Janssen Research & Development

Statistical Analysis Plan for the Phase 3 Maintenance Study

A Phase 2b/3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Protocol to Evaluate the Efficacy and Safety of Guselkumab in Participants with Moderately to Severely Active Ulcerative Colitis

QUASAR

Protocol CNTO1959UCO3001; Phase 2b and Phase 3
Amendment 2

CNTO1959 (guselkumab)

Status:	Approved
Date:	08 September 2023
Prepared by:	Janssen Research & Development, LLC
Document No.:	EDMS-RIM-413689, 6.0

Compliance: The study described in this report was performed according to the principles of Good Clinical Practice (GCP).

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

AMENDMENT HISTORY

DOCUMENT HISTORY	
Document	**Date**
Amendment 2	08 September 2023
Amendment 1	04 October 2022
Original SAP	04 May 2021

Amendment 2 (08 September 2023)

The main reasons for this amendment are to address comments from a Health Authority and to add additional analyses. In particular, the following high-level changes have been made:

- updated sensitivity analysis 1 for the primary endpoint: a tipping point analysis based on multiple imputation with Bernoulli draws instead of imputing clinical remission status in an increasing manner;
- addition of supplementary analyses for the primary and major secondary endpoints as appropriate, where the Mayo rectal bleeding and stool frequency subscores will be calculated based on all available diary data collected in the 7-day window prior to a clinical visit, excluding the day of an endoscopy and the day prior to an endoscopy;
- addition of analyses for the primary and major secondary endpoints for participants with an induction baseline modified Mayo score of 4.

The table below includes all changes made in this amendment.

Section Number and Name	Description of Change	Brief Rationale
	Added definitions of induction and maintenance baseline	To add clarification.
Section 5.3.2.2.1.	Replaced sensitivity analysis 1, imputing clinical remission status in an increasing manner with tipping point analysis based on multiple imputation with Bernoulli draws	Per Health Authority request
Section 5.3.2.2.2. Sensitivity Analysis 2: Multiple Imputation	Reduced the number of imputations due to computational challenges	To address computational issues.
Section 5.3.2.3. Supplementary Estimands for the Primary Endpoint; Section, 5.4.2. Supplementary Estimands for the Major Secondary Endpoints	Added Estimand 4 and Estimand 8 (Alternative Mayo Calculation 2) which considers all available Mayo diary data in the 7-day window, excluding the day of endoscopy and the day prior to an endoscopy.	Per Health Authority request
Section 5.5. Other Endpoints Analysis	• Added modified Mayo score to list of endpoints considered in association analyses of select histologic endpoints at Week M-0 with other endpoints at Week M-44. • Added analyses for primary and major secondary endpoints among participants with an induction baseline modified Mayo score of 4.	To gain a better understanding of the data

	• Added fecal calprotectin over time among participants with fecal calprotectin >150, ≤ 150 mg/kg at induction baseline • Added ≥7, 9-points improvement cut-offs to 7 domain T-scores of PROMIS-29 • PRO endpoints related to fatigue were added	
Section 5.7. Efficacy in Participants Who Had a Dose Adjustment	• Removed fecal calprotectin from Dose Adjustment analyses. • Added symptomatic response • Added partial Mayo score response	To gain a better understanding of the data.
Section 5.9.2. Adverse Events	• Added AE severity to summary of overall AEs	Per Health Authority request
Section 5.10.6. Definition of Subgroups	Added new subgroups based on modified Mayo score and fecal calprotectin cut-offs,	Examine subgroups to identify potential heterogeneity.
Section 6.3. Appendix 3 Demographics and Baseline Characteristics	Added additional categories for age and modified Mayo score	To gain a better understanding of the patient population
Section 6.9. Appendix 9 Laboratory Toxicity Grading	Criteria for select lab tests updated	Updated the criteria to avoid ambiguity or confusion.
Section 5.4.1.3. Figure 4a,4b; Section 5.5. Other Endpoints Analysis; Section 5.5.1. Definitions; Section 5.5.3. Analysis Methods for the Estimands for the Other Endpoints; Section 5.6. Efficacy Endpoints in the Nonrandomized Full Analysis Set; Section 5.9.2. Adverse Events; Section 5.10.6. Definition of Subgroups; Section 6.3. Appendix 3 Demographics and Baseline Characteristics; Attachment 2 Detailed Prohibited Changes in UC Medications Rules	Some minor updates to clarify endpoints as well as related hypotheses and statistical analyses, and to correct editorial mistakes.	To add further clarification.

