STATISTICAL ANALYSIS PLAN:
INDUCTION PERIOD ANALYSIS

Protocol VTX002-201

A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Clinical Efficacy and Safety of VTX002 in Subjects with Moderately to Severely Active Ulcerative Colitis

Protocol number:	VTX002-201
Compound:	VTX002
Sponsor:	Oppilan Pharma Ltd, a wholly owned subsidiary of Ventyx Biosciences, Inc. 5 New Street Square London UK EC4A 3TW
Version:	Final 1.1
Date:	25 Sep 2023
Author:	[redacted]

Document History

VERSION HISTORY

Version #	Version Date
1.0	19 Sep 2023
1.1	25 Sep 2023

REVISION HISTORY

Version #	Chapter	Revision Summary	Reason(s) for Revision
1.0	N/A	Initial release	N/A
1.1	8	Per Protocol Set definition clarified	To exclude from Per Protocol Set participants who received <80% or >120% of study treatment.
	15.1.2	Summary description clarified	To align with MockTFLs v1.0

APPROVAL SIGNATURES

STUDY TITLE: A Phase 2, Multicenter, Randomized, Double-Blind, Placebo-Controlled, Parallel-Group Study to Evaluate the Clinical Efficacy and Safety of VTX002 in Subjects with Moderately to Severely Active Ulcerative Colitis

PROTOCOL NUMBER: VTX002-201 Version 6.0, 21 July 2023

SAP Final 1.1 25 Sep 2023

PSI

Author [redacted]

Signature: ..... Date: .....

Project Manager [redacted]

Signature: ..... Date: .....

Peer-Review Statistician [redacted]

Signature: ..... Date: .....

Medical Writer [redacted]

Signature: ..... Date: .....

Ventyx Biosciences, Inc.
[redacted]

Signature: ..... Date: .....

[redacted]

Signature: ..... Date: .....

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. LIST OF ABBREVIATIONS

5-ASA	5-aminosalicylic acid
6-MP	6-mercaptopurine
AE	Adverse Event
ADaM	Analysis Data Model
ALC	Absolute Lymphocyte Count
ALT	Alanine Aminotransferase
ANC	Absolute Neutrophil Count
AST	Aspartate Aminotransferase
ATC	Anatomical Therapeutic Chemical
AZA	Azathioprine
BDRM	Blinded Data Review Meeting
BPM	Beats per minute
CI	Confidence Interval
CMH	Cochran-Mantel-Haenszel
CR	Copy Reference
CRP	C-reactive Protein
CSR	Clinical Study Report
CTCAE	Common Terminology Criteria for Adverse Events
CV	Coefficient of Variation
DLCO	Diffusing Capacity of the Lungs for Carbon Monoxide
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EDC	Electronic Data Capture Database
ES	Endoscopic Subscore
ET	Early Termination
FAS	Full Analysis Set
FAS-ET	Full Analysis Set-Extension Treatment
FCP	Fecal Calprotectin
FCS	Fully Conditional Specification
FEF	Forced Expiratory Flow
FEV	Forced Expiratory Volume
FVC	Forced Vital Capacity
JAK	Janus Kinase
HEMI	Histologic-endoscopic mucosal improvement
HRQoL	Health-Related Quality of Life
IBDQ	Inflammatory Bowel Disease Questionnaire
ICF	Informed Consent Form
INR	International Normalized Ratio

IWRS	Interactive Web Response System
LTE	Long-Term Extension
MAR	Missing at Random
MCS	Mayo Clinic Score
MCMC	Markov Chain Monte Carlo
MedDRA	Medical Dictionary for Regulatory Activities
mFAS	Modified Full Analysis Set
MMRM	Mixed Models with Repeated Measurements
MMS	Modified Mayo Score
MNAR	Missing Not at Random
OLE	Open-Label Extension
PDAP	Protocol Deviation Assessment Plan
PDP	Protocol Deviation Plan
PEFR	Peak Expiratory Flow Rate
PGA	Physician's Global Assessment
PK	Pharmacokinetic(s)
PMS	Partial Mayo Score
PP	Per Protocol
PT	Preferred Term
RB	Rectal Bleeding
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SF	Stool Frequency
SoA	Schedule of Activities
SOC	System Organ Class
TEAE	Treatment-Emergent Adverse Event
TLF	Tables, Listings and Figure
TMF	Trial Master File
TMS	Total Mayo Score
UC	Ulcerative Colitis
USA	United States of America
WHODD	WHO Drug Dictionary

2. INTRODUCTION

This Statistical Analysis Plan (SAP) covers the statistical analysis and reporting for the Induction Study Period for protocol VTX002-201 V.6.0, V.6.1, and V6.2, all dated 21 July 2023 and electronic Case Report Form (eCRF) V480 (V2.152 10AUG2023 0.81).

This is a Phase 2, multicenter, randomized, double-blind, placebo-controlled, parallel-group study to evaluate the clinical efficacy and safety of VTX002 in participants with moderately to severely active ulcerative colitis (UC).

This SAP includes a description of the planned pharmacokinetics (PK) analyses – see Section 30. The biomarker analyses are described in a standalone Biomarker Analysis Plan.

3. STUDY OBJECTIVES

3.1 Primary Objective

- Assess the efficacy of VTX002 when administered for 13 weeks on clinical remission

3.2 Secondary Objectives

- Assess the efficacy of VTX002 when administered for 13 weeks on endoscopic changes, symptomatic response and remission, histology, and mucosal healing
- Assess the safety and tolerability of VTX002
- Assess the PK of VTX002

3.3 Long-Term and Open-Label Extension Objectives

- Assess the efficacy of VTX002 through the Long-Term Extension (LTE) and Open-Label Extension (OLE) Treatment Periods on endoscopic changes, symptomatic response and remission, histology, and mucosal healing
- Assess the safety of VTX002 through the LTE and OLE Treatment Periods

3.4 Exploratory Objectives

- Assess the effect of VTX002 on health-related quality of life (HRQoL) outcomes and biomarkers

3.5 Estimands

The primary and key secondary estimands are described in Table 1 below. The analyses will be performed on the Full Analysis Set (FAS) as defined in Section 8. Other Secondary and Exploratory efficacy endpoints (specified in Sections 16.3 and 16.4) will generally be subject to the population, intercurrent event handling strategy, and population-level summary measures similar to the primary estimand. Supplementary analyses for the primary and key secondary

estimands will be performed using the FAS, Modified Full Analysis Set (mFAS), and Per Protocol (PP) populations.

Intercurrent events include exposure to rescue therapy for UC – see Section 25, Appendix IV for the list of rescue therapies applying to this study. Additional intercurrent events include missed endoscopy at Visit 7 Week 13 for participants available to have a Visit 7 Week 13 endoscopy, and study discontinuation for either disease worsening or treatment-emergent adverse event with a preferred term indicating ulcerative colitis. See Sections 16, 16.1.2, and 29 (Appendix VIII).

Participants will be evaluated for exposure to rescue therapy by Ventyx reviewers according to the guidelines presented in Section 25, Appendix IV. Participants deemed to have received rescue therapy, and not experiencing other deviations preventing contribution to the primary and key secondary analyses of the primary and key secondary efficacy endpoints, will be treated as non-responders in those analyses; they will also be included in the Per Protocol analysis set. Use of concomitant medications and procedures that are not potentially effective for the treatment of UC will not be considered intercurrent events.

Table 1. List of Primary and Key Secondary Estimands

Estimand	Definition	Attributes			
		Population	**Variable / Endpoint**	**Primary Intercurrent Event Handling Strategy**	**Population-Level Summary Measure**
Primary Estimand	Efficacy of VTX002 on Clinical Remission (Modified Mayo Score) at Week 13	FAS	Proportion of participants with stool frequency (SF) subscore = 0 or 1, rectal bleeding (RB) subscore = 0, and endoscopic subscore (ES) ≤ 1 (excluding friability)	Participants with any intercurrent events prior to Week 13 efficacy assessment will be treated as non-responders	Difference between VTX002 and Placebo with respect to population proportion of clinical remission at Week 13
Key Secondary Estimand 1	Efficacy of VTX002 on Endoscopic Improvement at Week 13	FAS	Proportion of participants with ES ≤ 1 (excluding friability)	Same as the intercurrent event strategy for the primary estimand	Difference between VTX002 and Placebo with respect to population proportion of endoscopic improvement at Week 13
Key Secondary Estimand 2	Efficacy of VTX002 on Symptomatic Remission at Week 13	FAS	Proportion of participants with SF subscore = 0 or 1 and RB subscore = 0	Same as the intercurrent event strategy for the primary estimand	Difference between VTX002 and Placebo with respect to population proportion of symptomatic remission at Week 13
Key Secondary Estimand 3	Efficacy of VTX002 on Histologic Remission at Week 13	FAS	Proportion of participants with Geboes Index score < 2.0	Same as the intercurrent event strategy for the primary estimand	Difference between VTX002 and Placebo with respect to population proportion of histologic remission at Week 13
Key Secondary Estimand 4	Efficacy of VTX002 on Endoscopic Improvement-Histologic Remission at Week 13	FAS	Proportion of participants with ES ≤ 1 (excluding friability) and Geboes Index score < 2.0	Same as the intercurrent event strategy for the primary estimand	Difference between VTX002 and Placebo with respect to population proportion of mucosal healing at Week 13

4. STUDY DESCRIPTION

4.1 Overall Design

This is a multicenter, randomized, double-blind, placebo-controlled study to evaluate the efficacy and safety of VTX002 30 mg and 60 mg in participants with moderately to severely active UC following daily oral administration of VTX002 as a tablet. Approximately 189 eligible participants will be randomized in a 1:1:1 ratio to receive VTX002 30 mg, VTX002 60 mg, or matching placebo, once daily (approximately 63 participants per treatment group).

The target patient population will include:

- Patients who have had an inadequate response, loss of response, or intolerance to conventional therapy and are naïve to biologic/Janus kinase (JAK) inhibitors (conventional failed).
- Patients who have had an inadequate response, loss of response, or intolerance to a biologic/JAK inhibitor (biologic/JAK inhibitor failed). Patients in this category may have received prior conventional therapy. It is expected that approximately 35% of participants in the study will have had an inadequate response to biologics.

Participant randomization under Protocol v.3 and later are stratified based on (a) biologic/ JAK inhibitor prior use status (yes/no), (b) baseline corticosteroid use (yes/no) and (c) baseline disease activity (Modified Mayo Score [MMS] 5-6 or 7-9). Participants randomized under Protocol v.1 were randomized using the following definitions of stratification factors: (a) biologic/JAK inhibitor ***failed*** status (yes/no), (b) baseline corticosteroid use (yes/no) and (c) baseline disease activity (MMS ***4***-6 or 7-9). Reconciliation of stratification factors will be performed via mapping at the Analysis Data Model (ADaM) dataset level prior to statistical analysis. See Section 10.2 for details.

Participants experiencing mis-stratification (i.e., randomized stratification factors in interactive web response system [IWRS] do not match stratification factors as collected via electronic data capture database [EDC] system) will be treated in the primary analysis of efficacy endpoints according to their randomized stratification factors. Supportive analyses based on actual strata membership (taking stable systemic corticosteroids, previously taking biologics and/or JAK inhibitors, and baseline MMS as collected in EDC) will be performed. See Section 16.1.7.

As shown in Figure 2, the study consists of a 28-day Screening Period, a 13-week double-blind Induction Treatment Period (including 7 days of titration followed by 12 weeks of treatment at the assigned dose), an LTE Treatment Period of up to 39 weeks, an OLE Treatment Period of up to 143 weeks, and a 2-week Follow-Up Period.

The duration of study participation for each participant is planned to be approximately 162 weeks, and the total duration of the study is planned to be approximately 60 months. Approximately 120 study centers in 15 countries are expected to participate.

4.1.1 Screening Period

Participant eligibility will be determined during a 4-week (28-day) Screening Period. Entry criteria will be based on confirmation of moderately to severely active UC, defined by an MMS of 5 to 9 with an endoscopic subscore (ES) ≥ 2 and rectal bleeding (RB) subscore ≥ 1.

4.1.2 Induction Treatment Period

On Day 1, eligible participants will be randomized in a 1:1:1 ratio to receive VTX002 30 mg, VTX002 60 mg, or matching placebo orally once a day (approximately 63 participants per treatment group). The treatment each participant will receive will not be disclosed to the Investigator, study center staff, participant, Sponsor, or study vendors (eg, central readers).

During the first week of the Induction Treatment Period, participants will undergo a 7-day dose titration up to the full assigned double-blind treatment dose (see Table 2 for more information).

At the end of the 13-week Induction Treatment Period, all participants will be evaluated for efficacy and safety.

4.1.3 Long-Term Extension Treatment Period

Participants who complete the Week 13 visit, and meet criteria for clinical response at Week 13, may continue into the single-blinded Long-Term Extension Treatment Period (LTE) at their previously assigned treatment for up to an additional 39 weeks, for a total of 52 weeks of treatment with assigned study treatment.

Study treatment assignment during this period will remain blinded to Investigator, study center staff, participant, and study vendors not engaged in data analysis associated with the Interim database lock for the Induction Treatment Period. The Sponsor will be unblinded to study treatment assignment for the LTE period after the Interim database lock (Section 5.1) and unblinding for the Induction Treatment Period.

Participants in the LTE experiencing loss of response during LTE will have the option to participate in the OLE. Loss of response in LTE will be assessed by the Investigator (Protocol Section 8.3.1.5).

Participants completing the Week 52 visit of the LTE may continue into the OLE.

4.1.4 Open-Label Extension Treatment Period

Section 4.1.4.1 only applies to study participants consented under Protocol v.1.
Sections 4.1.4.2 and 4.1.4.3 apply to study participants consented under Protocols v.3 or later.
The Sponsor will consider potential expansion of open-label access beyond the 36-month period, when the primary analysis results are available.

4.1.4.1 Completion of Induction Week 13 - Protocol Version 1.0

Participants consented under Protocol v.1 and not re-consented to Protocol v.3 or later before the Induction Treatment Period Week 13 visit may have participated in OLE regardless of clinical response at Induction Treatment Period Week 13. These participants may have

reconsented to Protocol v.3 or later during OLE, and their scheduled OLE visits followed the OLE Schedule of Activities (SoA) specified by the newer Protocol version. Otherwise, the scheduled OLE visits followed the 39-week SoA specified by Protocol v.1.

Participants consented under Protocol v.1 who did not have clinical response at Induction Treatment Period Week 13 must have been assessed at OLE Week 13 for symptomatic response. Such participants who failed to achieve symptomatic response by OLE Week 13 (defined as decrease from baseline ≥ 30% in composite RB and SF score) were discontinued from the study treatment and asked to complete the Early Termination (ET) and 1- and 2-Week Follow-up visits. However, participants who demonstrated symptomatic response at OLE Week 13 continued treatment with VTX002 60 mg in the OLE Treatment Period for up to an additional 26 weeks (a maximum OLE treatment duration of 39 weeks, and a maximum study treatment duration of 52 weeks).

4.1.4.2 Lack of Clinical Response at Induction Week 13 – Protocol Version 3.0 and Later

Participants who complete the Induction Week 13 visit, and do not meet the criteria for clinical response at Induction Week 13, will have the option to participate in the OLE for up to an additional 143 weeks of treatment (which includes a 7-day titration period followed by treatment with VTX002 60 mg for up to 142 weeks).

As non-responders, these participants will be assessed again for symptomatic response at OLE Week 13. At OLE Week 13, participants who fail to achieve symptomatic response will be discontinued from the study treatment and asked to complete the ET and 1- and 2-Week Follow-up visits. However, participants who respond to treatment by OLE Week 13 will continue treatment with VTX002 60 mg in the OLE Treatment Period for up to an additional 130 weeks (a maximum OLE treatment duration of 143 weeks, and a maximum study treatment duration of 156 weeks).

4.1.4.3 Loss of Clinical Response During LTE – Protocol Version 3.0 and Later

Participants in the LTE who lose response after Induction Week 13 will have the option to participate in the OLE. Loss of response is defined in Protocol VTX002-201 Section 8.3.1.5. Participants who enter the OLE from the LTE will undergo a 7-day titration period (see Section 4.3.3) and remain in the OLE for up to 143 weeks. These participants may have a maximum study treatment duration of 195 weeks (13 weeks for the Induction Treatment Period, 39 weeks for LTE, and 143 weeks for OLE).

4.1.5 Follow-Up Period

For all participants, Follow-Up visits will be performed at 1 and 2 weeks after the last dose of study treatment, as indicated in the Protocol SoA. All participants who discontinue treatment prematurely will be asked to return to the clinic for the ET visit. If an ET visit is ≥ 1 week after the last dose of study treatment, the 1-Week Follow-Up visit is not required; however, the 2-Week Follow-Up visit should be scheduled and completed.

Participants will be evaluated for safety and efficacy in accordance with the SoA. For PK, blood samples will be collected as shown in the SoA.

4.1.6 End of Study Definition

A participant is considered to have completed the study after completing all pertinent phases of the study including the Follow-Up visits.

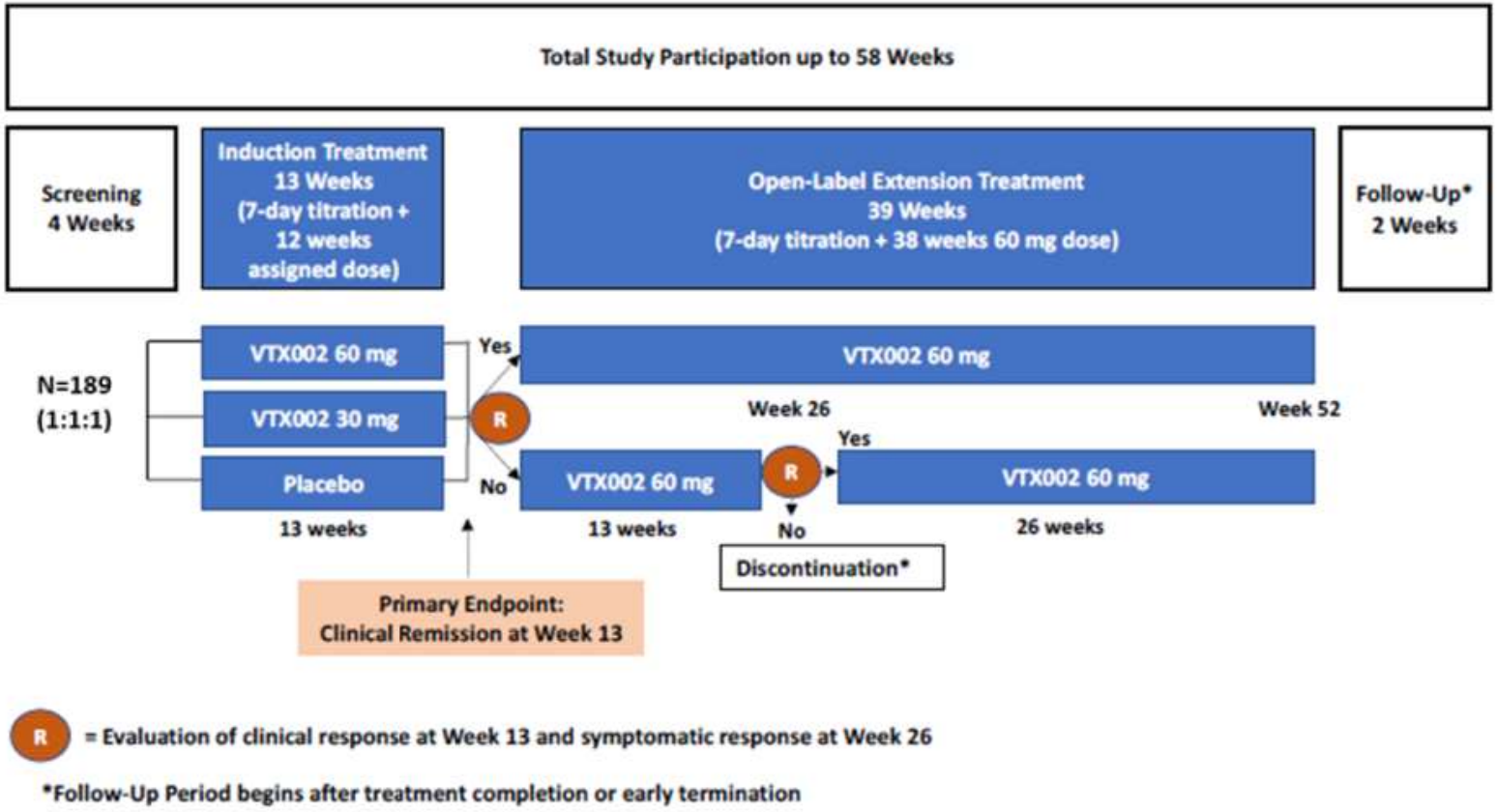
End of study is defined as when a participant completes the study or withdraws from the study for any reason prior to completion.

