234 Substudy 1 and Substudy 2. All bio-IR placebo subjects from M14-234 Substudy 1 and Substudy 2 will be identified.

E2. Propensity Score Estimation

Propensity score method will be used to balance the population-level baseline characteristics between in trial data and historical placebo group. In order to calculate propensity scores, selected common baseline demographics and characteristics variables among Study M15-722 and the two historical Studies M14-234 Substudy 1 and Substudy 2 bio-IR will be used as covariates in propensity score estimation.

Main baseline demographic characteristics can include:

- age (years)*,
- weight (kg)*,
- body mass index (BMI) (kg/ m^2).
- sex (male, female)*,
- ethnicity (Hispanic or Latino, not Hispanic or Latino)
- race (white, non-white),
- region (Europe, North America, or rest of world),
- tobacco user (current, former, never, unknown),
- alcohol user (current, former, never, unknown).

* indicating those demographic characteristics must be included.

Main baseline disease characteristics can include

- adapted Mayo score*,
- hs-CRP (mg/L),
- fecal calprotectin (mg/kg),
- albumin (g/L),
- UC disease duration in years,
- UCEIS,
- Geboes score,
- baseline steroid use (yes, no)*,
- baseline immunosuppressant use (yes, no)*,
- baseline aminosalicylates use (yes, no),
- prior exposure to biologic/JAK therapy for UC (1, >1)
- prior exposure to anti-TNF use for UC (yes, no),
- mayo endoscopic subscore (2, 3)*,
- disease extent (left-sided, extensive/pancolitis, limited to rectum)*

* indicating those baseline disease characteristics must be included due to their clinical relevance. Other baseline characteristics (and their specific levels of categorical variables) will be assessed in model diagnostics if data does not allow large number of covariates when estimating propensity score.

Propensity score is the probability of a patient being assigned to a treatment instead of the other one conditional on a given set of baseline characteristics. It is determined by a logistic regression with selected above common baseline covariates. Overlap and balance of propensity score between the in-trial data and synthetic placebo control will be assessed using histogram and descriptive statistics.

Specifically, covariate balance will be checked between all subjects enrolled in Study M15-722 and historical placebo subjects from the two historical trials using standardized mean difference (SMDs) and a threshold of 0.1 for each of the baseline characteristics. To assess overall performance of different synthetic control algorithms, the mean of absolute SMDs across all baseline characteristics that are included in estimating propensity score is the main measure with smaller values indicating better balance. Considering the modest sample size in the current trial, sample size preservation in the matched patients is another key consideration in determining the final recommendation.

E3. Analysis using placebo synthetic control with propensity score matching

Propensity score matching method (Rosenbaum and Rubin, 1985) matches treatment patients with synthetic placebo subjects based on propensity scores such that matched treatment subjects and placebo subjects are comparable in terms of covariates used in propensity score determination. This process mimics randomization to create two comparable groups with a caution of limitation that the matching is conditional on included covariates.

A 1:1 matched placebo group will be generated from the above set of synthetic placebo subjects using greedy matching to match the treatment subjects in Study M15-722. We will start the matching with the optimal caliper 0.2, which was recommended by Austin (2011), with the intention to match 40 placebo subjects. Other caliper values or matching ratios may also be considered to improve balance and preserve sample size in matched patients. If multiple placebo subjects have propensity scores that are equally close to that of the treated subject, one of these untreated subjects will be selected at random. Subjects who are not selected in the matching process will be excluded from further analysis.

Following matching, baseline covariates will be summarized for subjects in Study M15-722 and matched historical placebo subjects to ensure balance is generally achieved.