1. INTRODUCTION

The Phase 2b/3 clinical development program for guselkumab in ulcerative colitis (UC; QUASAR) will evaluate the safety and efficacy of guselkumab compared with placebo and will be conducted under a single protocol. Under this single protocol there will be 3 separate studies: a Phase 2b induction dose-ranging study (Induction Study 1), a Phase 3 induction study (Induction Study 2), and a Phase 3 maintenance study (Maintenance Study).

This Statistical Analysis Plan (SAP) contains definitions of analysis sets, derived variables, and statistical methods for all planned analyses for the Maintenance Study for protocol CNTO1959UCO3001. Separate SAPs are prepared for Induction Study 1 and Induction Study 2.

1.1. Objectives

Primary Objectives

The primary objectives of this study are, in participants with moderately to severely active UC who were induced into clinical response with guselkumab:

- To evaluate the efficacy of maintenance regimens of guselkumab.
- To evaluate the safety of maintenance regimens of guselkumab.

Secondary Objectives

The secondary objectives of this study are, in participants with moderately to severely active UC who were induced into clinical response with guselkumab:

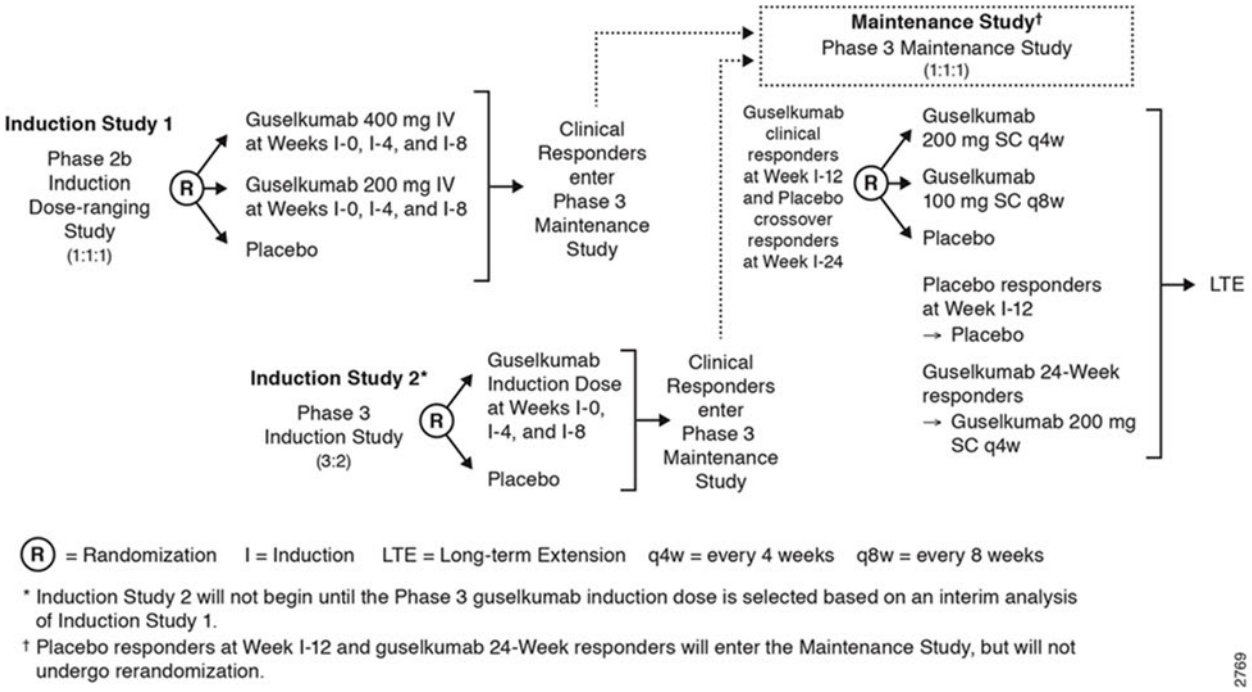
- To evaluate the impact of guselkumab on health-related quality of life (HRQoL) and health economics outcome measures.
- To evaluate the Pharmacokinetics (PK), immunogenicity, and Pharmacodynamics (PD) of guselkumab therapy, including changes in CRP and fecal calprotectin.

1.2. Study Design

The Phase 2b/3 clinical development program for guselkumab in UC (QUASAR) will evaluate the safety and efficacy of guselkumab compared with placebo. As described above, it will be conducted under a single protocol with a Phase 2b induction dose-ranging study (Induction Study 1), a Phase 3 induction study (Induction Study 2), and a Phase 3 maintenance study (Maintenance Study).

An overview of this clinical development program is presented in Figure 1.