The end of the whole study is defined as the date of the last visit of the last participant in the study or last scheduled procedure shown in the SoA for the last participant in the study globally.

4.2 Study Schema

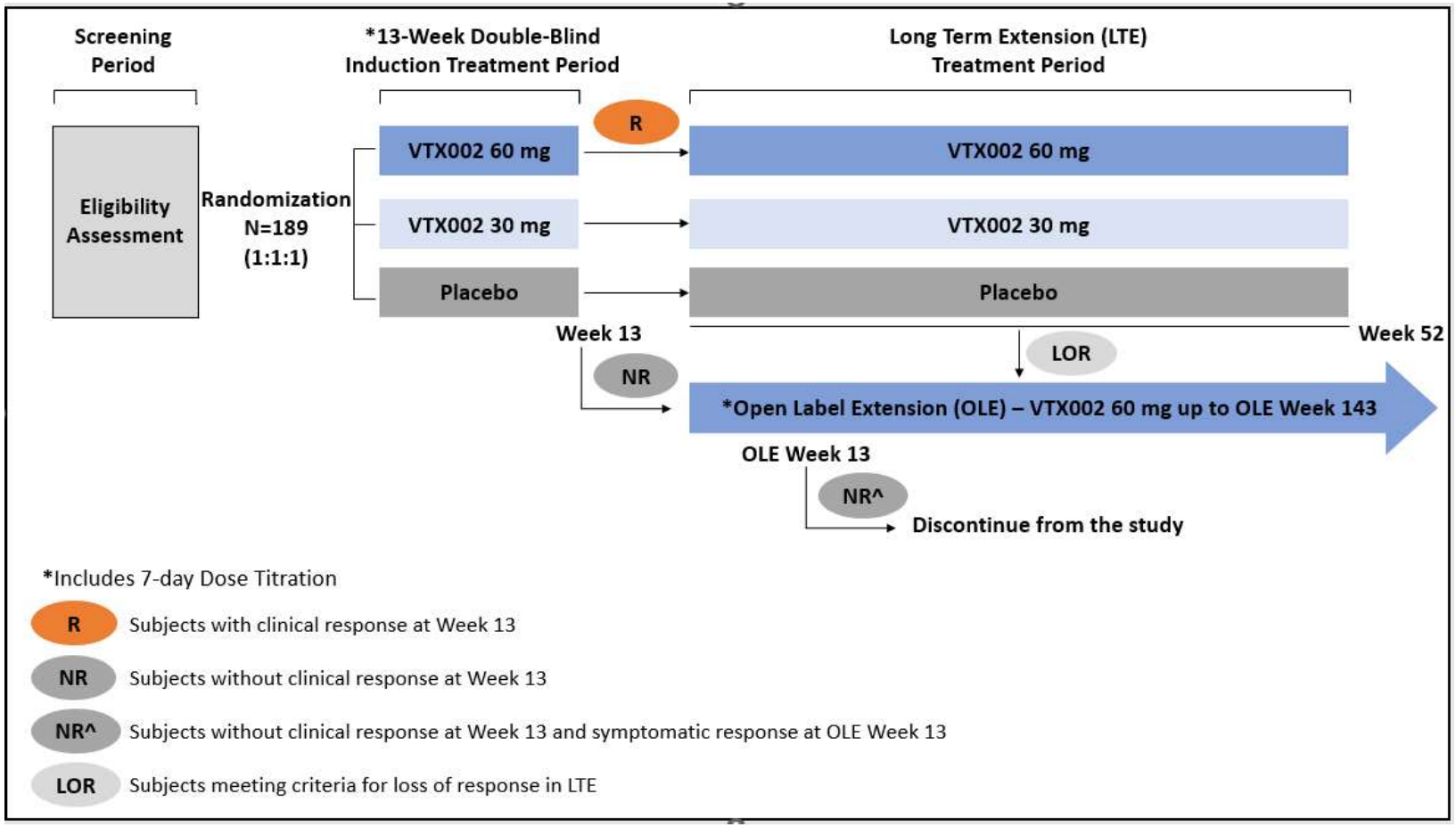
4.2.1 Protocol Version 1.0

Figure 1. Study Schema, Protocol Version 1.0



4.2.2 Protocol Versions 3.0 and Later

Figure 2. Study Schema, Protocol Versions 3.0 and Later



4.3 Study Treatment

Participants will be randomized in a 1:1:1 ratio to receive VTX002 30 mg, VTX002 60 mg, or matching placebo. Study medication compliance will be measured by tablet count. Overall study noncompliance is defined as taking less than 80% or more than 120% of study treatment during the entire treatment period.

4.3.1 Induction Treatment Weeks 0 to 13

Randomized participants will undergo a 7-day dose titration to receive their respective double-blinded drug doses from Day 1 to Day 7. The double-blind doses received by participants per day during a 7-day titration period and onwards will be as following for all treatment groups:

Table 2. Dose Titration in the Induction Treatment Period

Study treatment	Day 1 (Visit 2)	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8 and Onward (Visit 3)
VTX002 30 mg	3 mg	7 mg	10 mg	14 mg	20 mg	30 mg	30 mg	30 mg
VTX002 60 mg	3 mg	7 mg	10 mg	14 mg	20 mg	30 mg	50 mg	60 mg
Placebo	Placebo	Placebo	Placebo	Placebo	Placebo	Placebo	Placebo	Placebo

4.3.2 Long-Term Extension Treatment Weeks 13 (Visit 8) to 52

Dosing configuration for Weeks 13 (Visit 8) to 52 will be similar to dosing configuration during Induction Treatment Period starting from Day 8. Participants will take 4 tablets per day.

4.3.3 Open-Label Extension Treatment

Participants entering the OLE Treatment Period will undergo a 7-day dose titration (OLE Days 1 to 8) to receive open-label VTX002 60 mg following the same titration schedule for the VTX002 60 mg as shown in Table 2. Although dose titration would not be needed for participants who received VTX002 60 mg during the Induction Period/LTE Period, all participants entering the OLE Treatment Period will go through a full dose titration regimen to maintain the blind of the Induction Period/LTE treatment assignments.

5. PLANNED ANALYSES

Each planned analysis identified in this SAP will be performed after

- SAP is authorized by the study sponsor
- Database lock is completed as appropriate for the given planned analysis
- Study treatment is unblinded (if necessary) as appropriate for the given planned analysis

Any post hoc exploratory analyses completed to support planned study analyses not identified in this SAP will be documented and reported in the Clinical Study Report (CSR). Any results from these unplanned analyses will be clearly identified in the text of the CSR.

5.1 Interim Analyses

An interim database lock (iDBL), treatment unblinding, and data analysis will be conducted after all randomized participants have either completed the Induction Treatment Period or are no longer participating in the Induction Treatment Period for one of the following reasons:

- Participant enters OLE
- Participant enters LTE
- Participant withdraws from study

This interim analysis will be referred to as the Induction Period Analysis. The following rules will be used to determine which participant records are included in the Induction Period Analysis:

- For study participants entering one of the extension periods (either LTE or OLE), the Induction Period Analysis will be limited to data on or before the later of [a] the last available date of study drug administration during the Induction Treatment Period, and [b] latest visit date as reported in EDC, up to and including Visit 7 Week 13; and in cases where Visit 8/Week 13 occurs on the same day as Visit 7 Week 13, pre-dose Visit 8 data shall also be included. For concomitant medications and adverse events the start date will determine inclusion in Induction Period Analysis (or exclusion).
- For study participants not entering an extension period, all data is included

The focus of the Induction Period Analysis will be on data associated with the Induction Treatment Period – no OLE or LTE data will be incorporated into the Induction Period Analysis other than participant disposition information related to participant flow between study periods. The Induction Period Analysis will be limited to eCRF pages associated with the Induction Treatment Period as well as external vendor data (such as central laboratory data) associated with the Induction Treatment Period.

All data for the 13-week double-blind Induction Treatment Period will be analyzed. The primary and key secondary endpoints are measured and assessed at Induction Week 13, so this analysis will be considered final for the primary and key secondary endpoints. Outputs related to the Induction Treatment Period will be clearly labeled as such. The locked data used in support of the Induction Period Analysis will be treated as final and not subject to further data alterations. The SDTM and ADaM datasets derived from the locked data and used to support Tables, Listings, and Figures will be incorporated into a final Induction Treatment Period data package.

5.2 Final Analysis

After the last participant completes the study, data accumulated over the entire study including LTE/OLE Treatment Periods will be presented in summaries and listings.

6. DETERMINATION OF SAMPLE SIZE

Approximately 189 eligible participants will be randomized in the study. The study is powered to show superiority of VTX002 60 mg to placebo with respect to clinical remission at Week 13.

Based on literature data review, the expected proportion of clinical remission at Induction Week 13 is 28% for VTX002 60 mg and 8% for placebo (Sandborn 2016; Sandborn 2020). Under these assumptions, and with a 1:1 randomization ratio, two-group chi-squared test, and two-sided significance level of 5%, a sample of 57 participants per treatment group will be sufficient to achieve at least 80% power. Adding a 10% inflation for dropouts results in 63 participants per treatment group and 189 participants in total.

All sample size calculations were conducted in SAS 9.4.

7. ANALYSIS ENDPOINTS

7.1 Primary Endpoint

- The proportion of participants with clinical remission at Week 13 using MMS

7.2 Secondary Endpoints

Key Secondary Endpoints

- The proportion of participants with endoscopic improvement at Week 13
- The proportion of participants with symptomatic remission at Week 13
- The proportion of participants with histologic remission at Week 13
- The proportion of participants with endoscopic improvement-histologic remission at Week 13

Other Secondary Endpoints

- The proportion of participants with clinical response using MMS at Week 13
- The proportion of participants with endoscopic remission at Week 13
- The proportion of participants with endoscopic and histologic remission at Week 13
- The proportion of participants with endoscopic and clinical remission at Week 13
- The proportion of participants with endoscopic, histologic, and clinical remission at Week 13

- The proportion of participants with symptomatic remission at Weeks 4, 8, 10
- The proportion of participants with symptomatic response at Weeks 4, 8, 10, 13
- The proportion of participants with clinical remission using total Mayo Clinic score (MCS) at Week 13
- The proportion of participants with clinical response using total MCS at Week 13
- The proportion of participants with histologic improvement at Week 13
- The proportion of participants with histologic-endoscopic mucosal improvement (HEMI) at Week 13
- Proportion of participants with any decrease from baseline in Geboes Index score at Week 13
- The proportion of participants with UC-related hospitalizations
- The proportion of participants requiring UC-related surgeries, including colectomy
- Plasma concentrations of VTX002, assessed from samples collected predose at Weeks 0, 1, 4, 8, and 13 in the Induction Treatment Period, and Weeks 18, 26, 36, and 52 in the LTE Treatment Period; at the 1-Week and 2-Week Follow-Up visits; and at 2, 4, and 6 hours (± 15 minutes) postdose at Week 0

7.3 Safety Endpoints

- Incidence and severity of adverse events (AEs)
- Incidence and severity of laboratory abnormalities, and change from baseline in laboratory values (e.g., hematology and serum chemistry)
- Incidence of clinically significant vital sign abnormalities and changes from baseline

7.4 Exploratory Endpoints

The endpoints described here will be analyzed separately, where appropriate, for the Induction Treatment Period, LTE, and OLE. LTE and OLE analyses for the same endpoint may differ not only with respect to the participant set but also with respect to statistical methodology.

- Proportion of participants with histologic remission at Week 13 using RHI
- Proportion of participants with histologic remission at Week 13 using NHI
- Proportion of participants with histologic improvement at Week 13 using RHI
- Proportion of participants with histologic improvement at Week 13 using NHI
- Proportion of participants with clinical remission after 52 weeks of treatment
- Proportion of participants with clinical remission after 13 and 52 weeks of treatment

- Proportion of participants with symptomatic remission at Weeks 18, 26, 36, and 52
- Proportion of participants with endoscopic remission with histologic remission after 52 weeks of treatment
- Proportion of participants with endoscopic improvement-histologic remission after 52 weeks of treatment
- Proportion of participants with endoscopic and clinical remission after 52 weeks of treatment
- Proportion of participants with endoscopic remission, histologic remission, and clinical remission after 52 weeks of treatment
- Proportion of participants with HEMI after 52 weeks of treatment
- Proportion of participants with clinical response after 52 weeks of treatment
- Proportion of participants with symptomatic response at Weeks 18, 26, 36, and 52
- Proportion of participants with symptomatic response at OLE Week 13, and at each additional OLE visit
- Proportion of participants with any decrease from baseline in Geboes Index score at Week 52
- Change from baseline in fecal calprotectin (FCP) at Weeks 1, 4, 8, 13, 26, 36, 52, and at visits during the OLE
- Change from baseline in C-reactive protein (CRP) at Weeks 1, 4, 8, 13, 26, 36, 52, and at visits during the OLE
- Change from baseline in partial Mayo score (PMS) at Weeks 4, 8, 13, 18, 26, 36, 52, and at visits during the OLE
- Change from baseline in total MCS at Week 52
- Change from baseline in Inflammatory Bowel Disease Questionnaire (IBDQ) score at Weeks 13 and 52
- Change and percentage change from baseline in lymphocyte counts at Weeks 1, 4, 8, 13, 26, 36, and 52

7.5 Endpoint Definitions

Clinical remission: SF subscore = 0 or 1, RB subscore = 0, and ES ≤ 1 (excluding friability)

Clinical response using MMS: A ≥ 2-point and ≥ 30% decrease from baseline in MMS, and a ≥ 1-point decrease from baseline in RB subscore or an absolute RB subscore ≤ 1

Clinical remission using total MCS: Total MCS ≤ 2 points with no individual subscore > 1 point (SF subscore = 0 or 1, RB subscore = 0, ES ≤ 1 [excluding friability], Physician Global Assessment [PGA] ≤ 1, and SF + RB + ES + PGA ≤ 2)

Clinical response using total MCS: A ≥ 3-point and ≥ 30% decrease from baseline in total MCS, and a ≥ 1-point decrease from baseline in RB subscore or an absolute RB ≤ 1.

Endoscopic improvement: ES ≤ 1 (excluding friability)

Endoscopic remission: ES = 0

Histologic improvement

- Primary: Geboes Index score < 3.1
- Sensitivity: Robarts Histopathology Index (RHI) ≥ 50% reduction from baseline or RHI ≤ 3
- Sensitivity: Nancy Histopathology Index (NHI) ≥ 1 point reduction from baseline

Histologic remission

- Primary: Geboes Index score < 2.0
- Sensitivity: RHI ≤ 3 with scores of 0 for both Geboes Grade 2B and Grade 3
- Sensitivity: NHI ≤ 1

Endoscopic improvement-histologic remission: ES ≤ 1 (excluding friability) and histologic remission measured by a Geboes Index score < 2.0

Endoscopic remission with histologic remission: ES = 0 and histologic remission measured by a Geboes Index score < 2.0

Histologic-endoscopic mucosal improvement (HEMI): ES ≤ 1 (excluding friability) and histologic remission measured by a Geboes Index score < 3.1

Symptomatic remission

- Primary: SF subscore = 0 or 1 and RB subscore = 0
- Sensitivity: SF subscore = 0 (or =1 with a ≥ 1-point decrease from baseline) and RB subscore = 0

Symptomatic response: Decrease from baseline ≥ 30% of the combined RB and SF subscores

Endoscopic and clinical remission: SF subscore ≤ 1, RB subscore = 0, and ES subscore = 0

Endoscopic remission with histologic remission and clinical remission: SF subscore ≤ 1, RB subscore = 0, ES subscore = 0, and histologic remission measured by a Geboes Index score < 2.0

8. ANALYSIS SETS

The following analysis sets for the Induction Treatment Period are defined as follows:

Full Analysis Set (FAS): The FAS consists of all randomized participants receiving at least 1 dose of study treatment. Under this approach, participants are counted in the treatment group to which they were randomized, regardless of the treatment received during the study.

Modified Full Analysis Set (mFAS): For a given endpoint (see Section 23, Appendix II), mFAS consists of all randomized participants who received at least 1 dose of study treatment and contribute at least one complete endpoint assessment for at least one post-baseline Induction visit. Participants will be analyzed according to the treatment to which they were randomized.

Per Protocol (PP) Set: The PP Set consists of all participants in the primary endpoint mFAS who received ≥80% and ≤120% of study treatment and do not have a protocol deviation preventing contribution to the primary analysis of the primary efficacy endpoint. Participants experiencing an intercurrent event during the Induction Treatment Period and not experiencing other deviations preventing contribution to the primary and key secondary analyses of the primary and key secondary efficacy endpoints will be included in the PP set and treated as non-responders in these analyses (Section 3.5).

Safety Set: The Safety Set consists of all randomized participants who received at least 1 dose of study treatment. For this population, participants are analyzed according to the treatment received, regardless of randomization. The Safety Set will be used for all safety analyses in the Induction Treatment Period.