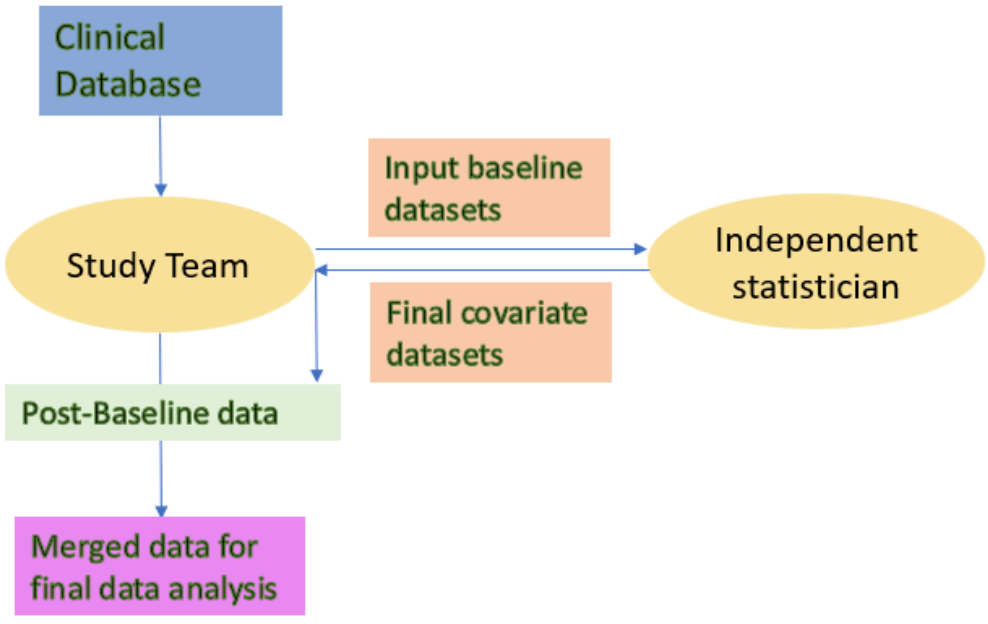
The optimal matching will also be performed based on a distance measure, caliper, calculated from the propensity scores estimates.

E4. Implementation

To reduce operational bias and maintain study validity, the propensity score calculation and matching described in Appendix E will be conducted by an independent statistician who is removed from Study M15-722. Prior to the first dose of the 40th subject, the study team and the independent statistician will meet as needed and agree on the statistical details of each step above, with any question (must be unrelated to treatment assignment information and in trial and historical subjects outcome) from the independent statistician resolved during the meeting(s).

The process to implement the propensity score analysis is summarized in Figure 4. Each step is described in detail below. These steps should happen after 40 subjects enrolled into Study M15-722 and prior to the primary database lock.

Figure 4. Flowchart of Implementation



1. Input baseline datasets of Study M15-722

A dataset of Study M15-722 containing study ID, de-identified subject IDs, and baseline characteristics will be delivered to the independent statistician by the study team. No post-baseline data should be contained in this dataset.

2. Input integrated baseline dataset of historical placebo subjects

A dataset integrating the baseline characteristics from the two studies with historical placebo subjects will be delivered to the independent statistician by the study team. No post-baseline data should be contained. In addition, study IDs and subject IDs in this dataset will be de-identified by the study programing team before delivering to the independent statistician.

3. Post-baseline dataset of historical placebo subjects

A dataset of historical placebo subjects integrating endoscopic improvement at Week 8 from Study M14-234 Substudy 1 and Substudy 2 will be generated. Subject ID will be included. This dataset is to be kept by study team for the final analysis and should not be accessible to the independent statistician.

4. Propensity score analysis and communication with study team

Prior to enrollment completion, the study team and the independent statistician will meet as needed and agree on the final propensity score analysis rules (for example, matching ratio, matching caliper) of each step. Any question (must be unrelated to treatment assignment information) from the independent statistician will also be addressed during these meeting(s).

After Study M15-722 enrolled 40 subjects, the input datasets will be delivered to the independent statistician for propensity score estimation and matching, without accessing the subject-level post-baseline data (including primary endpoint data) of selected historical placebo subjects and Study M15-722 subjects. The independent statistician will

check the covariate balance before and after matching is performed to make sure the balance is improved following the matching.

5. Final covariate datasets

The resultant datasets (matched) containing subject-level covariate data will be delivered by the independent statistician to the study team prior to the primary database lock for analysis. Matched datasets should discard those unmatched subjects. The analysis-ready datasets should include variables such as re-identified subject ID, baseline characteristics, propensity score.

There is no plan to share interim data with external experts or with Regulatory Agencies (prior to study completion and availability of final study results).

6. Final data analysis

After database lock, the study programming team will merge the final output datasets with datasets containing post-baseline data of ravagalimab and historical placebo subjects and perform the pre-specified analyses.