Figure 1: Overview of the Quasar Phase 2b/3 Study Design



R = Randomization I = Induction LTE = Long-term Extension q4w = every 4 weeks q8w = every 8 weeks

* Induction Study 2 will not begin until the Phase 3 guselkumab induction dose is selected based on an interim analysis of Induction Study 1.

† Placebo responders at Week I-12 and guselkumab 24-Week responders will enter the Maintenance Study, but will not undergo rerandomization.

The protocol will target participants 18 years of age or older with moderately to severely active UC who have demonstrated an inadequate response or failure to tolerate conventional (i.e., 6-mercaptopurine [6-MP], azathioprine [AZA], or corticosteroids) or advanced therapy (ADT) (i.e., TNFα antagonists, vedolizumab, or tofacitinib). The protocol enrolled participants with a baseline (Week I-0) modified Mayo score of 4 to 9, inclusive, a baseline Mayo rectal bleeding subscore ≥ 1, and a baseline Mayo endoscopy subscore ≥ 2, using the endoscopy score obtained during the central review of the video of the endoscopy.

The primary analysis population for **all three studies** will be randomized and treated participants with a modified Mayo score of 5 to 9 at induction baseline. However, the program also allows for the enrollment of participants with a modified Mayo score of 4, which is capped at ≤5% of the total population.

This program begins with randomization of participants into Induction Study 1. During Induction Study 1, an interim analysis of the first 150 randomized participants who have completed the Week I (induction)-12 visit or have terminated study participation prior to Week I-12 will be performed. The purpose of this interim analysis is to select a single induction dose for confirmatory evaluation in the Phase 3 induction study (Induction Study 2).

Transition from Induction Study 1 to Induction Study 2 of the protocol will occur once the dose decision (in Induction Study 1) has been made and implemented. All participants randomized after the dose decision has been implemented will be part of Induction Study 2.

Participants who have demonstrated a clinical response (defined as a decrease from induction baseline in the modified Mayo score [Section 5.3.1.1] by ≥ 30% and ≥ 2 points, with either a ≥ 1-point decrease from baseline in the rectal bleeding subscore or a rectal bleeding subscore of 0 or 1) in Induction Study 1 or Induction Study 2 will be eligible to enter the randomized-withdrawal maintenance study.

Participants who complete the safety and efficacy evaluations at Week M (maintenance)-44 of the Maintenance Study and who may benefit from continued study intervention, in the opinion of the investigator, will have the opportunity to participate in the long-term extension (LTE) of the Maintenance Study for up to an additional 4 years of treatment to evaluate the efficacy and safety of long-term maintenance treatment.

The detailed study design of the maintenance study is presented below in Section 1.3. See protocol Section 4.1. for more details on the study design for Induction Study 1 and Induction Study 2.

In this document, the time points mentioned for each study or phase refer to Week 0 of that study or phase (induction or maintenance). For example, Week I-12 refers to Week 12 of induction (Study 1 or Study 2). Similarly, Week M-44 refers to 44 weeks after the first maintenance visit (Week M-0) and not 44 weeks after the first induction visit.

1.3. Phase 3 Maintenance Study

In the randomized-withdrawal maintenance study, all participants enrolled will be clinical responders from Induction Study 1 or Induction Study 2. The schema for the maintenance study is shown in Figure 2.

The randomized population of the maintenance study is composed of the following participants:

- Guselkumab clinical responders at Week I-12.
- Placebo crossover responders at Week I-24: Participants initially randomized to placebo who are not in clinical response at Week I-12 who then crossover to guselkumab induction IV treatment and achieve clinical response at Week I-24.

These participants will be randomized in a 1:1:1 ratio to one of three maintenance groups:

- guselkumab 200 mg subcutaneous (SC) q4w
- guselkumab 100 mg SC q8w
- placebo SC

Participants will be allocated to an intervention group using permuted block randomization stratified by the following factors:

- clinical remission status at maintenance baseline (Yes/No) based on the local endoscopy subscore,

- concomitant use of corticosteroids at maintenance baseline (Yes/No), and
- induction treatment (guselkumab 400 mg IV, guselkumab 200 mg IV, placebo crossover to guselkumab 200 mg IV).