The following analysis sets for the LTE and OLE Treatment Periods are subsets of the analysis sets of the Induction Treatment Period, as defined below:

FAS-Extension Treatment (FAS-ET) Set: The FAS-ET Set consists of all randomized participants who received at least 1 dose of study treatment in the LTE or the OLE Treatment Period. Under this approach, the original treatment groups as assigned at randomization will be used throughout LTE. In general, participants will be analyzed according to protocol-specified intended study treatment dose in FAS-ET analyses. Additional details are provided for specific analyses described elsewhere in this SAP.

Safety-Extension Treatment (Safety-ET) Set: The Safety-ET Set includes all participants who received at least 1 dose of study treatment in the LTE or the OLE Treatment Periods. The Safety-ET Set will be used for all safety analyses. In general, participants will be analyzed according to actual study treatment dose in Safety-ET analyses. Additional details are provided for specific analyses described elsewhere in this SAP.

9. GENERAL CONSIDERATIONS

9.1 Summary Statistics

In general, continuous variables will be summarized using the following standard descriptive summary statistics: number of observations, mean, standard deviation (SD), minimum, first quartile, median, third quartile, maximum, and number of missing observations. Use of decimal places in descriptive statistics will be as follows:

- Min, Max: same as the actual data (but not more than 4 decimals)

- Mean, Geometric Mean, Median, First Quartile, Third Quartile: actual data + 1 decimal (but not more than 4 decimals)
- SD, Geometric SD: actual data + 2 decimals (but not more than 4 decimals)
- Geometric CV: 2 decimals (but not more than 4 decimals)

Categorical data will be described using frequency counts and percentage. Percentages will be based on the number of participants in the selected analysis set and relevant treatment group unless otherwise stated. Percentages will be rounded to one decimal place except for one hundred percent, which will be presented as 100%. Tests will use a two-sided alpha=0.05 unless otherwise specified.

9.2 OLE Visit Labels

Visit labels for the OLE in the EDC system are different between Protocol version 1.0 and later protocol versions. Refer to Section 24 (i.e., Appendix III) for a visit matrix mapping that allows a common OLE visit label schema across all participants regardless of the protocol version.

9.3 Reference Start Date and Study Day

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

- If the date of the event is on or after the reference start date, then:

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

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

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

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

9.4 Body Mass Index

Body mass index is calculated from Screening weight (kg) and height (m) and retains one significant digit to the right of the decimal prior to summarization:

$$\text{BMI } (kg/m^2) = [\text{weight in kilograms}] / [(\text{height in meters})^2]$$

9.5 Baseline

The last non-missing assessment made prior to the first dose of Induction Treatment Period study treatment (including unscheduled assessments) will be used as baseline for both the Induction Treatment Period and LTE. In the event a participant is randomized but never treated, the randomization date will be used instead of the date of first dose of study product. OLE may utilize two different baselines:

- Last non-missing assessment made prior to the first dose of Induction Treatment Period study treatment (including unscheduled assessments)
- Last non-missing assessment prior to the first dose of OLE study treatment (including unscheduled assessments)

If measurements include time, the date/time will be used to define Baseline. Otherwise, only dates will be compared. In the event the last non-missing observation and date of first dose have the same date and time of day is not collected, if the measurement was planned per protocol to be captured prior to the first dose, the measurement will be considered Baseline.

9.6 Windowing Conventions

All scheduled study visits are defined relative to Study Day 1, the date of first Induction Treatment Period dose.

Scheduled visit windows for the Induction Treatment Period are defined in Protocol VTX002-201 Section 1.2 Table 1. Scheduled visit windows for the LTE are defined in Protocol VTX002-201 Section 1.2 Table 2. Scheduled visit windows for the OLE are defined in the Protocol VT002-201 Section 1.2 Tables 3, 4, 5, and 6.

A windowing convention will be used to determine the analysis visit value for a given measurement and will be applicable for all by-visit summaries and analyses for efficacy and safety data. Refer to Table 3 for specific Induction Treatment Period visit windows.

Windowing will be applied prior to any missing data calculations. The last non-missing measurement taken prior to Day 1 (including unscheduled assessments) will be labeled as Baseline. Unless stated otherwise, data from scheduled, unscheduled, and early termination site visits will be eligible for allocation to an analysis visit. Follow-up visit data is handled as-is.

Table 3. Analysis Visit Windows

	Visit Label	Study Week	Ideal Study Day (Study Day)	Protocol Visit Window (Study Day)	Analysis Visit Window (Study Day)
Non-Efficacy, Physician's Global Assessment (PGA), IBDQ [a]	Visit 1/Baseline	n/a	-14	[-28, -1]	≤1
	Visit 2	0	1	[1]	[1, 4]
	Visit 3	1	8	[8, 10]	[5, 18]
	Visit 4	4	29	[26, 32]	[19, 42]
	Visit 5	8	57	[54, 60]	[43, 63]
	Visit 6	10	71	[68, 74]	[64, 81]
	Visit 7	13	92	[89, 95]	[82, 102]
Endoscopy and Colonic Biopsy [b]	Visit 1/Baseline	n/a	-14	[-28, -1]	≤ 1
	Visit 7	13	92	[89, 95]	[76, 108]
SF and RB calculated from eDiary entries [b, c]	*see footnote*	*see footnote*	*see footnote*	*see footnote*	*see footnote*
Early Term /Unscheduled [d]	*see footnote*	*see footnote*	*see footnote*	*see footnote*	*see footnote*
1-Week Follow-up [d]	1-Week Follow-up	n/a	X+7	n/a	n/a
2-Week Follow-up [d]	2-Week Follow-up	n/a	X+14	n/a	n/a

[a] Analysis visit windows and Ideal Study Days are intended to map, to Study Visits, all by-visit data collected during the Induction Treatment Period.

[b] When endoscopy occurs: SF and RB scores are computed using a 7-day look back window whose ending Study Day is immediately prior to the first day of bowel preparation, and these SF and RB scores are attached to the endoscopy. If no bowel preparation is performed for a given endoscopy, the 7-day look back window ends at the Study Day immediately prior to the endoscopy. If the endoscopy occurs outside of both Visit 1/Baseline and Visit 7 analysis windows for endoscopy, the SF, RB, ES, and histopathology data for the endoscopy are not associated with any analysis visit and are listed only. If the endoscopy occurs inside either the Visit 1/Baseline or Visit 7 endoscopy analysis window, the SF, RB, ES, and histopathology data are associated with Visit 1/Baseline or Visit 7 as appropriate. If multiple endoscopies are associated with the same Visit 1 or Visit 7 analysis visit window, the following sequential steps will be applied to choose a unique endoscopy to associate with the analysis visit: [i] endoscopies with missing ES scores are dropped from consideration and, [ii] if there are remaining endoscopies to choose from, the most recent endoscopy is kept. After determining a unique endoscopy to associate with the analysis visit, if the 7-day look back window overlaps with a previous set of endoscopy and/or bowel preparation dates, the eDiary SF and RB values for those overlapping dates will be set to missing for computational purposes.

[c] For a given post-baseline analysis visit, if no endoscopy is associated with that analysis visit SF and RB will be computed using a 7-day look back window whose ending Study Day is immediately prior to the site visit closest to the Ideal Study Day when a within-window site visit exists. If two site visits are equidistant from the Ideal Study Day, the earlier site visit is chosen. If no site visit is associated with the given post-baseline analysis visit, SF and RB will be not be computed for that visit. If the 7-day look back window overlaps with a previous set of endoscopy and/or bowel preparation dates, the eDiary SF and RB values for those overlapping dates will be set to missing for computational purposes.

[d] Post-baseline ET and Unscheduled visits may include endoscopy and analysis visit windows are applied to ET and Unscheduled visit data in a manner identical to that of other visits. Endoscopy data from ET and Unscheduled visits may not be associated with any analysis visit other than Visit 7. Endoscopy

data from ET and Unscheduled visits not associated with Visit 7 will be listed only. 1-Week and 2-Week Follow-up data is not windowed and presented according to nominal visit label in tables.

For composite scores such as MMS, if one or more components is missing from an analysis visit window the composite score is missing. If one or more results for a variable are assigned to the same analysis visit, the result with the date closest to the protocol Ideal Study Day will be used in the analysis, except for rectal bleeding and stool frequency component subscores used in the determination of Symptomatic Remission, Symptomatic Response, MMS, or total MCS. In this case, 7-day look-back windows anchored to either an endoscopy or a site visit are used to determine the component subscores.

If two measurements in the same analysis visit window are equidistant from the protocol scheduled study day, the earliest measurement will be used in the analysis. If multiple assessments are available on the same day, then the average of the assessment will be used in the analysis, except for laboratory data where the assessment at the earliest time of the same day will be used. Central laboratory assessments will be used in both summaries and listings, while local laboratory assessments will only be separately listed. Local laboratory data may be used by investigators to help inform participant safety decisions and adverse event reporting.

9.7 Statistical Tests

The default statistical significance level will be 0.05 for two-sided hypothesis tests; one-sided hypothesis tests will not be performed unless explicitly stated otherwise. The default for confidence intervals (CIs) will be to generate two-sided 95% CIs.

9.8 Common Calculations

For quantitative measurements:

- Change from Baseline = Test Value at Visit X – Baseline Value
- Percent change from Baseline = (Test Value at Visit X – Baseline Value) / Baseline Value × 100
- Proportion at Visit X = Number of participants satisfying criteria at Visit X / Total number of participants at Visit X

9.9 Statistical Software and CDISC Implementation

All analyses will be performed and data appendices will be created using SAS® software (version 9.4 or higher; SAS Institute Inc.; Cary, NC, USA). Analysis Datasets produced for analysis will follow CDISC ADaM standard IG 1.3. ADaM specifications will follow conformance rules for Define.xml v.2.0. CDISC compliance of datasets will be validated using Pinnacle 21® Enterprise software (Certara; Blue Bell, PA, USA) and Pinnacle reports will be produced.

9.10 Study Information

A general table with summary of study information will be generated, including the date of first participant signed informed consent form (ICF), and the last participant visit date and the database lock date for each unique database lock.

10. STATISTICAL CONSIDERATIONS

10.1 Missing Data

The following rules will be applied for AEs, prior medications, and concomitant medications with incomplete start dates:

- If day is missing for start date and month and year are the same as month and year for the first treatment date, then start date is set to the treatment start date.
- If day is missing for start date and month and year are not the same as month and year for the first treatment date, then impute Day 1 of the month.
- If day and month are missing for start date and the year is the same as the year for the first treatment date, then start date is set to the first treatment date.
- If day and month are missing for start date and the year is less than year for the first treatment date, then impute July 15th of the year.
- If day and month are missing for start date and the year is greater than year for the first treatment date, then impute January 1st of the year.
- If the imputed start date is greater than or equal to the corresponding non-missing end date, then start date equals the end date.

The following rules will be applied for AEs, prior medications and concomitant medications with incomplete end dates:

- If day is missing for end date and month and year are non-missing, then day will be imputed as last day of the month.
- If day and month are missing for end date and year is non-missing, then day and month will be imputed as December 31st.

The completely missing AEs and CMs start and end dates will not be imputed.

The handling of missing efficacy data is discussed in Section 16.

10.2 Randomization Stratification

For the Induction Treatment Period, all participants are randomized to study treatment using permuted block randomization with block sizes of six, as specified in the Randomization

Specification Form (28 July 2021). This will be done via a IWRS. Under version 1.0 of the Protocol, the stratification factors were the following:

- biologic/JAK inhibitor ***failed*** status (yes/no)
- baseline corticosteroid use (yes/no)
- baseline disease activity (MMS 4-6 or 7-9)

Under subsequent Protocol versions the stratification factors are the following:

- biologic/JAK inhibitor prior use status (yes/no)
- baseline corticosteroid use (yes/no)
- baseline disease activity (MMS 5-6 or 7-9)

The randomization list was not updated to reflect these changes, and the IWRS system was kept as-is (Randomization Plan, version 2.0; 25 July 2022). For most statistical analysis purposes the following will be used:

- biologic/JAK inhibitor prior use status (yes/no)
- baseline corticosteroid use (yes/no)
- baseline disease activity (MMS 5-6 or 7-9) (i.e., participants with MMS of 4 at baseline according to IWRS are excluded from most efficacy analyses; however, see Sections 10.5 and 16.1.5 for several analyses that include them)

In all cases, randomized stratum will be used in the primary analysis of primary and key secondary efficacy analyses. A sensitivity analysis of the primary analysis of the primary endpoint will be generated using actual stratum as collected in EDC. Descriptive statistics based on actual stratum will be generated. See Section 16.1.7.

Data will be pooled across geographic region, country, and investigational site. No adjustment for these factors will be included in any statistical analysis.

10.3 Rectal Bleeding and Stool Frequency Subscores from eDiary

The eDiary approach described here is intended to match the Ulcerative Colitis Guidance Document (2022) issued by FDA, which recommends the following algorithm for computing the RB and SF subscores for the MMS outcome that is used in the Primary Endpoint:

- Use a 7-day period during which daily subscores are collected *immediately prior* to bowel preparation (or the day of endoscopy, if bowel preparation was not performed) for a study visit that includes endoscopy
- The RB and SF subscores should be calculated by averaging the daily subscores from within this 7-day period; see main text below Table 4 for a description of how this is done

- A minimum of 3 consecutive days of completed diary entries or 4 nonconsecutive days are necessary (otherwise the score should be considered missing and the participant's result imputed as non-responder)

Table 4 shows how the eDiary records should be selected for use in RB and SF calculation:

Table 4. Calculation of RB and SF Subscores from eDiary

Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Total Number of Days	Most Recent Consecutive Days	Implementation
Y	Y	Y	Y	Y	Y	Y	7	7	Use all 7
Y	Y	Y		Y	Y	Y	6	3	Use 6 available
Y	Y	Y	Y		Y	Y	4	2	Use 6 available
			Y		Y	Y	3	2	Cannot be used, RB and SF subscores not determined
	Y		Y		Y	Y	4	2	Use 4 non-consecutive

NOTE: Cells with 'Y' indicate that eDiary data is recorded

For study visits including an endoscopy (Visit 1, Visit 7, and possibly Unscheduled and/or Early Termination visits), a seven-day look-back period beginning the day immediately prior to bowel preparation will be used. For study visits not including an endoscopy, a seven-day look-back period beginning the day immediately prior to the actual study visit will be used when a study visit exists for a given analysis visit. If an analysis visit does not have a study visit, a seven-day look-back period beginning the day immediately prior to the Ideal Study Day will be used.

The calculation of SF for a given study participant is as follows:

a. eDiary days are selected according to Table 4.
b. For each eDiary day, the difference between the daily SF value and historical 'normal' is computed
c. Map each daily SF difference to an appropriate SF subscore as documented in FDA Guidance
d. Compute the numerical average SF subscore
e. Round to the nearest integer

RB is computed by taking the numerical average and rounding to the nearest integer; RB does not require comparison to a historical 'normal'.

10.4 Multiple Comparisons and Family-wise Type I Error Rate

All formal, controlled, pre-planned statistical analyses in this Protocol (primary endpoint and key secondary endpoints) are associated with Induction Week 13 – the end of the Induction Treatment Period.

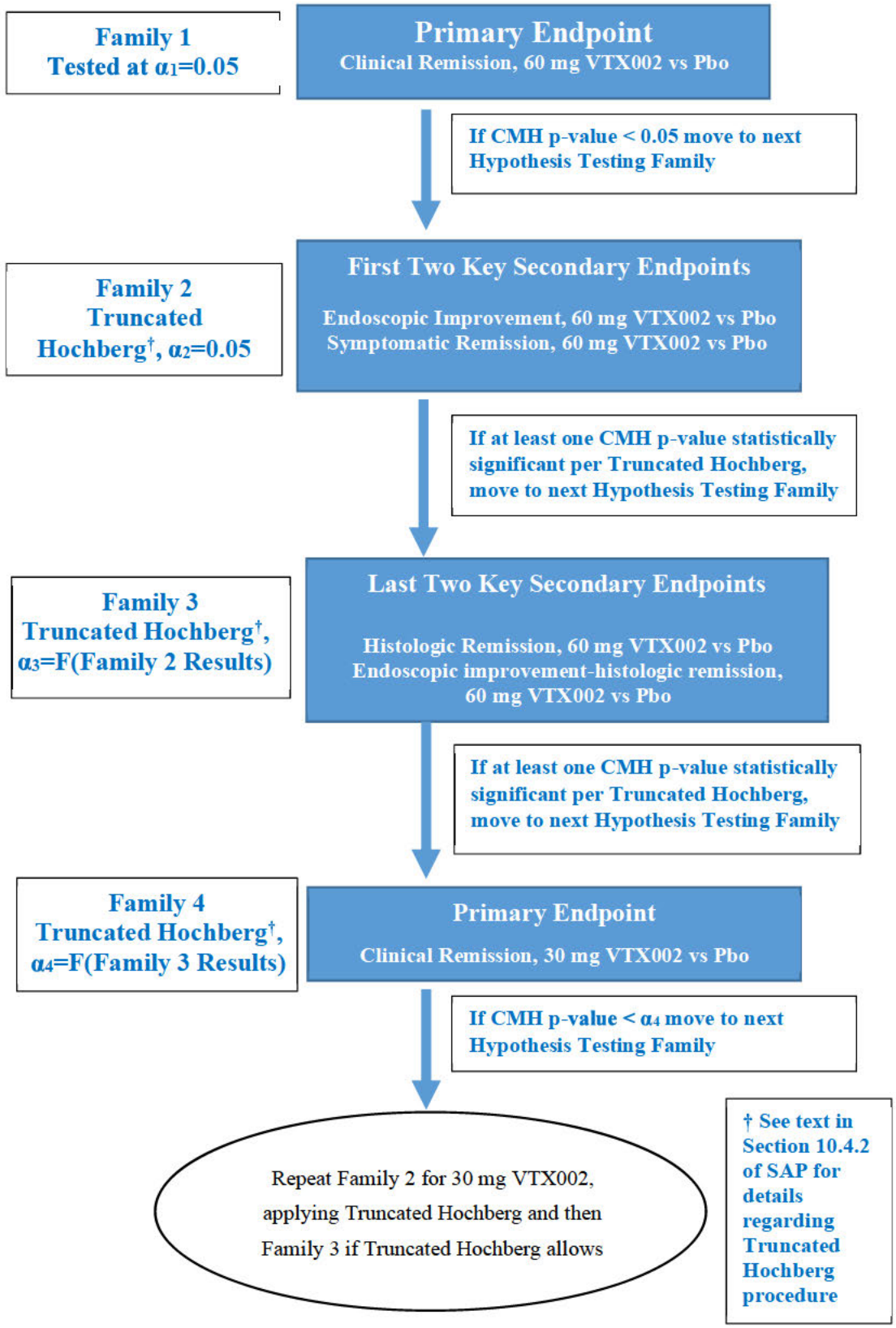
Preservation of an overall 5% Type I error rate for the primary analysis of the primary endpoint and key secondary endpoints will rely on the primary analysis of the primary efficacy endpoint as a gatekeeper, followed by sequential application of a truncated Hochberg (Dmitrienko 2008; Dmitrienko 2011a; Dmitrienko 2011b) procedure (i.e., sequentially rejective method) to families of hypothesis tests. Figure 3 illustrates the procedure.

The primary analysis of the primary efficacy endpoint will consist of an assessment of VTX002 60 mg versus Placebo for clinical remission at Induction Week 13. The null hypothesis associated with this analysis is denoted Family 1 and is a gatekeeper for the collection of Key Secondary Analyses comprised of hypothesis test families 2 through 6.

If this results in rejection of the null hypothesis at $\alpha = 0.05$, analysis proceeds to Family 2, consisting of null hypotheses for endoscopic improvement and symptomatic remission for VTX002 60 mg versus Placebo. If at least one of the hypotheses in this family is rejected in accordance with a truncated Hochberg procedure, some or all of the Type 1 error rate available for Family 2 is passed to Family 3 for additional hypothesis testing. If at least one of the hypotheses in Family 3 are rejected in accordance with a truncated Hochberg procedure, some or all of the Type 1 error rate available for Family 3 is passed to Family 4 – a gatekeeper hypothesis test assessing VTX002 30 mg versus Placebo with respect to the primary endpoint. This process continues until all hypothesis testing families (1 through 6) have been assessed or testing stops due to failure.

All other hypothesis tests (sensitivity analyses, supportive analyses, exploratory hypothesis tests) described in this SAP will be performed in an uncontrolled manner with nominal $\alpha = 0.05$, unless otherwise specified.

Figure 3. Sequentially Rejective Method: Truncated Hochberg Procedure



10.4.1 Hochberg Procedure

Consider testing the family of hypotheses H_{0i}; i = 1… k. Let p_i, for i = 1… k, denote the sample *p*-values of tests for H_{0i}; i = 1…k, computed without multiplicity adjustment. Let [1] … [k] denote the random indices such that $p[1] \leq \ldots \leq p[k]$, such that *p*[1] is the smallest *p*-value and *p*[k] is the largest *p*-value. The Hochberg procedure is implemented as follows:

Step 1: if $p[k] < \alpha$, reject $H_0[i]$, i=1, …, k and stop; otherwise proceed to next step.

Step 2: if $p[k-1] < \alpha/2$, reject $H_0[i]$, i=1, …, (k-1), and stop; otherwise proceed to next step.

…

Step k: if $p[1] < \alpha/k$, reject $H_0[1]$, and stop; otherwise no null hypothesis is rejected.

10.4.2 Truncated Hochberg Procedure

Consider testing the family of hypotheses H_{0i}; i = 1… k. Let p_i, for i = 1… k, denote the sample *p*-values of tests for H_{0i}; i = 1…k, computed without multiplicity adjustment. Let [1] … [k] denote the random indices such that $p[1] \leq \ldots \leq p[k]$, such that *p*[1] is the smallest *p*-value and *p*[k] is the largest *p*-value. The truncated Hochberg procedure employs a linear combination of a Bonferroni-adjusted significance level (α/k for all steps) with traditional Hochberg-adjusted significance levels (α at step 1, α/2 at step 2, … , α/k at step k) at each step via a truncation fraction, $f \in [0, 1]$, and is implemented as follows:

Step 1: if $p[k] < f\alpha+(1-f)*\alpha/k$, reject $H_{0[i]}$, i = 1, …, k and stop; otherwise proceed to next step.

Step 2: if $p[k-1] < f\alpha/2+(1-f)*\alpha/k$, reject $H_{0[i]}$, i = 1, …, k-1, and stop; otherwise proceed to next step.

…

Step k: if $p[1] < f\alpha/k+(1-f)*\alpha/k = \alpha/k$, reject $H_{0[1]}$, and stop; otherwise no null hypothesis is rejected.

A truncation fraction of f=1 yields a traditional Hochberg procedure, whereas a truncation fraction of f=0 yields a traditional Bonferroni procedure.

A truncation fraction of 0.7 will be applied in this study. Its application is demonstrated for Family 2 in Figure 3:

- At Step 1, the cutoff for the largest p-value is 0.7*(0.05) + 0.3*(0.025) = 0.035 + 0.0075 = 0.0425.
    - Suppose the largest p-value is 0.04. Then both hypotheses are rejected and α=0.05 is preserved and is passed in its entirety to Family 3 for use in hypothesis testing.

  - Suppose the largest p-value is 0.045. Then the null hypothesis associated with this p-value is not rejected and we proceed to Step 2

- At Step 2, the cutoff for the next largest p-value is 0.7*(0.025) + 0.3*(0.025) = 0.025
  - Suppose this p-value is 0.01. Then the hypothesis associated with this p-value is rejected and $\alpha_3 = \alpha_2-[f\alpha_2+(1-f)*\alpha_2/2] = 0.05 – [0.035 + 0.0075] = 0.0075$ is passed to Family 3 for use in further hypothesis testing.
  - Suppose this p-value is 0.03. Then the hypothesis associated with this p-value cannot be rejected and no Type 1 error rate may be passed to Family 3 (or any other hypothesis families) for additional controlled testing.

10.5 Subgroup Analyses

The primary and key secondary endpoints will be assessed for prespecified subgroups for the Induction Treatment Period (Induction Week 13). These subgroup analyses will be uncontrolled, are considered supportive, and will be based on the FAS. The following subgroups will be assessed:

- Randomization Stratification Factors – for these assessments, the stratification factor being assessed will be removed from any formal statistical tests
  - Biologic/JAK inhibitor prior use status (yes or no)
    - Number of prior biologic or JAK inhibitor therapies (1 or >1)
    - Prior UC treatment failure of anti-tumor necrosis factor alpha (anti-TNFα) (yes or no)
  - Baseline corticosteroid use (yes or no)
  - Baseline disease activity
    - MMS 5-6 or MMS 7-9
    - MMS 4-6 or MMS 7-9
- Other Factors
  - Sex (Female or Male)
  - Age (18-64, 65-74, 75-80, 18-44, 45-65, >65, ≥65, ≥75)
  - Race (White or Non-White)
  - Region (North America, Western Europe, Eastern Europe, Other)
  - Baseline endoscopic subscore (2 or 3)
  - Baseline FCP (≤ or > 250mg/kg)
  - Baseline high-sensitivity C-reactive protein (CRP) (≤ or > 3mg/L)

- Baseline total Mayo score (TMS) (≤ or > 8)
- Duration of UC at Screening (≤ or > Median)
- Extent of disease at Screening (as collected in eCRFs)
  - Proctosigmoiditis
  - Pancolitis
  - Proctitis
- Prior UC treatment of oral 5-aminosalicylic acid (5-ASA) only (Yes or No)
- Prior UC treatment failure of oral 5-ASA only (Yes or No)

Additional subgroups may be assessed, if deemed necessary. The medians will be derived based on the FAS for all subgroups cut at the median. This study was not designed to detect treatment differences within subgroups. If a subgroup includes < 10% of all participants, no inferential statistics will be generated.

The actual stratum at Baseline (as collected in EDC) will be used for all subgroup analyses, including randomization stratification factors. Randomization stratification factors not being treated as a subgroup for a particular subgroup analysis will also be handled according to actual Baseline stratum as collected in EDC.

Forest plots of the results associated not only with the overall FAS, but also all considered subgroups, will be presented for the primary endpoint and each key secondary analysis. Inference from the analyses described in Sections 16.1.3 and Section 16.2.1 Table 6, columns 2 and 3 will be used. Differences in responder percentages along with 95% CIs will be displayed. Sensitivity and/or Supplementary analyses will not be considered.

11. DISPOSITION AND PROTOCOL DEVIATIONS

11.1 Participant Disposition

All study participants providing informed consent will be accounted for in this study. Treatment for the Induction Treatment Period and LTE will be represented as-randomized. Treatment for OLE will be represented as-received. Percentages for participants will be based on the number of participants randomized in the Induction Treatment Period.

Participants screened and reason for screen fail will be summarized by number and percentage. Participants meeting all eligibility criteria but not randomized will be included in this summary as well. Inclusion criteria not met and exclusion criteria met will be included in this tabulation and will also be listed in participant listings.

Among study participants, disposition will be tabulated so that each protocol study period is separately represented (Induction Treatment Period, LTE, OLE). LTE and OLE are available

only for participants initially randomized into the Induction Treatment Period. Participant disposition tabulation for these study periods will display participant flow through all three study periods. A CONSORT diagram will also be developed.

There are three study period-specific completion status assignments a participant may have:

- A participant flagged with a status of "Treatment completed" on the End of Induction Treatment Period form is considered to be an Induction completer
- A participant flagged with a status of "Treatment completed" on the End of Long Term Extension Period form is considered to be an LTE completer
- A participant flagged with a status of "Treatment completed" on the End of Open Label Extension Period form is considered to be an OLE completer

Within a given study period, the number and percent of participants who completed/discontinued treatment, primary reason for discontinuation of treatment, the number and percent of participants who completed/discontinued the study period, and primary reason off study period will be summarized by number and percentage. This summary will also be provided by region and country. The number and percent of participants in each analysis set will be summarized.

Death is a reason for study discontinuation and will also be summarized for a given study period by number and percentage. The summary of death will include all SAEs with a fatal outcome.

Listings will be provided for participants' study disposition, including reasons for exclusion and discontinuation from the study treatment, discontinuation from study periods and date/cause of death.

11.2 Protocol Deviations

After the last participant completes the Induction Treatment Period, the protocol deviation log for the Induction Treatment Period will be provided to Sponsor for review at the final Induction Treatment Period Blinded Data Review Meeting (BDRM). Sponsor will classify protocol deviations as either major or minor per the Protocol Deviation Plan (PDP).

Similarly, at the end of the study, the final project-level protocol deviation log will be provided to the sponsor prior to the final database lock. Upon the study end, the final project-level protocol deviation log will be approved by the sponsor at the final study BDRM and filed in trial master file (TMF).

For the Induction Treatment Period, major deviations will be separately summarized (number and percentage) by treatment group for the FAS according to the following categories, and all deviations will be listed:

- Withdrawal criteria
- Randomization procedure

- Subject visit completion or timing
- Study procedure or assessment
- Study medication
- (Serious) AE reporting
- Other

Note that participants experiencing an intercurrent event during the Induction Treatment Period and not experiencing other deviations preventing contribution to the primary and key secondary analyses of the primary and key secondary endpoints will be included in the PP set and treated as non-responders in those analyses.

A separate summary will be provided for protocol deviations occurring during the entire study. The FAS-ET set and planned treatment group will be used for this purpose.

12. DEMOGRAPHIC AND BASELINE CHARACTERISTICS

The following demographic variables will be summarized, as collected in EDC:

- Baseline Age (yr)
- Baseline Age Group (18-<65 ≥65)
- Baseline Height (cm)
- Baseline Weight (kg)
- BMI (kg/m^2), derived from Screening weight (kg) and height (m)
- Baseline eGFR (mL/min/1.73m^2), derived from serum creatinine using the 2021 update to the CKD-EPI formula
- Country
- Sex (M/F)
- Female of Childbearing potential (Y/N)
  If No:
    - Surgical Sterilization or Hysterectomy (Y/N)
    - Post-menopausal for more than 1 year (Y/N)
- Ethnicity
- Race

Separately, Baseline [1] UC signs and symptoms will be presented, where Baseline is understood to be the most recently available measurement prior to administration of study treatment. For these UC Baseline signs and symptoms, stool frequency and rectal bleeding data derived from eDiary will be presented, as opposed to what is available from EDC. Refer to Section 10.3 for a description of how eDiary data is processed to determine the stool frequency and rectal bleeding subscores. Derived stool frequency and rectal bleeding data will also be used for efficacy endpoint evaluation, and if these values are present in EDC but not calculable from eDiary, they will be treated as missing. The following UC Baseline signs and symptoms will be summarized descriptively:

- Duration of Ulcerative Colitis (yr): (Informed consent date – Date of diagnosis + 1) / 365.25
- Location and extent of Ulcerative Colitis (Proctosigmoiditis, Pancolitis, Proctitus, Other)
- Any surgeries related to Ulcerative Colitis? (Y/N)
  If Yes:
    - Number of surgeries
- Colonoscopy within past 12 months? (Y/N)
- Endoscopic Score (ES)
- Stool Frequency (SF)
- Rectal Bleeding (RB)
- Physician's Global Assessment (PGA)
- Mayo Clinic Score (MCS)
- Modified Mayo Score (MMS)
- MMS Group (5-6, 7-9)
    - As collected in IWRS
    - As collected in EDC
- Partial Mayo Score (PMS)
- Stable corticosteroids (Y/N)
    - As collected in IWRS
    - As collected in EDC

1 Baseline refers to the most recently available measurement(s) prior to administration of study treatment. For some UC signs and symptoms, such as Endoscopic Score, this is strictly a Screening data point. For others such as Physician's Global Assessment (PGA) and MCS (which uses PGA as a component), these are Study Day 1 pre-dosing values.

- Prior biologic/JAK (Y/N)
    - As collected in IWRS
    - As collected in EDC
    - Prior failure of anti-TNF (Y/N) as collected in EDC
    - Prior failure of vedolizumab (Y/N) as collected in EDC
    - Prior failure of both anti-TNF and vedolizumab (Y/N) as collected in EDC
    - Prior failure of JAK inhibitor (Y/N) as collected in EDC
- Prior failure of oral 5-ASA only (Y/N) as collected in EDC

For the Induction Period Analysis, summaries will be presented by planned treatment group and overall for the FAS, mFAS, and PP populations. At study end, summaries will be presented by planned treatment group and overall for FAS-ET population. Listings will be provided for all demographic and baseline variables, and derived SF and RB values will be presented along with SF and RB as collected in EDC, and in this listing differences between EDC and programmatically-derived SF and RB scores will be flagged.

13. MEDICAL HISTORY

The following variables will be summarized:

- Any past or concomitant diseases? (Y/N)
- Any past surgeries for reasons other than Ulcerative Colitis? (Y/N)

Additionally, medical history and concomitant medical conditions will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) and summarized by System Organ Class (SOC) and Preferred Term (PT) within each of these categories (historical vs concurrent). A concomitant condition is one having ongoing status = Yes, as collected in the Medical History eCRF. Adverse events occurring prior to randomization and on or after signing ICF will be captured as Medical History.

For the Induction Period Analysis, these summaries will be presented by planned treatment group and overall for the FAS, mFAS, and PP Set.

Listings will be provided for all medical history findings.

14. PRIOR AND CONCOMITANT MEDICATIONS

Medications will be coded using the WHO Drug Dictionary, utilizing Anatomical Therapeutic Chemical (ATC) Therapeutic Main Group (Level 2) and Chemical Subgroup (Level 4). Listings will be provided and will additionally include ATC Level 5, separately, for all prior and concomitant medications.

14.1 UC MEDICATIONS

Medications taken for UC will be considered prior UC medications if the start date of the medication is prior to the Informed Consent date. Otherwise, they are considered concomitant UC medications.

For the Induction Period Analysis, summaries of prior and concomitant UC medications will be separately presented by planned treatment group and overall for the Safety, FAS, mFAS, and PP sets. At study end, summaries will be presented by planned treatment group for FAS-ET population. Medications flagged as 'rescue medications' (see Section 25, Appendix IV) will be considered concomitant UC medications and will be flagged in listings.

14.2 OTHER MEDICATIONS

Medications not taken for Ulcerative Colitis will be considered prior medications if the end date of the medication is prior to the date of first dose of study treatment. Otherwise, they are considered concomitant medications.

A medication is considered concomitant for the Induction Treatment Period if it meets the following conditions:

a) Medication started or was ongoing on or after the first study treatment administration date

and

b) Medication started or was ongoing on or before Visit 7
or
Medication started or was ongoing on or before any of the Early Termination/1-Week Follow-up/2-Week Follow-up visits for participants that permanently discontinue study treatment during the Induction Treatment Period (and are therefore discontinued from the study).

A medication is considered concomitant for the entire study if meets the following conditions:

c) Medication started or was ongoing on or after the first study treatment administration date

and

d) Medication study treatment was ongoing on or before the two-week follow-up visit.

Imputation rules specified in Section 10.1 will be applied to partial/missing dates.

Three summaries will be created:

- Prior medications will be summarized by planned treatment group and overall for the FAS, mFAS, and PP sets.
- Concomitant medications taken during the Induction Treatment Period will be summarized by planned treatment group and overall for the Safety, FAS mFAS, and PP sets. This summary will be generated for the Induction Period Analysis.

- Concomitant medications taken during the entire study summaries will be summarized by planned treatment group and overall for FAS ET and PP populations. This summary will be generated at study end.

Additionally, the use of P-gp or BCRP transporters during the study will be flagged in participant-level dataset(s) and in listings. Please see Section 25 (Appendix IV) for WHO Drug Dictionary (WHODD) ATC codes specifying P-gp and BCRP transporters.

If prior medications for UC are inadvertently entered on eCRF pages intended for concomitant medications, these records will be treated as prior medications.

15. EXPOSURE AND TREATMENT COMPLIANCE

The Safety Set will be used for all exposure and compliance tabulations and listings.

Titration blister packs contain a varying and nondecreasing number of tablets for each study day of titration. Maintenance blister packs contain a fixed number of tablets for each study day of maintenance. Bottles contain a fixed number of tablets intended to provide the study participant with three tablets per day between site visits, plus overage to account for visit scheduling.

Blister packs are dispensed to study participants for the Induction Treatment Period and LTE – no bottles are dispensed. During OLE, participants are dispensed blister packs for titration purposes only; after titration, participants are dispensed study treatment in bottles.

Participant eDiary data will not be used for any purpose related to exposure or treatment compliance. Exposure and compliance tables and listings will rely exclusively on data as collected in EDC and IWRS.

15.1 Treatment Exposure

Treatment exposure (duration in days, number of tables expected, number of tablets taken, number of tablets missed, compliance, dose interruptions, and overdoses) will be summarized descriptively. For the interim database lock and unblinding associated with the Induction Treatment Period, treatment exposure will be tabulated for the Induction Treatment Period only. Participant listings will be generated.

15.1.1 Duration

Per-participant Treatment Duration will be calculated as (Participant Date of Last Study treatment – Participant Date of First Study treatment) + 1, using appropriate first and last dates, and expressed in days. Interruptions, compliance, and dose changes are not taken into account for duration of exposure.

For specific study periods (such as Induction, LTE, or OLE), the per-participant first and last dates will be the first and last treatment dates within the given study period. The per-participant

first and last treatment dates will be the least recent and most recent treatment dates across all applicable study periods for that participant.