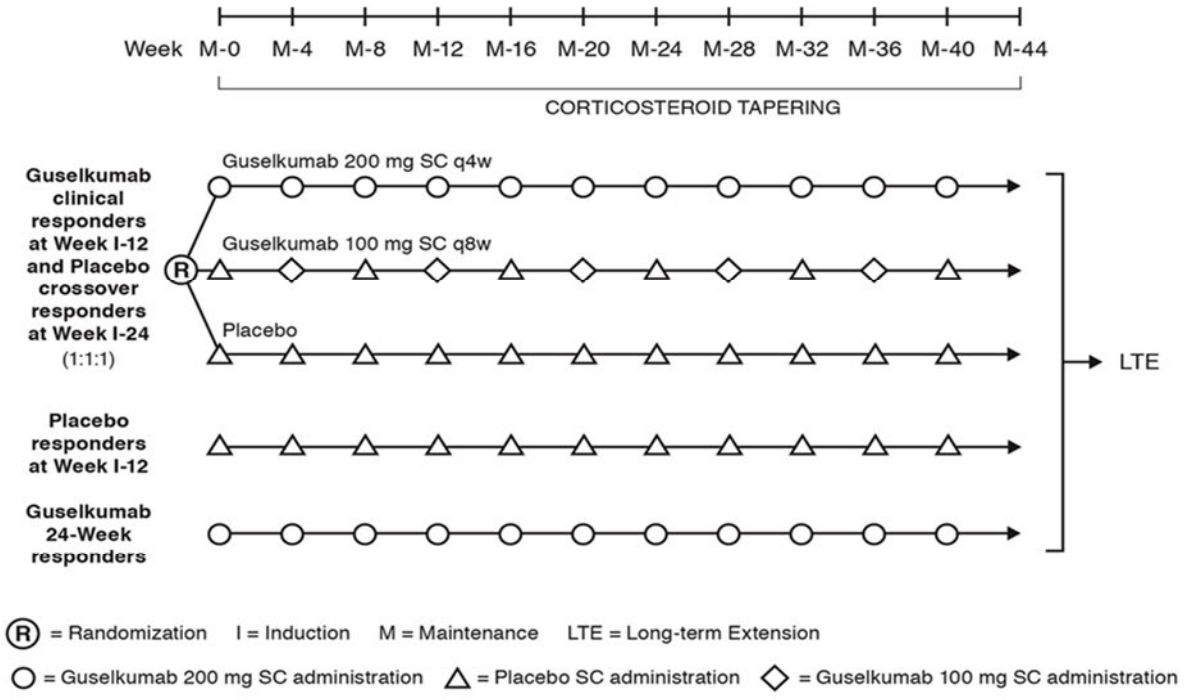
In addition to the aforementioned randomized population, guselkumab 24-week responders (i.e., participants initially randomized to guselkumab IV in an induction study who are not in clinical response at Week I-12 who then receive 3 doses of guselkumab 200 mg SC and achieve clinical response at Week I-24) and placebo responders at Week I-12 from Induction Study 1 or Induction Study 2 will enter the Maintenance Study but will not be randomized. Induction placebo responders will receive placebo SC, and guselkumab 24-week clinical responders will receive guselkumab 200 mg SC q4w.

To maintain the blind, all participants will receive study intervention at all scheduled study intervention visits i.e., M-0 to M-40 every 4 weeks. Participants who discontinue from study intervention administration in maintenance should have a final safety follow-up visit approximately 12 weeks after their last dose of study intervention.

Participants in the randomized population who subsequently lose response at any scheduled visit between Week M-8 and Week M-32 will be eligible to have a one-time dose adjustment (Section 1.3.1.)

With the exception of corticosteroids, which should be tapered, UC-specific medical therapies (i.e., oral 5-ASA compounds, 6-MP, AZA, or MTX) must be maintained at stable doses from Week I-0 through Week M-44 unless the investigator determines that the therapy be discontinued, or the dose reduced because of toxicity or medical necessity. Corticosteroids must be tapered beginning at the M-0 visit for all participants who enter the maintenance study, unless medically not feasible.

Figure 2: Study Schema for Phase 3 Maintenance Study



Guselkumab 24-Week responders: Participants initially randomized to guselkumab who are not in clinical response at Week I-12 who then receive 3 doses of guselkumab 200 mg SC and achieve clinical response at Week I-24.
Placebo crossover responders: Participants initially randomized to placebo who are not in clinical response at Week I-12 who then crossover to guselkumab induction IV dose treatment and achieve clinical response at Week I-24.

2771_v2

A DBL is planned for Week M-44 when all participants in the Maintenance Study have either completed the Week M-44 visit or have terminated study participation before Week M-44.