- Induction
    - Date of first study treatment will be taken from the Study treatment Administration on Site form for the Visit 2 Week 0 site visit.
    - Date of final study treatment will be taken from the End of Induction Treatment Period form.
- LTE
    - Date of first study treatment will be taken from the "Start Date of Administration" field from the "Study Drug Compliance – LTE" form.
    - Date of final study treatment will be taken from the End of Long Term Extension Period form.
- OLE
    - Date of first study treatment will be taken from the "Start Date of Administration" field from the "Study Drug Compliance – OLE" form.
    - Date of final study treatment will be taken from the End of Open Label Extension Period form

15.1.2 Tablets Taken, Missed Doses, and Dose Interruptions

The Study treatment Titration (and Study treatment Re-Titration for participants needing it), Study treatment Compliance, Study treatment Overdose, and Study treatment Administration on Site forms will be used to capture drug exposure. Based on these data for the Induction Treatment Period, the following will be calculated:

- Number of tablets taken = total dispensed – total returned
- Number of tablets expected to be taken = number of tablets a participant should have taken between their first and last study treatment administration
- Number of incomplete doses = number of instances where the number of tablets taken in a study day is less than the number tablets the participant should have taken for that study day
- Number of missed doses = number of instances where the participant takes no tablets for a given study day

These will be summarized for the Induction Treatment Period and will also be listed.

The frequency and percentage of participants who had at least 1 dose interruption, who had at least 1 dose interruption of > 7 days, who had at least 1 dose interruption of > 14 days, who had 1

overdose, and who had > 1 overdose will be summarized for the Induction Treatment Period. Interruption is defined as at least one day of missed dose.

15.2 Compliance

For the Induction Period Analysis, treatment compliance is calculated as (Number of Tablets Taken*100)/(Number of Tablets Expected to be Taken). A participant is non-compliant with study treatment if their treatment compliance is less than 80% or more than 120%. For all bottles not returned, it will be assumed that all dispensed tablets were taken.

This will be summarized by treatment and overall for the Induction Treatment Period. Participant listings will be generated.

16. EFFICACY

Controlled efficacy analyses are limited to the Induction Treatment Period. Efficacy endpoints may be computed for LTE and OLE study visits, but all hypothesis testing will be uncontrolled, resulting in nominal p-values and CIs only.

The study is designed to show superiority of VTX002 60 mg to placebo for the primary efficacy endpoint at Induction Week 13. Four key secondary endpoints are defined for this study and VTX002 30 mg is an important additional active drug dose of interest. Across the primary endpoint and key secondary endpoints, a total of ten (10) Induction Week 13 controlled hypothesis tests involving the two randomized VTX002 doses and placebo are planned. The family-wise error rate will be maintained at $\alpha = 0.05$ – see Section 10.3 for details.

All controlled efficacy analyses (Week 13, Induction) will be primarily based on participant stratum at randomization (see Section 10.2) according to IWRS and for the FAS. Participants having MMS = 4 at baseline will be excluded from all efficacy analyses except the sensitivity analysis described in Section 16.1.5. Additional sensitivity and/or supportive analyses will be performed as described elsewhere in this section. For composite scores such as MMS, if one or more components is missing from an analysis visit window the composite score is missing.

Blinded study team review will be used prior to Induction interim database lock and unblinding, to attribute a known Induction Week 13 non-response to participants having specific on-study experiences. Participants will be classified as non-responders in all responder-based efficacy analyses, including the primary endpoint, at all timepoints subsequent to the experience of the earliest of any of the following intercurrent events:

- Missed Visit 7 Week 13 endoscopy for participants eligible to have Visit 7 Week 13 endoscopy
- Discontinuation of the study during the Induction Treatment Period for either disease worsening or treatment-emergent adverse event having a preferred term indicating ulcerative colitis (UC)

- Exposure to rescue therapy for UC. See Appendix IV, Section 25.

For a given participant, if efficacy outcomes (as opposed to components of outcomes) are available subsequent to an intercurrent event, they will be flagged as such. Participants may be included in the PP Set as long as other Per Protocol criteria are not violated.

Analysis visits will be mapped according to Section 9.6 Table 3 before any missing data imputation is applied.

16.1 Primary Efficacy

16.1.1 Primary Efficacy Variable and Derivation

Primary efficacy analyses will compare VTX002 60 mg to placebo at Week 13 of the Induction Treatment Period. This will be done for the proportion of participants achieving clinical remission as defined in Section 7.5. Participants who achieve clinical remission will be referred to as responders. Participants who do not achieve clinical remission will be referred to as non-responders.

16.1.2 Missing Data Methods for the Primary Efficacy Variable

The concept of non-response for participants with non-missing Week 13 Induction primary efficacy outcomes is described at the beginning of Section 16. For participants with missing Week 13 Induction Treatment Period primary efficacy outcomes, the following missing data methods will be used:

- Primary method: single imputation as non-responder
- Sensitivity analyses
    - multiple imputation under missing at random (MAR)
    - multiple imputation with Copy Reference (CR) under missing not at random (MNAR)
    - tipping point analysis
- Supplementary analyses: mFAS and PP set analysis with data as observed

These methods are discussed in more detail below. See Section 29 (Appendix VIII) for a tabular presentation of the relationship between intercurrent events, missing data, and response imputation.

16.1.3 Primary Analysis of Primary Efficacy Endpoint

The primary analysis of the primary efficacy endpoint will be completed based on the FAS. The primary endpoint is the proportion of participants with clinical remission at Induction Week 13. The following are the null and alternate hypotheses as stated in the Protocol:

H_0: p(VTX002 60 mg) – p(Placebo) = 0

H_A: p(VTX002 60 mg) – p(Placebo) ≠ 0.

For statistical analysis purposes, the following statement will be considered logically equivalent:

H_0: p(VTX002 60 mg) = p(Placebo)

H_A: p(VTX002 60 mg) ≠ p(Placebo)

The null hypothesis will be tested at a 5% level of significance. Testing will be done using a Cochran-Mantel-Haenszel (CMH) test with biologic/JAK inhibitor prior use status ('Yes' vs 'No'), baseline corticosteroid use ('Yes' vs 'No'), and baseline disease activity ('Low' vs 'High') as the stratification factors. The stratification factors will be treated as-randomized regardless of actual participant stratification. The CMH chi-square *p*-value, stratified risk difference ("common" risk difference) with 95% CI using the Newcombe method, Mantel-Haenszel odds ratio, and 95% CI for the Mantel-Haenszel odds ratio will be reported for pairwise comparisons between VTX002 60 mg and placebo, along with number and percentage of participants in clinical remission in each as-randomized treatment group.

The following SAS pseudo-code will be used for computation of Cochrane-Mantel-Haenszel 'general association' p-values (based on table scores), exact odds ratios and confidence intervals, and Newcombe common risk differences in the presence of randomization stratification factors:

[redacted]

Additionally, unstratified common risk differences will be generated and presented along with the Mantel-Fleiss criterion (Mantel 1980):

[redacted]

In all cases, statistical analyses accounting for randomization stratification factors will be considered primary.

Additional adhoc analysis methods may be employed subsequent to the Induction Period Analysis database lock and unblinding if the presence of zeroes in 2x2 responder frequency tables (one for each stratum, see Section 10.2) creates difficulties for implementation of the pre-specified analysis approach.

16.1.4 Missing Data Sensitivity Analyses of Primary Efficacy Endpoint

Three sensitivity analyses will be implemented to explore different types of missing data approaches, as defined in Section 16.1.2. For purposes of multiple imputation, SAS default common risk differences (as opposed to Newcombe) will be generated in order to capture risk difference standard errors.

16.1.4.1 Multiple Imputation Under MAR

Any missing component scores of MMS at the planned assessments will be imputed using multiple imputation under MAR. MAR assumes the missing value is independent of unobserved outcomes given observed data (i.e., participants with missing RB subscores can be modeled based on participants with observed RB subscores) (Rubin 1987).

Fully Conditional Specification (FCS) with predictive mean matching via PROC MI will be used in a "Just Another Variable" (JAV) approach in which all longitudinal variables are put into a "wide format" dataset (Bergland 2015, Bergland 2018).

Variables with potentially missing values at post-baseline visits include SF, RB, ES, Geboes index score, absolute lymphocyte count (ALC), CRP, and PGA. These variables are continuous, and after multiple imputation per-participant responder values ('Y' or 'N') will be derived for the primary endpoint for every multiply-imputed dataset.

Variables with no missing values allowed are the randomization stratification factors, prior UC treatment failure of oral 5-ASA only (participant had inadequate response, loss of response, or intolerance to previous treatment with oral 5-ASA and did not have an inadequate response, loss of response, or intolerance to any other previous UC medication), and all of the previously mentioned continuous variables at baseline. If a participant has missing baseline values in any of these variables, the participant will be excluded from multiple imputation.

The following steps will be implemented:

Step 1:

Regardless of the arbitrary missing data pattern (i.e., non-monotone or monotone), a FCS method with predictive mean matching for continuous variables will be used to impute the missing data at all timepoints. The FCS method allows for separate conditional distributions for each imputed variable. The predictive mean matching approach creates a regression model using parameters sampled from the posterior distribution and then a predicted value for each missing value is computed. The missing value is replaced by randomly selecting an observation from a set of 'k' values that are the closest predicted values to the missing predicted value. Missing data imputation will be performed using the SAS PROC MI procedure.

Natural logarithm will be applied to ALC and CRP prior to multiple imputation. The number of imputations will be 40 and is based on a projected overall dropout rate of 10% and relative efficiency of > 99%. A separate imputation model will be used for each treatment group.

Refer to Section 26 (Appendix V) for the order of variables and timepoints in the model as well as sample PROC MI code.

Step 2:

For each of the 40 imputed datasets at Induction Week 13, participants will be classified as responder or non-responder based on the criteria specified in Section 16. For each dataset, the proportion of responders will be calculated. For endpoints requiring MMS, MMS will be recomputed by summing the scores of ES, SF, and RB.

Step 3:

The PROC FREQ code given in Section 16.1.3 will be applied to the 40 multiply-imputed datasets, resulting in 40 sets of CMH chi-square p-values, Newcombe risk differences, and CMH odds ratios.

Step 4:

PROC MIANALYZE requires sample statistics to have approximately normal distributions. The following table describes transformations applied to the three output statistics prior to handling with PROC MIANALYZE (Ratitch 2013):

Table 5. Transformations Prior to Using PROC MIANALYZE

Statistic	Transformed Value	Transformed Standard Error or Standard Deviation
Risk Difference (p_1-p_2)	*Identity*	*Identity*
Odds Ratio ($Odds_1/Odds_2$) †	$\ln(Odds_1) - \ln(Odds_2)$	[ln(95% UCL) - ln(95% LCL)] / (2*1.96)
CMH Test Statistic ("VALUE") ††	[(VALUE/df)^(1/3) - (1-2/(9*df))] / ((2/(9*df))^(1/2))	1.0

† After application of PROC MIANALYZE, the results are back-transformed to the original scale
†† This Wilson-Hilferty transformation yields test statistics having an approximate normal distribution with standard deviation of 1.0. The multiply-imputed CMH p-value is obtained by computing one-sided t-statistic p-values after application of PROC MIANALYZE.

Refer to Section 26 (Appendix V) for variable ordering, PROC MI code, sample SAS transformation code, PROC MIANALYZE code, and post-processing code.

It is noted there could be certain adjustments due to unexpected data issues after unblinding treatment. All post-unblinding modifications to the multiple imputation model or approaches to address missing data will be described in the CSR.

16.1.4.2 Multiple Imputation with Copy Reference

Missing data will also be explored assuming an MNAR approach. MNAR assumes the missing values depend on unobserved outcomes even after accounting for the observed data. Therefore, participants cannot be modelled based on participants with observed data, and more assumptions are needed. Copy reference is a type of MNAR approach where each missing value for participants in the treatment group will be imputed using observed data in the placebo group, and missing values for participants in the placebo group will be imputed under the MAR approach.

The following steps will be implemented:

Step 1:

Step 1 (a):

Intermittent missing data (before discontinuation from study) at each scheduled visit will be imputed, separately for each treatment group, under the MAR assumption using a multivariate normal imputation model and the Markov Chain Monte Carlo (MCMC) method with multiple chains. The number of imputations will be 40. The resulting data will have a monotone missing data pattern. The imputed data will be used together with observed data to impute post-discontinuation missing values.

Step 1 (b):

Only after Step 1 (a) is completed, the remaining monotone missing data after discontinuation will be imputed using the FCS predictive mean matching method based on data from the placebo group only. The RB, SF, ES, and PGA subscores, stratification variables, a binary variable indicating if a participant is a treatment failure, and biomarker measurements (ALC,R CRP, log-transformed) at each planned visit will be included in the imputation model.

Step 2:

Same as Step 2 described in Section 16.1.4.1.

Step 3:

Same as Step 3 described in Section 16.1.4.1.

Step 4:

Same as Step 4 described in Section 16.1.4.1.

16.1.4.3 Tipping Point Analysis

A tipping point approach based on multiple imputation of Visit 7 Week 13 responses with respect to the primary endpoint will be used [2]. This will be an MNAR-based imputation approach in

[2] https://www.pharmasug.org/proceedings/2023/SD/PharmaSUG-2023-SD-069.pdf

which direct Bernoulli simulation is used to evaluate the robustness of the obtained primary efficacy results with respect to various assumed response rates in each treatment arm.

The following steps will be implemented:

Step 1:

Participants not contributing observed RB, SF, and ES at Visit 7 Week 13 will be assumed to have primary efficacy response rates that vary according to the following thirty-element lists:

VTX002 60 mg: p_{VTX} ε {0%, 2%, 4%,…, 30%, 35%, 40%,…, 95%, 100%}

Placebo: p_{PBO} ε {0%, 2%, 4%,…, 30%, 35%, 40%,…, 95%, 100%}

Step 2:

a. Consider all possible combinations of primary efficacy response rates. This yields 900 pairs of response rates.

b. Fix attention to a specific pair of response rates. An example of such a pair would be (p_{VTX} = 14%, p_{PBO} = 35%).

   i. Imputation #1: Assign responder or non-responder status to each participant not contributing observed RB, SF, and ES at Visit 7 Week 13. This will be done, per participant, by drawing N=1 random draws from a Binomial distribution having an appropriate probability from the response rate pair. An example of such a random draw would be generating $X_i \sim$ Bin(n=1, p_{PBO} = 35%) for each i, where i indexes all placebo participants not contributing observed RB, SF, and ES at Visit 7 Week 13.

   ii. Repeat until 40 total imputations have been generated

c. Repeat 2b until all possible pairs of response rates have been handled.

Step 3:

Same as Step 3 described in Section 16.1.4.1.

Step 4:

Same as Step 4 described in Section 16.1.4.1.

Results will be plotted as a heat map on a two-dimensional plot with axes corresponding to the assumed primary efficacy response rates applied to each treatment group respectively, and different colors will be used to represent the magnitude of the 2-sided *p*-values corresponding to analysis with each combination of delta values. The clinical interpretation about the plausibility of the assumptions underlying the tipping point will be provided in the CSR.

16.1.5 Baseline Disease Activity Sensitivity Analysis of Primary Efficacy Variable

A total of four (4) participants randomized have baseline disease activity MMS = 4 according to IWRS. These four participants will be included in a sensitivity analysis of the Primary Efficacy

Variable. No imputation will be performed for this sensitivity analyses, non-responders will be handled as defined at the beginning of Section 16, FAS will be used, and this analysis will use the same CMH statistical methodology with randomized stratification factors as other responder analyses. VTX002 60 mg vs Placebo will be the only treatment comparison considered.

16.1.6 Supplementary Analysis of Primary Efficacy Variable

The primary model as described in Section 16.1.3 will be repeated using the mFAS and Per Protocol sets as separate analyses.

16.1.7 Stratification Factor Sensitivity Analysis of Primary Efficacy Variable

FAS will also be used in re-analyses of the primary model but using stratification as-collected in EDC instead of as-randomized in IWRS. No imputation beyond non-responder imputation as described at the top of Section 16 will be done.

16.2 Key Secondary Analyses

Key secondary efficacy endpoints are defined in Section 7.2:

- Hypothesis testing of these endpoints for VTX002 60 mg versus placebo are considered key secondary analyses
- Hypothesis testing of the primary efficacy endpoint for VTX002 30 mg versus placebo, in addition to hypothesis testing of the key secondary efficacy endpoints for VTX002 30 mg versus placebo, are also considered key secondary analyses

The primary analyses of these hypothesis tests are subject to the family-wise Type 1 error rate handling and gatekeeping (i.e., multiplicity adjustments) described in Section 10.3. After attribution of known outcome at Induction Week 13 (see introduction to Section 16), the primary planned Key Secondary Analyses as described Table 6 will be conducted.

The CMH methodology used for the primary analysis of the primary endpoint will be used for all Key Secondary Analyses. The additional primary endpoint analysis methods described in Sections 16.1.4 and 16.1.6 – missing data methods (MAR multiple imputation, copy reference multiple imputation, tipping point), randomization stratification factor (EDC) sensitivity analysis, and supplementary analyses (application of mFAS and PP analysis sets to primary analysis methods) – will also be applied to the Key Secondary endpoints. See Table 6.

For tipping point analyses, identification of participants not contributing observed data at Visit 7 Week 13 will rely on the types of data needed for a given Key Secondary endpoint. For example, endoscopic improvement-histologic remission (the fourth Key Secondary endpoint) requires Geboes and ES to both be available. If they are not available for a given participant, that participant will be subject to tipping point analysis. The set of participants subject to tipping point analysis may differ from one endpoint to another.

16.2.1 Key Secondary Efficacy Endpoints and Derivations

16.2.1.1 Endoscopic Improvement at Week 13

The null hypothesis is that the proportion of participants achieving endoscopic improvement at Induction Week 13 is the same between VTX002 and placebo. The alternative hypothesis is that the proportion of participants achieving endoscopic improvement at Induction Week 13 is different between VTX002 and placebo. Subject to successful rejection of the primary efficacy endpoint for VTX002 60 mg vs placebo, this key secondary endpoint is evaluated according to the sequential Hypothesis Family approach illustrated in Section 10.3, Figure 3.

16.2.1.2 Symptomatic Remission at Week 13

The null hypothesis is that the proportion of participants achieving symptomatic remission at Induction Week 13 is the same between VTX002 and placebo. The alternative hypothesis is that the proportion of participants achieving symptomatic remission at Induction Week 13 is different between VTX002 and placebo. Subject to successful rejection of the primary efficacy endpoint for VTX002 60 mg vs placebo, this key secondary endpoint is evaluated according to the sequential Hypothesis Family approach illustrated in Section 10.3 Figure 3.

16.2.1.3 Histologic Remission at Week 13

The null hypothesis is that the proportion of participants achieving histologic remission at Induction Week 13 is the same between VTX002 and placebo. The alternative hypothesis is that the proportion of participants achieving histologic remission at Induction Week 13 is different between VTX002 and placebo. Subject to successful rejection of the primary efficacy endpoint for VTX002 60 mg vs placebo, this key secondary endpoint is evaluated according to the sequential Hypothesis Family approach illustrated in Section 10.3 Figure 3.

16.2.1.4 Endoscopic Improvement-Histologic Remission at Week 13

The null hypothesis is that the proportion of participants achieving endoscopic improvement-histologic remission at Induction Week 13 is the same between VTX002 and placebo. The alternative hypothesis is that the proportion of participants achieving endoscopic improvement-histologic remission at Induction Week 13 is different between VTX002 and placebo. Subject to successful rejection of the primary efficacy endpoint for VTX002 60 mg vs placebo, this key secondary endpoint is evaluated according to the sequential Hypothesis Family approach illustrated in Section 10.3, Figure 3.

Table 6. Planned Primary and Key Secondary Analyses
X = VTX002 60 mg vs Placebo
Y = VTX002 30 mg vs Placebo

Endpoint	Primary Analysis with Stratified CMH (FAS)	Key Secondary Analyses with Stratified CMH [b] (FAS)	Section 16.1.4 Missing Data Sensitivity Analyses (FAS: MAR MI, CR MI, Tipping Point)	Section 16.1.5 Sensitivity Analysis (FAS)	Section 16.1.6 Supplementary Analyses (mFAS, PP Sets)	Section 16.1.7 Sensitivity Analyses (FAS)
Primary endpoint – Clinical Remission	X [a] Family 1	Y Family 4	X, Y	X	X, Y	X, Y
Key Secondary Endpoints						
Endoscopic Improvement		X: Family 2 Y: Family 5	X, Y		X, Y	X, Y
Symptomatic Remission		X: Family 2 Y: Family 5	X, Y		X, Y	X, Y
Histologic Remission		X: Family 3 Y: Family 6	X, Y		X, Y	X, Y
Endoscopic improvement – Histologic Remission		X: Family 3 Y: Family 6	X, Y		X, Y	X, Y

[a] Serves as primary analysis gatekeeper for testing all Key Secondary Analyses with Stratified CMH
[b] The collection of Key Secondary Analyses with Stratified CMH subject to a truncated Hochberg multiplicity adjustment to control the family-wise Type 1 error rate at α=0.05. Figure 3, Section 10.3 presents a flow chart demonstrating how this is to be done.

CMH = Cochrane-Mantel-Haenszel; CR = copy reference; FAS = Full Analysis Set; mFAS = modified Full Analysis Set; PP = Per Protocol set; MI = multiple imputation; MAR = missing at random

16.3 Other Secondary Efficacy

Plasma concentrations of VTX002 will be handled in a separate PK Analysis Plan. Other secondary endpoints addressed in this SAP include the following (see Section 7.2):

- The proportion of participants with clinical response using MMS at Week 13
- The proportion of participants with endoscopic remission at Week 13
- The proportion of participants with endoscopic and histologic remission at Week 13
- The proportion of participants with endoscopic and clinical remission at Week 13
- The proportion of participants with endoscopic, histologic, and clinical remission at Week 13
- The proportion of participants with symptomatic remission at Weeks 4, 8, 10
- The proportion of participants with symptomatic response at Weeks 4, 8, 10, 13
- The proportion of participants with clinical remission using total MCS at Week 13
- The proportion of participants with clinical response using total MCS at Week 13
- The proportion of participants with histologic improvement at Week 13
- The proportion of participants with HEMI at Week 13
- The proportion of participants with any decrease from baseline in Geboes Index score at Week 13
- The proportion of participants with UC-related hospitalizations
- The proportion of participants requiring UC-related surgeries, including colectomy

These endpoints will be analyzed similarly to the primary endpoint, per timepoint, using the CMH test with the same as-randomized stratification factors. The FAS set will be used for this purpose and treatment group comparisons (VTX002 60 mg vs placebo; VTX002 30 mg vs placebo) will be as-randomized. No adjustment for multiplicity will be made.

Additionally, for clinical remission, endoscopic improvement, endoscopic remission, histologic improvement (Geboes), and histologic remission (Geboes) at Week 13, separate logistic regression models with stepwise selection will be built to assess the impact of the following covariates measured at baseline:

- ALC derived from blinded hematology laboratory results
- c-RP

Randomization stratification factors (from IWRS) will also be included in the stepwise selection. The alpha level for both entering and removing a covariate will be 0.10. Point estimates and the corresponding 95% CIs for the odds ratios will be calculated.

Repeated measures Induction/LTE data on the same endpoints (clinical remission, endoscopic improvement, endoscopic remission, histologic improvement (Geboes), and histologic remission (Geboes) at Weeks 4, 8, 10, 13, 18, 26, 36, and 52 will be analyzed. This will be done using repeated measurements logistic regression specifying the distribution as binomial, the link function as logit and randomization stratification factors added as covariates in addition to randomized treatment. Two summaries will be provided: summary for the Induction Treatment Period will include only data from Weeks 4, 8, 10 and 13, whereas a comprehensive summary for both Induction and LTE will include data through Week 52. Sample SAS code is presented below:

16.4 Exploratory Efficacy

The endpoints described here will be analyzed separately, as appropriate, for the Induction Treatment Period, Long-Term Extension Treatment Period (LTE), and Open-Label Extension (OLE). In general, when an endpoint refers to Induction Study Period visits, that endpoint will be included in the Induction Period Analysis, but restricted to visits associated with the Induction Study Period. For example, CRP change from baseline analysis will be included in the Induction Period Analysis, but restricted to Weeks 1, 4, 8, and 13. LTE and OLE analyses for the same endpoint may differ not only with respect to the participant set but also with respect to statistical methodology.

- Proportion of participants with histologic remission at Week 13 using RHI
- Proportion of participants with histologic remission at Week 13 using NHI
- Proportion of participants with histologic improvement at Week 13 using RHI
- Proportion of participants with histologic improvement at Week 13 using NHI
- Proportion of participants with clinical remission after 52 weeks of treatment
- Proportion of participants with clinical remission after 13 and 52 weeks of treatment
- Proportion of participants with symptomatic remission at Weeks 18, 26, 36, and 52
- Proportion of participants with endoscopic remission with histologic remission after 52 weeks of treatment
- Proportion of participants with endoscopic improvement-histologic remission after 52 weeks of treatment
- Proportion of participants with endoscopic and clinical remission after 52 weeks of treatment

- Proportion of participants with endoscopic remission, histologic remission, and clinical remission after 52 weeks of treatment
- Proportion of participants with HEMI after 52 weeks of treatment
- Proportion of participants with clinical response after 52 weeks of treatment
- Proportion of participants with symptomatic response at Weeks 18, 26, 36, and 52
- Proportion of participants with symptomatic response at OLE Week 13, and at each additional OLE visit
- Proportion of participants with any decrease from baseline in Geboes Index score at Week 52
- Change from baseline in FCP at Weeks 1, 4, 8, 13, 26, 36, 52, and at visits during the OLE
- Change from baseline in CRP at Weeks 1, 4, 8, 13, 26, 36, 52, and at visits during the OLE
- Change from baseline in PMS at Weeks 4, 8, 13, 18, 26, 36, 52, and at visits during the OLE
- Change from baseline in total MCS at Week 52
- Change from baseline in IBDQ score at Weeks 13 and 52
- Change and percentage change from baseline in lymphocyte counts at Weeks 1, 4, 8, 13, 26, 36, and 52
- Change and percentage change from baseline in Rectal Bleeding at Weeks 1, 4, 8, 10, 13, 18, 26, 36, and 52
- Change and percentage change from baseline in Stool Frequency at Weeks 1, 4, 8, 10, 13, 18, 26, 36, and 52

The first fifteen exploratory endpoints will be handled as responder analyses and analyzed similarly to the primary endpoint, per timepoint, using the CMH test with the same as-randomized stratification factors. The FAS set will be used for this purpose and treatment group comparisons (VTX002 60 mg vs placebo; VTX002 30 mg vs placebo) will be as-randomized. No adjustment for multiplicity will be made and the nominal Type 1 error rate for each hypothesis test is α=0.05.

The last eight exploratory endpoints are quantitative. Table 7 describes the analyses and analysis sets to be used for treatment group comparisons for these endpoints.

For the parameters listed in Table 7, when Summary Statistics are indicated absolute values, change from baseline, and percent change from baseline will be summarized. The analysis set in the table denotes how treatment group is handled in these summaries. Additionally, figures for mean (SD) absolute value - time curve during study will be created for the following parameters:

CRP, lymphocyte counts, and neutrophil counts using normalized pooled data from both central laboratories. Separate figures will be generated for induction, LTE and OLE Treatment Periods.

Table 7. Quantitative Exploratory Endpoints and Analysis Sets

Parameter	Timepoint	Comparison	Methodology	Analysis Set
FCP (central laboratory)	Induction treatment: Weeks 1, 4, 8, 13	VTX002 60 mg vs Placebo	MMRM [a], Summary Statistics [b]	FAS
		VTX002 30 mg vs Placebo		
	LTE/OLE: Weeks 26, 36, 52 and each visit during the OLE	VTX002 60 mg vs Placebo		FAS
		VTX002 30 mg vs Placebo		
CRP (central laboratory)	Induction treatment: Weeks 1, 4, 8, 13	VTX002 60 mg vs Placebo		FAS
		VTX002 30 mg vs Placebo		
	LTE/OLE: Weeks 26, 36, 52 and each visit during the OLE	VTX002 60 mg vs Placebo		FAS
		VTX002 30 mg vs Placebo		
PMS	Induction treatment: Weeks 4, 8, 13	VTX002 60 mg vs Placebo		FAS, PP
		VTX002 30 mg vs Placebo		
	LTE/OLE: Weeks 18, 26, 36, 52 and each visit during the OLE	VTX002 60 mg vs Placebo		FAS-ET
		VTX002 30 mg vs Placebo		
Total MCS	LTE: Week 52	No formal comparisons, VTX002 60 mg, VTX002 30 mg, and placebo presented by visit	Summary Statistics [b]	FAS ET
IBDQ	Induction treatment: Week 13			FAS
	LTE: Week 52			FAS ET
Lymphocyte counts (central laboratory)	Induction treatment: Weeks 1, 4, 8, 13			FAS
	LTE/OLE: Weeks 26, 36, and 52			FAS ET
Stool Frequency	Induction treatment: Weeks 0, 1, 4, 8, 10, 13			FAS
	LTE/OLE: Weeks 18, 26, 36, 52 and each visit during the OLE			FAS ET

Parameter	Timepoint	Comparison	Methodology	Analysis Set
Rectal Bleeding	Induction treatment: Weeks 0, 1, 4, 8, 10, 13			FAS
	LTE/OLE: Weeks 18, 26, 36, 52 and each visit during the OLE			FAS ET
Symptomatic Response and Symptomatic Remission	Induction treatment: Week 13			FAS
	LTE/OLE: Weeks 18, 26, 36, and 52			FAS ET

[a] Mixed Model for Repeated Measures, used to model change from baseline for Induction Treatment Period separately from the pooled visits across both the Induction and LTE study periods. MMRM generates statistical comparisons between treatment groups, by visit and overall.
[b] Summary statistics are produced by visit for each treatment group and statistical comparisons between treatment groups are not produced.

Where mixed models with repeated measurements (MMRM) are indicated in Table 7, the analysis will be applied to change from baseline in FCP (separately for Cerba and LabCorp central laboratories, Weeks 1, 4, 8, 13, 26, 36, and 52), CRP (Weeks 1, 4, 8, 13, 26, 36, and 52) and PMS (Weeks 4, 8, 13, 26, 36, and 52). The Induction Treatment Period will be handled separately from the pooled visits across both the Induction and LTE study periods. For the Induction Treatment Period, MMRM model(s) will include data from weeks 1, 4, 8 and 13. For the combined Induction and LTE study periods, MMRM model(s) will include all available data from scheduled visits across both study periods.

No imputation of missing results will be done. If a participant has any missing data whatever non-missing data is available for that participant will be used in the MMRM.

Least square means for both change from baseline within treatment group and difference between treatment groups with respect to the outcome variable, together with standard errors, 95% Cis, and pairwise treatment group p-values (VTX002 60 mg versus placebo; VTX002 30 mg versus placebo) will be reported, by visit. Stratification factors will be added as covariates and will be handled as-randomized in all cases regardless of the analysis set being used.

Sample SAS code is presented below:

An Unstructured Covariance structure will be used in Proc Mixed. If that fails to converge, Autoregressive Covariance structure will be used. If that fails to converge, Compound Symmetry will be used. The contain method of computing degrees of freedom will be applied.

16.4.1 Mayo Scoring System

All components of Mayo scoring system (RB, SF, ES and PGA) as well as three Mayo scores (MMS, MCS and PMS) will be listed by visit. RB and SF will be summarized by visit.

16.5 Symptomatic Remission Sensitivity Analyses

Ten endpoints having a direct relationship to Symptomatic Remission will be subjected to an additional sensitivity analysis in which the way Stool Frequency (SF) is handled is altered:

- Primary definition: "Stool frequency (SF) subscore = 0 or 1"
- Sensitivity analysis definition: "Stool frequency (SF) subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline)"

The ten endpoints subjected to this sensitivity analysis are given in Table 8:

Table 8. Symptomatic Remission Sensitivity Analyses

Endpoint	Definition	Sensitivity Analysis Derivation
Primary Endpoint	Proportion of participants with clinical remission at Week 13 using modified Mayo score (MMS)	Stool frequency (SF) subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), RB subscore = 0, and endoscopic subscore (ES) ≤ 1 (excluding friability)
Key Secondary	Proportion of participants with symptomatic remission at Week 13	SF = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0
Other Secondary Endpoint	Proportion of participants with endoscopic and clinical remission at Week 13	ES = 0, SF subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0
Other Secondary Endpoint	Proportion of participants with endoscopic remission, histologic remission, and clinical remission at Week 13	ES= 0, Geboes Index score < 2.0, SF subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0
Other Secondary Endpoint	Proportion of participants with symptomatic remission at Weeks 4, 8, 10	SF = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0
Other Secondary Endpoint	Proportion of participants with clinical remission at Week 13 using total Mayo Clinic score (MCS)	Stool frequency (SF) subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), RB subscore = 0, endoscopic subscore (ES) ≤ 1 (excluding friability), PGA ≤ 1, and SF + RB + ES + PGA ≤ 2
Exploratory	Proportion of participants with clinical remission at Week 52 using modified Mayo score (MMS)	Stool frequency (SF) subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), RB subscore = 0, and endoscopic subscore (ES) ≤ 1 (excluding friability)

Endpoint	Definition	Sensitivity Analysis Derivation
Exploratory	Proportion of participants with clinical remission at *both* Weeks 13 and 52 using modified Mayo score (MMS)	Stool frequency (SF) subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), RB subscore = 0, and endoscopic subscore (ES) ≤ 1 (excluding friability)
Exploratory	Proportion of participants with symptomatic remission at Week 18, Week 26, Week 36, and Week 52 (separately)	SF = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0
Exploratory	Proportion of participants with endoscopic and clinical remission after 52 weeks of treatment	ES = 0, SF subscore = 0 (or = 1 with a ≥ 1-point decrease from baseline), and RB subscore = 0

No imputation will be performed for these sensitivity analyses, non-responders will be handled as defined at the beginning of Section 16, FAS will be used, and these analyses will use the same CMH statistical methodology with randomized stratification factors as other responder analyses. VTX002 60 mg vs Placebo will be handled separately from VTX002 30 mg vs Placebo.

16.6 Pharmacokinetics Analysis and Plasma Concentration

See Section 29.

17. SAFETY

17.1 Adverse Events

17.1.1 General Considerations

Adverse events are collected from the date and time of signing ICF and will be coded according to the MedDRA version 26.0 or higher. Adverse events with an onset prior to first dose of study treatment are collected as part of medical history. Adverse events with an onset on or after first dose of study treatment are collected on adverse event eCRFs. All AE tables will be summarized according to the actual treatment group with an appropriate safety set.

A treatment-emergent adverse event (TEAE) is either an AE occurring on or after the date and time of first study treatment intake, or an ongoing AE worsening in severity after the date and time of the first study treatment. If an AE date coincides with the date of first dose of study treatment and either the AE onset time or drug administration time is missing then the AE will be considered treatment-emergent. AEs having both start and end dates missing will also be considered treatment-emergent. AEs with a missing onset date and non-missing end date will be considered as TEAE unless end date is prior to the first drug administration. Adverse events with missing seriousness will be considered serious. Adverse events with missing severity will be considered to have CTCAE grade 3. Adverse events with missing relatedness will be considered related.

AEs with an onset prior to the date and time of first study treatment intake (i.e., during Screening) will be listed but will not be summarized unless they either result in a TEAE due to AE worsening in severity or are serious. TEAE summaries will be separately prepared for study periods of Induction Treatment Period, LTE, OLE, and for the entire study:

- TEAEs with an onset on or after date and time of first study treatment intake during the Induction Treatment Period but prior to last participation in the Induction Treatment Period will be associated with Induction. Also, TEAEs resulting from worsening of severity during Induction in an AE having onset during Screening will be associated with Induction. Summaries for the Induction Treatment Period will use the Safety Set. This set of TEAEs will be associated with the Induction Period Analysis.
- TEAEs with an onset on or after the date and time of first study treatment intake during the LTE but prior to last participation in LTE will be associated with LTE. Also, TEAEs resulting from worsening of severity during LTE in an AE having onset prior to LTE will be associated with LTE. Summaries for the LTE will use the Safety-ET Set.
- TEAEs with an onset on or after the date and time of first study treatment intake during OLE but prior to last participation in OLE will be associated with OLE. Also, TEAEs resulting from worsening of severity during OLE in an AE having onset prior to OLE will be associated with OLE. Summaries for the LTE will use the Safety-ET Set.
- TEAEs with an onset on or after the date and time of first study treatment intake will be used for entire-study AE summarization. Note that a TEAE may also be an AE having onset during Screening for this purpose, but having worsening severity at some point after first study treatment intake. An AE considered as TEAE because of worsening severity will only be counted once, at maximum severity, for this purpose. Summaries for the entire study will use the Safety-ET Set.

17.1.2 TEAE Summaries

An overall TEAE summary table will be presented separately for each of the study periods described in Section 17.1.1. It will include the number and percentage of participants with at least one event of specified category. The following categories will be displayed:

- TEAE
- Treatment-emergent serious AE (SAE)
- Non-serious TEAEs
- TEAE/treatment-emergent SAE related to study treatment
- AE/TEAE leading to death.
- TEAEs Grade 4 or higher
- TEAEs Grade 3 or higher

- TEAEs Grade 2 or higher
- TEAEs leading to study treatment discontinuation
- TEAEs leading to study treatment interruption
- TEAEs by maximum CTCAE severity
- Related TEAEs by maximum CTCAE severity

The incidence of treatment-emergent SAEs, TEAEs, and non-serious TEAEs will be summarized separately, by the following, for each of the study periods described in Section 17.1.1:

- SOC and PT, ordered by descending frequency of PT in VTX002 60 mg group
- SOC, PT, and relatedness to study treatment, ordered by decreasing frequency in VTX002 60 mg group
- SOC, PT, and maximum Common Terminology Criteria for Adverse Events (CTCAE) toxicity grade and ordered by decreasing frequency in VTX002 60 mg group.
- PT, ordered by descending frequency of PT in VTX002 60 mg group

Additionally, the incidence of treatment-related TEAEs, treatment-related SAEs, treatment-related non-serious TEAEs, TEAEs leading to study treatment interruption, TEAEs leading to premature discontinuation of study treatment, and TEAEs leading to death will be summarized separately, by SOC and PT, for each of the four types of TEAE summaries described in Section 17.1.1 and ordered by descending frequency of PT in VTX002 60 mg group.

In summaries associated with a specific study period – either Induction, LTE, or OLE – if the same TEAE is reported for the same participant more than once, the TEAE is counted only once for each SOC by PT combination. In these study period tabulations, study treatment shall be considered to be the treatment a study participant actually receives during a given study period. If a study participant receives multiple study treatments during a given study period, adverse events for that participant will be tabulated according to the highest dose of study treatment received.

For entire study summaries, if the same TEAE is reported for the same participant more than once, the TEAE is counted only once, at the date of TEAE onset, for each SOC, PT and study treatment combination. For entire study summary purposes, study treatment shall be considered the same in each study period as what is used for TEAE summaries by study period, and a TEAE having onset in a given study period will be linked to the treatment associated with that study period.

Separate listings will be prepared for all AEs, SAEs, AEs leading to study treatment interruption, AEs leading to permanent discontinuation of study treatment and AEs leading to death, with flags for TEAE (Y/N).

17.1.3 Targeted Medical Events (TMEs)

Targeted treatment-emergent adverse events will be separately tabulated and labeled as "targeted medical events". Categories of TMEs will include the following:

- Cardiac
    - Bradycardia
    - AV conduction delay
    - Hypertension
- Macular edema
- Infections
    - Serious infections
    - Opportunistic infection
    - Herpes simplex and herpes zoster infection
- Liver
    - Elevated aspartate aminotransferase (AST)
    - Elevated alanine aminotransferase (ALT)
    - Elevated bilirubin
- Malignancies
- Posterior reversible encephalopathy syndrome (PRES)
- Pulmonary disorders with preferred terms indicating airflow obstruction
- Pulmonary disorders with preferred terms indicating decreased gas exchange

See Section 27 (Appendix VI) for MedDRA terms specifying the identification of Targeted Medical Events (TMEs).

17.2 Laboratory Data

Where data transformation is needed, all available significant digits will be retained prior to populating ADaM dataset(s) with production lab data values.

17.2.1 eGFR

For participants contributing age, sex, and race demographic information, estimated glomerular flow rate (mL/min/1.73m^2) will be calculated from serum creatinine using the 2021 update to the CKD-EPI formula [3]:

$$\text{eGFR} = 141 * \frac{\left(min\left(\frac{\text{SCR (mg/dL)}}{\kappa}, 1\right)\right)^{\alpha}}{\left(max\left(\frac{\text{SCR (mg/dL)}}{\kappa}, 1\right)\right)^{1.209}} * 0.993^{\text{AGE}} * X * Y$$

where

κ = 0.7 for females, 0.9 for males

α=-0.329 for females, -0.411 for males

X=1.018 for females, 1.0 for males

Y=1.159 for Black/African-American, 1.0 otherwise

Serum creatinine expressed in µmol/L (SI units) will be converted to mg/dL prior to application of the CKD-EPI formula. Missing serum creatinine will result in missing eGFR value(s). Participants not contributing any one of age, sex, or race will have eGFR set to missing for all visits. Per-participant eGFR normal range will be determined by passing the relevant central lab serum creatinine normal range through the CKD-EPI formula. It is noted that the threshold for 'High' levels of serum creatinine will be mapped to a 'Low' threshold of eGFR, and the threshold for 'Low' levels of serum creatinine will be mapped to a 'High' threshold of eGFR. Lab values will be flagged as low, normal, or high on the basis of these calculated values, and these flags will be used in tabulations (as opposed to eGFR flags embedded in lab data transferred from central labs). Calculated eGFR will be included in lab data listings and differences between central lab eGFR flags and calculated eGFR flags, if any, will be highlighted in the listings.

17.2.2 Lab Values and Normal Ranges

[redacted]For purposes of data analysis, ADAM.ADLB (and any other ADaM datasets with laboratory data) will put a limited number of central lab values and normal ranges on a common footing, with[redacted] normal ranges serving as the reference ranges. Local lab data will not be transformed in this way. The analytes subject to normalization are absolute lymphocyte count (ALC), absolute neutrophil count (ANC), and high-specificity c-reactive protein (CRP).

[3] https://reference.medscape.com/calculator/251/egfr-using-ckd-epi-2021-update

[redacted] data for ALC, ANC, and CRP will be mapped to [redacted] normal ranges. In some cases age and/or sex may need to be taken into account for a given central laboratory and analyte.

To match the manner in which the two central laboratories handle participant Age, the following rules will be applied during the analyte normalization process:

- [redacted] participant Age at collection during Screening / Baseline will be incremented by +1 every 30 Jun
- [redacted] participant Age at collection during Screening / Baseline will be incremented by +1 every 01 Jan

These Age rules are specific to the analyte normalization process.

The analyte mapping will be done according to the following formulas and the notation in these formulas suppresses both age and sex demographic characteristics for simplicity:

[redacted]

[redacted]

Transformation

Let X_{OBS} = observed lab value

Let X_{REF} = transformed lab value

- [redacted]
- [redacted]

Flag values for individual lab assays (e.g., 'low' vs 'normal' vs 'high', or 'Abnormal, clinically significant' vs 'Abnormal, not clinically significant' vs 'Normal') will not be altered as a result of mapping lab values and normal ranges. These flag values, as determined by central laboratory, will be used in listings and tables. Composite events relying on multiple cutoff values, such as cases of potential drug-induced liver injury, will also be determined using original central lab values and normal ranges.

17.2.3 Summaries

For tabulation purposes, values below assay validity cutoffs will be set equal (e.g., if 1.112 mmol/L is the assay validity lower bound, a reported value of LDL-C < 1.112 mmol/L is set equal to LDL-C = 1.112 mmol/L prior to rounding) and values above assay validity cutoffs will be set equal (e.g., if 13.9 mmol/L is the assay validity upper bound, a reported value of LDL-C > 13.9 mmol/L is set equal to LDL-C = 13.9 mmol/L prior to rounding).

Four general sets of laboratory data summaries will be prepared: Induction Treatment Period, LTE, OLE, and the entire study. Safety Set will be used for induction treatment summaries and Safety-ET Set for LTE/OLE summaries.

Values, change from baseline, and percent change from baseline will be summarized descriptively, by visit, for each treatment group, and only central laboratory data will be tabulated.

- Summary Tables
    - Summary tabulations will only be prepared for central laboratory data using international system (SI) units
    - The normalized assay results (ALC, ANC, and CRP; see Sections 17.2.1 and 17.2.2) will be summarized separately from all other assay results and [redacted]
    - Fecal calprotectin (FCP) will only be summarized separately for [redacted] FCP will not be pooled, except for shift table purposes. See Section 10.5 for clinically meaningful cutoffs applicable to FCP, regardless of central laboratory.
    - For the non-normalized assay results other than FCP, eGFR and hematology and serum chemistry data will be considered separately and will each be summarized [redacted] Visit windowing will be used to associate laboratory data with study visits – see Section 9.6 Table 3.
- Shift Tables
    - Shift tables will include both central and local laboratory data whenever possible.
    - Normalized assay results (ALC, ANC, and CRP) will be considered only on a pooled basis
    - All non-normalized assay results, including FCP and eGFR, will be presented both separately for each central laboratory and also on a pooled basis
    - Shift tables (number and percentage) from baseline to every post-baseline visit will be presented
        - Shift tables for the change in laboratory results interpretation (e.g., flags such as 'Normal' vs 'Low' vs 'High' or 'Abnormal, clinically significant' vs 'Abnormal, not clinically significant' vs 'Normal') from baseline to every post-baseline visit and worst post-baseline visit in Induction Treatment Period
        - Shift tables for the change in CTCAE toxicity grades (with separate parts for decreased and increased grades and Grade = 0 for normal results) from baseline to every post-baseline visit and worse post-baseline visit in Induction Treatment Period

  - The number of participants in a particular treatment group having both baseline and post-baseline results available will be used as the denominator.

- Listings
  - Listings will present data in both conventional (CN) and international system (SI) units
  - Hematology, serum chemistry, urinalysis, coagulation data, and pregnancy test data will be listed and these listings will include both central laboratory and local laboratory data.
  - Laboratory values outside normal limits will be identified in the participant data listings with appropriate flags (e.g., 'Normal' vs 'Low' vs 'High' or 'Abnormal, clinically significant' vs 'Abnormal, not clinically significant' vs 'Normal'). Abnormal lab values and CTCAE grades over time will summarized by number and percentage.

Additionally, the number and percentage of participants meeting the following criteria will be presented by treatment group and visit:

- Elevated ALT or AST levels (>3 x ULN and total bilirubin > 2 × ULN or international normalized ratio (INR) > 1.5; >5 x ULN for > 2 weeks and >8 x ULN).
- ALT or AST > 8x ULN
- ALT or AST > 5x ULN
- ALT or AST > 3x ULN
- ALT or AST > 3x ULN and total bilirubin > 2x ULN
- ALT or AST > 3x ULN and INR > 1.5x ULN

The denominator will be the number of participants in treatment group with non-missing data at specified visit.

17.3 Vital Signs

Similar to laboratory data analysis, three sets of vital signs summaries will be prepared: for Induction Treatment Period, LTE and OLE Treatment Periods. Safety Set will be used for induction treatment summaries and Safety-ET Set for LTE/OLE summaries.

Vital signs include data on pulse rate (bpm), systolic/diastolic blood pressure (mmHg), respiratory rate (breaths/min), temperature (°C), weight (kg) and height (cm). Descriptive statistics will be provided for observed values, change from baseline, and percent change from baseline to post-baseline time points for every parameter. Visit windowing will be used to associate vital sign data with study visits – see Section 9.6 Table 3.

A listing of vital signs data will be presented.

17.4 Physical Examination

A listing will be provided for physical examination findings including description of any abnormalities.

17.5 Electrocardiogram (ECG)

Similar to laboratory data analysis, three sets of ECG summaries will be prepared: for Induction Treatment Period (Safety Set), LTE and OLE Treatment Periods (Safety ET Set).

17.5.1 Twelve-Lead ECG

ECG data will be presented by treatment group for each study period (Induction, LTE, OLE) and overall.

The 12-lead ECG data will include heart rate (bpm), RR (msec), PR (msec), QRS (msec), QT (msec) and QTcF (msec) intervals. Absolute values and changes from baseline to post-baseline time points will be summarized using descriptive statistics for quantitative variables. Shift tables from baseline ('Normal', 'Abnormal, not clinically significant' and 'Abnormal, clinically significant') to each visit will be provided. ECG data will be read locally as well as centrally. Only central ECG data will be used for summaries, all available ECG data from local and central readers will be listed. Actual results will not be reconciled. In case of several assessment within timepoint, the earlier assessment will be used for the analysis.

The following incidences will be summarized by number and percentage at every post-baseline visit:

- QT interval
    - ≥ 500 msec
    - Change from baseline > +30 msec
    - Change from baseline > +60 msec
- QTcF interval
    - ≥ 450 msec (male) or ≥ 470 msec (female)
    - Change from baseline > +30 msec
    - Change from baseline > +60 msec
- PR interval
    - PR > 200 msec
    - PR > 230 msec
- HR
    - < 40 beats/min (bpm)

  - < 50 bpm
  - > 110
  - Change from baseline < -10 bpm
- AV Block
  - First-degree AV Block
  - Second-degree AV Block Type 1
  - Second-degree AV Block Type 2
  - Third-degree AV Block

The denominator will be the number of participants in treatment group with non-missing results at the specified visit.

17.5.2 Holter ECG monitoring

On Day 1 of Induction Treatment Period and OLE Day 1 continuous Holter ECG monitoring will be performed for at least 1 hour predose and at least 24 hours postdose.

Overall interpretation ('Normal', 'Abnormal', 'Not Evaluable') will be summarized by number and percentage. Denominator will be the number of participants available at specified visit in treatment group.

Parameters of interest will be listed.

17.6 Optical Coherence Tomography

Four summaries will be prepared: for Induction Treatment Period (Safety Set), LTE, OLE, and overall, using the Safety and Safety-ET sets.

Shift tables from baseline ('Normal', 'Abnormal, not clinically significant' and 'Abnormal, clinically significant') to each visit will be provided. Denominator will be the number of participants in treatment group with non-missing results at specified visit.

17.7 Pulmonary Function Test

Similar to laboratory data analysis, four sets of summaries will be prepared: for Induction Treatment Period (Safety Set), LTE, OLE, and overall, using the Safety and Safety-ET sets.

The pulmonary function test data will include the following parameters: Forced Expiratory Volume (FEV1) (L), Predicted FEV1 (L), Percentage of predicted FEV1 (%), Forced Vital Capacity (FVC) (L), FVC (%), Peak Expiratory Flow Rate (PEFR) (L/sec), Forced Expiratory Flow (FEF) (L/sec), Total Lung Capacity (L), Percentage of Total Lung Capacity (%), Predicted Total Lung Capacity (L), DLCO, predicted DLCO, percentage of predicted DLCO.

Absolute values and changes from baseline to post-baseline time points will be summarized using descriptive statistics for quantitative variables. Abnormal findings and incidences of FEV1, FVC, and/or FEV1/FVC ratio will be summarized by number and percentage. Denominators will be the number of participants in treatment group with non-missing results at specified visit.

All pulmonary function test data will be listed. The following criteria will be used to define abnormal findings and incidences:

- Decrease from baseline > 20% in FEV1 (absolute value)
- Decrease from baseline > 20% in FVC (absolute value)
- Decrease from baseline > 20% in DLCO (absolute value)
- Decrease from baseline in % predicted FEV1 > 20%
- Decrease from baseline in % predicted FVC > 20%
- Decrease from baseline in % predicted DLCO > 20%
- % Predicted FEV1 < 50%
- % Predicted FVC < 50%
- % Predicted FEV1/FVC ratio < 50%

18. DEVIATIONS FROM ANALYSIS AS DESCRIBED IN THE PROTOCOL

Study Day for Screening will run from Day -28 to Day -1, versus the Protocol SOA which says Screening will be from Day -28 to Day 0.

FCP [redacted] will be *separately* modelled using MMRM models due to major differences in analytical methods between these laboratories. See Section 16.4.

19. PROGRAMMING SPECIFICATIONS

All outputs will be produced using SAS version 9.4 or a later version. Detailed programming specifications are provided in the Mock Tables, Listings and Figures (TFLs) file.

20. LIST OF TABLES, LISTINGS, AND FIGURES

See Mock TFLs file.

21. REFERENCES

Bergland, P (2015). Multiple Imputation Using the Fully Conditional Specification Method: A Comparison of SAS®, Stata, IVEware, and R. SAS Global Forum Proceedings, Paper 2018-2015 (https://support.sas.com/resources/papers/proceedings15/2081-2015.pdf).

Bergland, P (2018). Using SAS® for Multiple Imputation and Analysis of Longitudinal Data. SAS Global Forum Proceedings, Paper 1738-2018 (https://support.sas.com/resources/papers/proceedings18/1738-2018.pdf).

Dmitrienko A, Tamhane AC, Wiens BL. (2008). General multistage gatekeeping procedures, Biometrical Journal. 50, 667–677

Dmitrienko A, Millen BA, Brechenmacher T, Paux G. (2011a). Development of gatekeeping strategies in confirmatory clinical trials. Biometrical Journal, 53, 875–893.

Dmitrienko A, Tamhane AC. (2011b). Mixtures of multiple testing procedures for gatekeeping applications in clinical trials. Statistics in Medicine, 30, 1473–1488.

FDA Center for Drug Evaluation and Research (2022). Ulcerative Colitis: Developing Drugs for Treatment.

ICH Guidance for Industry (2013): E3 Structure and Content of Clinical Study Reports: Questions and Answers (R1), pages 6-7, Question Q7 and Answer A7

Mantel N and Fleiss JL (1980). Minimum expected cell size requirements for the mantel-haenszel one-degree-of-freedom chi-square test and a related rapid procedure. Am. J. of Epidemiology. 112:1;129-134.

Ratitch B, Lipkovich I, O'Kelly M (2013). Combining Analysis Results from Multiply Imputed Categorical Data. Pharmaceutical SAS Users Group, Paper SP03 (https://www.lexjansen.com/pharmasug/2013/SP/PharmaSUG-2013-SP03.pdf).

Rubin DB (1987). Multiple imputation for nonresponse in surveys. New York, NY: John Wiley and Sons.

Sandborn WJ, Feagan BG, Wolf DC, et al. (2016) Ozanimod induction and maintenance treatment for ulcerative colitis. N Engl J Med. 374;1754-62.

Sandborn WJ, Peyrin-Biroulet L, Zhang J, et al. (2020) Efficacy and safety of etrasimod in a Phase 2 randomized trial of participants with ulcerative colitis. Gastroenterol. 158;550-61.

22. APPENDIX I. MAYO SCORING SYSTEM

The Mayo score ranges from 0 to 12, with higher scores indicating more severe disease.

Stool frequency subscore: Each participant serves as his or her own control to establish the degree of abnormality of the stool frequency.

0=Normal number of stools for this participant

1=1 to 2 stools more than normal

2=3 to 4 stools more than normal

3=5 or more stools more than normal

Subscore: 0 to 3

Rectal bleeding subscore: The daily bleeding score represents the most severe bleeding of the day.

0=No blood seen

1=Streaks of blood with stool less than half the time

2=Obvious blood with stool most of the time

3=Blood alone passes

Subscore: 0 to 3

Findings on endoscopy (endoscopy subscore): The endoscopy subscore will be determined by qualified personnel at a central laboratory.

0=Normal or inactive disease

1=Mild disease (erythema, decreased vascular pattern)

2=Moderate disease (marked erythema, lack of vascular pattern, friability, erosions)

3=Severe disease (spontaneous bleeding, ulceration)

Subscore: 0 to 3

Physician's Global Assessment: The Physician's Global Assessment acknowledges the 3 other criteria, the participant's daily recollection of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the participant's performance status.

0=Normal

1=Mild disease

2=Moderate disease

3=Severe disease

Subscore: 0 to 3.

The following combinations of subscores are used in efficacy endpoints definitions:

Modified Mayo score (MMS): Endoscopy subscore + Rectal bleeding subscore + Stool frequency subscore;

Mayo Clinic score (MCS): Endoscopy subscore + Rectal bleeding subscore + Stool frequency subscore + Physician's Global Assessment;

Partial Mayo Score (PMS): Rectal bleeding subscore + Stool frequency subscore + Physician's Global Assessment.

23. APPENDIX II. RELATIONSHIP OF MFAS TO ENDPOINTS

Table 9. Relationship of Primary Endpoint and Key Secondary Endpoints to Post-Baseline Assessments Occurring at the Same Visit

	Primary Endpoint	Key Secondary Endpoints			
Non-Missing Assessments at the Same Post-Baseline Visit	**Clinical Remission**	**Endoscopic Improvement**	**Symptomatic Remission**	**Histologic Remission**	**Endoscopic improvement-histologic remission**
ES		☑			
Geboes				☑	
ES, Geboes					☑
RB, SF			☑		
ES, RB, SF	☑				

For example, in order for a participant to contribute non-trivially to Symptomatic Remission at Induction Treatment Period Week 13 (Visit 7), their RB and SF scores must both be available at the Week 13 visit, according to the Visit Windows described in Section 9.6. Because such a participant contributes *at least one* complete post-baseline Induction Treatment Period set of assessments for Symptomatic Remission, they are members of mFAS at *all* Induction Treatment Period post-baseline visits for Symptomatic Remission. This is true even if (say) Symptomatic Remission cannot be determined at Week 8 (Day 57) because SF is missing at Week 8.

24. APPENDIX III. VISIT LABEL MAPPING

EDC underwent significant changes between Protocol versions 1.0/2.0 and ≥ 3.0, and especially for the OLE study period. The following visit mapping is intended to ensure that ADaM datasets utilize a consistent set of visit labeling conventions for tabulation, listing, and graphical purposes.

Table 10. Visit Label Mapping

Period	FolderName (Visit)		FolderName Mapped	PK Endpoints per Protocol	PK Sampling		
					Predose	2, 4, and 6 hours postdose (1, 2, 4 hrs postdose for V1)	24 hours postdose
Induction	Screening		Screening				
	Visit 2 Week 0		Visit 2 Week 0	Week 0	X	X	
	Visit 3 Week 1		Visit 3 Week 1	Week 1	X		
	Visit 4 Week 4		Visit 4 Week 4	Week 4	X		
	Visit 5 Week 8		Visit 5 Week 8	Week 8	X		
	Visit 6 Week 10		Visit 6 Week 10				
	Visit 7 Week 13		Visit 7 Week 13	Week 13	X - PV >2		
Long Term Extention (LTE)	LTE Visit 8 Week 13		LTE Visit 8 Week 13				
	LTE Visit 9 Week 18		LTE Visit 9 Week 18	Week 18	X		
	LTE Visit 10 Week 26		LTE Visit 10 Week 26	Week 26	X		
	LTE Visit 11 Week 36		LTE Visit 11 Week 36	Week 36	X		
	LTE Visit 12 Week 52		LTE Visit 12 Week 52	Week 52	X		
Open Label Extention (OLE)		**OLE Week**					
Protocol > V2	OLE Visit 1 Week 0	0	OLE Visit 1 Week 0		X	X	X
Protocol V1	Visit 8 Week 13	0	OLE Visit 1 Week 0		X		
	OLE Visit 2 Week 1	1	OLE Visit 2 Week 1		X	X	X
Protocol V1	Visit 9 Week 14	1	OLE Visit 2 Week 1				
	OLE Visit 3 Week 4	4	OLE Visit 3 Week 4		X	X	X
Protocol V1	Visit 10 Week 18	5	**OLE Visit 3X Week 5**				
	OLE Visit 4 Week 8	8	OLE Visit 4 Week 8		X		

Period	FolderName (Visit)		FolderName Mapped	PK Endpoints per Protocol	PK Sampling		
					Predose	2, 4, and 6 hours postdose (1, 2, 4 hrs postdose for V1)	24 hours postdose
	OLE Visit 5 Week 13	13	OLE Visit 5 Week 13		X		
Protocol V1	Visit 11 Week 26	13	OLE Visit 5 Week 13				
	OLE Visit 6 Week 18	18	OLE Visit 6 Week 18		X		
Protocol V1	Visit 12 Week 36	23	**OLE Visit 6X Week 23**				
	OLE Visit 7 Week 26	26	OLE Visit 7 Week 26		X		
	OLE Visit 8 Week 36	36	OLE Visit 8 Week 36		X		
	OLE Visit 9 Week 39	39	OLE Visit 9 Week 39		X		
	Visit 13 Week 52 / End of Treatment	39	OLE Visit 9 Week 39				
	OLE Visit 10 Week 51	51	OLE Visit 10 Week 51				
	OLE Visit 11 Week 63	63	OLE Visit 11 Week 63				
	OLE Visit 12 Week 75	75	OLE Visit 12 Week 75				
	OLE Visit 13 Week 91	91	OLE Visit 13 Week 91		X		
	OLE Visit 14 Week 103	103	OLE Visit 14 Week 103				
	OLE Visit 15 Week 115	115	OLE Visit 15 Week 115				
	OLE Visit 16 Week 127	127	OLE Visit 16 Week 127				
	OLE Visit 17 Week 143	143	OLE Visit 17 Week 143		X		
Other							
	Unscheduled Visit		Unscheduled Visit				
	Early Termination Visit		Early Termination Visit		X		
	1-Week Follow-Up		1-Week Follow-Up		X		
	2-Week Follow-Up		2-Week Follow-Up		X		

25. APPENDIX IV. RESCUE THERAPY

Ventyx team members will classify therapies (medications and/or procedures) as 'rescue' for UC prior to interim database lock and in a blinded manner. Rescue therapy identification will occur prior to the Induction Period Analysis database lock and read-out. The method for Ventyx team members to use to perform this classification is described in this section. Rescue therapy identification will assist in establishing intercurrent events and Per Protocol analysis set membership.

Rescue therapies identified by Ventyx reviewers will be imported in statistical programming. Identified rescue therapies applicable to the Induction Period Analysis will be finalized before interim database lock and study unblinding. Only medications and medical procedures reported on the eCRFs will be assessed to determine whether they are rescue therapy for UC.

If rescue therapy exposure does not occur prior to efficacy assessment, then it is not considered a rescue therapy for the purposes of efficacy assessments occurring prior to rescue therapy exposure. The rules outlined below apply to both new use and increase in dose from Baseline.

Biologics with immunomodulatory properties

- Rule:
  - Any exposure after first dose of study treatment
- List of medications:
  - AntiTNFα antibodies:
    - ADALIMUMAB
    - CERTOLIZUMAB
    - CERTOLIZUMAB PEGOL
    - GOLIMUMAB
    - INFLIXIMAB
  - Other Biologics:
    - USTEKINUMAB
    - VEDOLIZUMAB

Non-biologics with immunomodulatory properties

- Immunosuppressants
  - Rules to consider for Visit 7 Week 13 efficacy endpoints
    - Must be discontinued > 2 weeks prior to Screening endoscopy

        - Any treatment during the duration of the study
    - List of medications:
        - MERCAPTOPURINE
        - AZATHIOPRINE
        - TIOGUANINE
        - METHOTREXATE
        - METHOTREXATE SODIUM
- 5-ASA Compounds
    - Rules to consider for Visit 7 Week 13 efficacy endpoints
        - After Screening and throughout time on study: any change from baseline
            - Discontinuation or reduction in dose allowed per Investigator judgement deems subject safety due to toxicity or medical necessity
        - Rectal route of administration must be discontinued prior to Screening
    - List of medications:
        - MESALAZINE
        - BALSALAZIDE
        - BALSALAZIDE DISODIUM DIHYDRATE
        - BALSALAZIDE SODIUM
        - OLSALAZINE SODIUM
        - SULFASALAZINE
        - BECLOMETASONE W/MESALAZINE
    - Routes:
        - ORAL
        - RECTAL
- Other small molecule immunomodulatory active agents
    - Rules to consider for Visit 7 Week 13 efficacy endpoints:
        - Must be discontinued ≤ 8 Weeks prior to Screening

        - Any treatment during the duration of the study
    - List of medications:
        - CICLOSPORIN
        - TACROLIMUS
        - TOFACITINIB
        - TOFACITINIB CITRATE
        - FILGOTINIB
        - UPADACITINIB
- Systemic glucocorticoids
    - Systemic glucocorticoids given via oral or rectal routes of administration
        - Rules to consider for Visit 7 Week 13 efficacy endpoints
            - After first dose of study treatment and up to and including Week 13: any increase from baseline (including new use) for more than 7 days
            - After Week 8: Any dose above baseline
            - After Week 13: dose may be tapered; however, if corticosteroid taper cannot be tolerated without recurrence of clinical symptoms of either UC or steroid withdrawal, corticosteroid dose may be increased (up to the dose at study entry if required), but tapering should be attempted again within 2 weeks
        - List of medications:
            - BETAMETHASONE
            - BETAMETHASONE DIPROPIONATE
            - BETAMETHASONE SODIUM PHOSPHATE
            - DEXAMETHASONE
            - DEXAMETHASONE SODIUM PHOSPHATE
            - DEXAMETHASONE VALERATE
            - METHYLPREDNISOLONE
            - METHYLPREDNISOLONE SODIUM SUCCINATE

- PREDNISOLONE
- PREDNISOLONE SODIUM PHOSPHATE
- PREDNISOLONE METASULFOBENZOATE SODIUM
- PREDNISONE
- TRIAMCINOLONE
- HYDROCORTISONE
- HYDROCORTISONE ACETATE
- HYDROCORTISONE BUTYRATE
- HYDROCORTISONE SODIUM SUCCINATE

  - Routes:
    - ORAL
    - RECTAL

- Systemic glucocorticoids given via parenteral routes of administration
  - Rule for Visit 7 Week 13 efficacy endpoints:
    - Any exposure after ≤ 2 weeks prior to Screening endoscopy
  - List of medications:
    - BETAMETHASONE
    - BETAMETHASONE DIPROPIONATE
    - BETAMETHASONE SODIUM PHOSPHATE
    - DEXAMETHASONE
    - DEXAMETHASONE SODIUM PHOSPHATE
    - DEXAMETHASONE VALERATE
    - METHYLPREDNISOLONE
    - METHYLPREDNISOLONE SODIUM SUCCINATE
    - PREDNISOLONE
    - PREDNISOLONE SODIUM PHOSPHATE
    - PREDNISOLONE METASULFOBENZOATE SODIUM
    - PREDNISONE
    - TRIAMCINOLONE

        - HYDROCORTISONE
        - HYDROCORTISONE ACETATE
        - HYDROCORTISONE BUTYRATE
        - HYDROCORTISONE SODIUM SUCCINATE
    - Routes:
        - INTRAVENOUS
        - INTRAMUSCULAR
- Topical Glucocorticoids
    - Rules for Visit 7 Week 13 efficacy endpoints:
        - After first dose of study treatment and up to and including Week 8: any exposure above baseline (or new use) for more than 5 days
        - After Week 8: Any dose above baseline
    - List of medications:
        - BUDESONIDE (entocort, cortiment, uceris)
    - Routes:
        - ORAL
        - RECTAL
- Beclomethasone
    - Rules for Visit 7 Week 13 efficacy endpoints:
        - After first dose of study treatment and up to and including Week 8: Any increase above baseline for more than 5 days
        - After Week 8: Any dose above baseline
    - List of medications:
        - BECLOMETASONE
        - BECLOMETASONE DIPROPIONATE
        - BECLOMETASONE W/MESALAZINE
    - Routes:
        - ORAL
        - RECTAL

Medical procedures

- Leukocyte apheresis, other apheresis, and plasma exchange
  - Rule:
    - Any exposure after first dose of study treatment
  - List of medical procedures
    - APHERESIS
    - LEUKAPHERESIS
    - IMMUNOGLOBULIN THERAPY
    - IMMNOADSORPTION THERAPY
    - PLASMAPHERESIS
- Surgeries:
  - Rule:
    - Any exposure after first dose of study treatment
  - List of surgeries:
    - COLECTOMY (partial or total)
    - SIGMOIDECTOMY
    - COLOSTOMY
    - ILEOSTOMY

26. APPENDIX V. MULTIPLE IMPUTATION

26.1 JAV Multiple Imputation

26.1.1 Variable Ordering

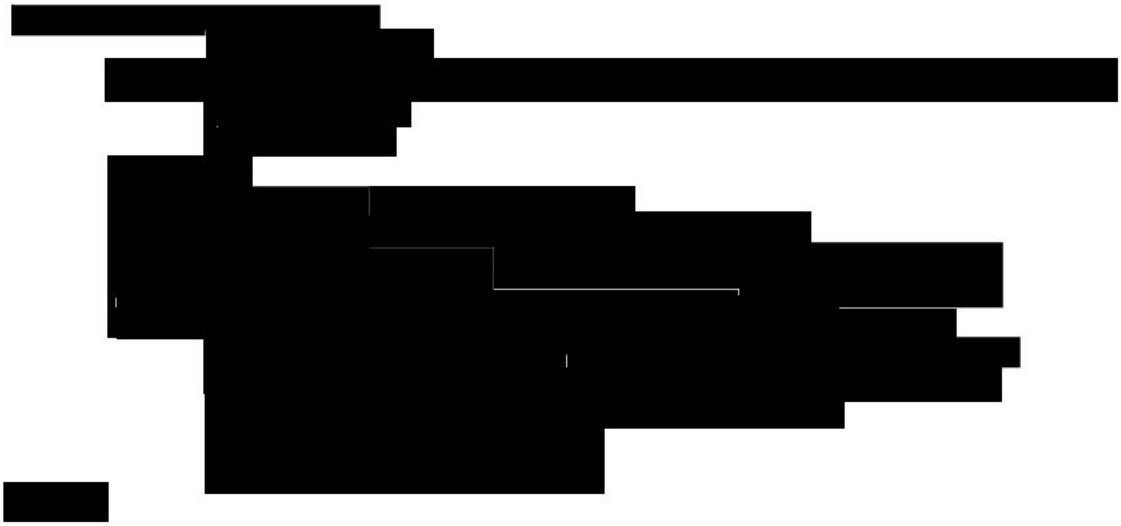
Table 11 presents the variable ordering to be used when performing "Just Another Value" multiple imputation with PROC MI. This ordering is intended to put the variables in order from those having the least number of missing values to those having the most missing values, from left to right, when structured as a "wide" dataset. If a different post-DB lock ordering is found to be more appropriate with respect to missing values, the variable ordering in Table 11 may be altered:

Table 11. Variable Ordering for Just Another Value (JAV) Multiple Imputation

	Variable Ordering
1.	Biologic/JAK inhibitor prior use status (Y or N)
2.	Baseline corticosteroid use (Y or N)
3.	Baseline disease activity (MMS 5-6 or 7-9)
4.	Oral 5-ASA Failure
5.	ES_Baseline
6.	ES_Visit 7
7.	Geboes_Baseline
8.	Geboes_Visit 7
9.	PGA_Visit 2
10.	PGA_Visit 4
11.	PGA_Visit 5
12.	PGA_Visit 7
13.	ln(ALC_Visit 1)
14.	ln(ALC_Visit 2)
…	
19.	ln(ALC_Visit 7)
21.	ln(FCP_Visit 2)
…	
24.	ln(FCP_Visit 5)
25.	ln(FCP_Visit 7)
26.	ln(CRP_Visit 2)
…	
29.	ln(CRP_Visit 5)
30.	ln(CRP_Visit 7)
31.	SF_Baseline
32.	SF_Visit 2
…	
37.	SF_Visit 7
38.	RB_Baseline
39.	RB_Visit 2
…	
44.	RB_Visit 7

26.1.2 PROC MI Code

The following sample PROC MI code illustrates how PROC MI will be applied to the "wide format" data (one record per participant, with variables ordered according to Table 11):



26.2 Transformations

Note: it is assumed the 'MIOUT' dataset referred to in Section 26.1.2 has been post-processed with the PROC FREQ code in Section 16.1.3 and that the appropriate per-imputation test statistics have been extracted into 'ODDS' and 'CMH'. No transformation of risk differences is needed, so sample code for PROC MIANALYZE is not presented.

The following sample code illustrates how the transformation of odds ratios will be handled prior to and after PROC MIANALYZE:

The following sample code illustrates how the transformation of the CMH test statistic will be handled prior to and after PROC MIANALYZE:

27. APPENDIX VI. TARGETED MEDICAL EVENTS

Targeted Medical Events (TMEs) are TEAEs that are potentially of clinical interest that the Ventyx clinical development and biostatistics teams determined based on experience with VTX002 and on information from labels of similar approved compounds. TMEs will be determined using pre-defined Standardised MedDRA Queries (SMQs), and/or custom-defined MedDRA queries (CMQs) that are specific but are not pre-defined within MedDRA.

With regards to reporting of the VTX002-201 TMEs, once the database is locked (either the interim lock for the Induction Period Analysis or the final database lock), a SAS program can used to identify the reported AEs that are classified as VTX002-201 TMEs. A table of TMEs by SOC and PT will be generated as part of this statistical analysis plan (SAP). A standalone document called "Targeted Medical Events" is maintained by the Ventyx Clinical Development Department which outlines the process to identify TMEs from recorded TEAEs.

28. APPENDIX VII. DETERMINATION OF SAMPLE SIZE

Based on literature data review, the expected proportion of clinical remission at Induction Week 13 is 28% for VTX002 60 mg and 8% for placebo (Sandborn 2016; Sandborn 2020). Under these assumptions, and with a 1:1 randomization ratio, 2-group chi-squared test, and two-sided significance level of 5%, a sample of 57 participants per treatment group will be sufficient to achieve at least 80% power. Adding a 10% inflation for dropouts results in 63 participants per treatment group and 189 participants in total.

Software

SAS (version 9.4; SAS Institute Inc.; Cary, NC, USA).

Input & Output

[redacted]

The POWER Procedure
Pearson Chi-square Test for Proportion Difference

Fixed Scenario Elements	
Distribution	Asymptotic normal
Method	Normal approximation
Number of Sides	2
Alpha	0.05
Group 1 Proportion	0.08
Group 2 Proportion	0.28
Nominal Power	0.8
Null Proportion Difference	0

Computed N per Group	
Actual Power	N per Group
0.802	57

29. APPENDIX VIII. RELATIONSHIP OF INTERCURRENT EVENTS TO NON-RESPONDER IMPUTATION, MULTIPLE IMPUTATION, COPY REFERENCE IMPUTATION, AND TIPPING POINT

Table 12. Relationships Between Various Types of Imputation and Data Missingness

Type of Occurrence	Visits Affected	Effect of Protocol Deviations	Primary Analysis of Primary, Key Secondary, and Other Secondary Endpoints	Multiple Imputation	Copy Reference	Tipping Point
Intercurrent Event	All visits on/after intercurrent event, regardless of data missingness for other reasons. Intercurrent event handling takes priority over missing data handling.	An intercurrent event is a major protocol deviation. As long as other deviations don't prevent the participant from being Per Protocol, they will be included in the PP Set.	Non-responder imputation	Non-responder imputation	Non-responder imputation	Non-responder imputation
Missing Data	Visits w/ efficacy endpoints having missing values in one or more endpoint components.	As long as there are no intercurrent events, missing data is handled per this row in the table.	Non-responder imputation	Multiple imputation	Copy reference imputation	Tipping point imputation

30. APPENDIX IX. PK ANALYSIS PLAN

PK Analysis Plan was separately approved and final version 1.0 signed 12 September 2023