1.3.1. Dose Adjustment

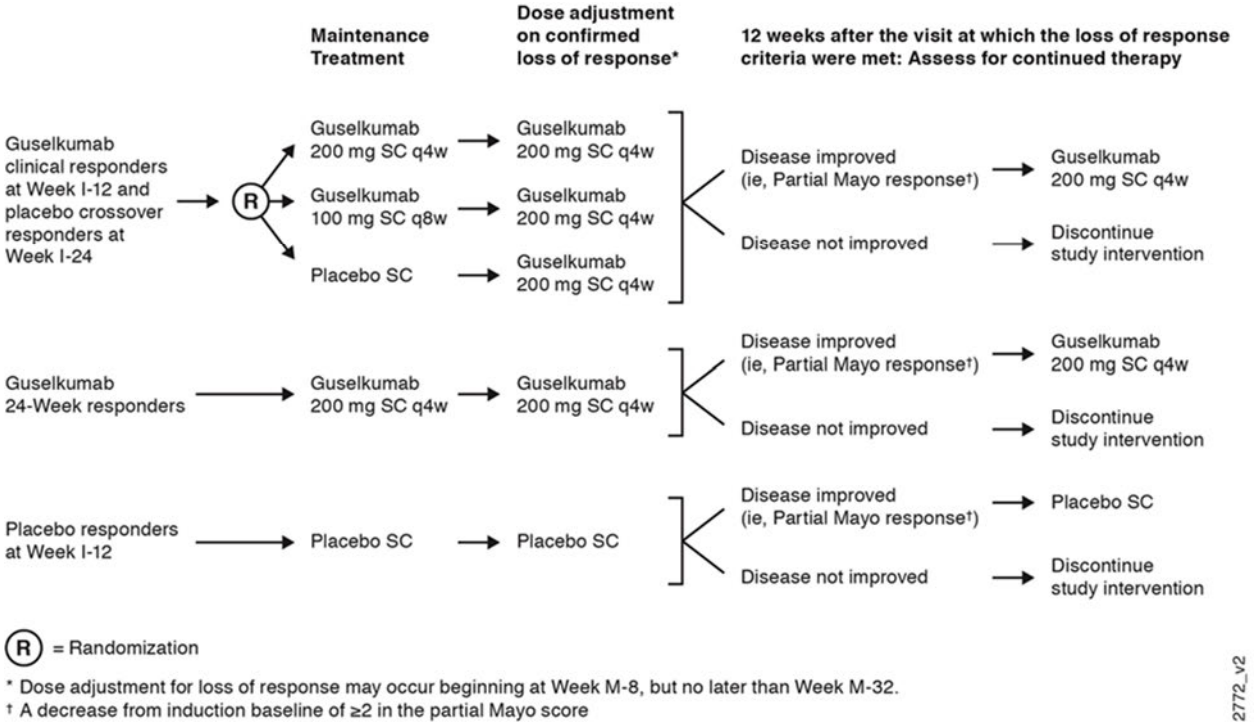
Participants in the randomized population with confirmed loss of clinical response (i.e., no longer satisfies the definition of clinical response as previously defined in Section 1.2) between Week M-8 and Week M-32 will be eligible to receive a single blinded dose adjustment as described below and shown in Figure 3:

- Guselkumab 200 mg SC q4w group: Participants will continue on guselkumab 200 mg SC q4w (i.e., sham dose adjustment).
- Guselkumab 100 mg SC q8w group: Participants will adjust to receive guselkumab 200 mg SC q4w.
- Placebo SC: Participants will adjust to receive guselkumab 200 mg SC q4w.

Guselkumab 24-week responders will continue to receive guselkumab 200 mg SC q4w and placebo responders at Week I-12 will continue to receive SC placebo. These participants are not eligible for a dose adjustment (they will undergo a sham dose adjustment to maintain the blind).

Participants who lost response will be assessed 12 weeks after the visit at which the loss of clinical response criteria are met and dose adjustment occurred. Participants who have not achieved a partial Mayo response (i.e., a decrease from induction baseline of ≥2 in the partial Mayo score) at 12 weeks after loss of clinical response will be discontinued from study intervention administration at that time.

Figure 3: Dose Adjustment Treatment Groups in the Maintenance Study for Participants in Clinical Response Following Induction



1.4. Long-term Extension

The long-term extension (LTE) of the maintenance study begins after the assessments listed for the M-44 visit of the Maintenance Study have been completed and will continue through approximately an additional 4 years of treatment or until the sponsor decides not to pursue an indication in UC, whichever occurs first.

Participants will continue to receive the same study intervention regimen during the LTE that they are receiving at the end of maintenance, with the first dose in the LTE being administered at Week M-44. During the LTE, all participants will be assessed for worsening of disease activity based on the clinical judgment of the investigator. Participants whose UC disease activity worsens during the LTE will be discontinued from study intervention and need to complete the final efficacy and safety visits.

During the LTE, all concomitant medications, including UC-specific medications (except for the prohibited medications listed in protocol Section 6.5.2), may be administered at the discretion of the investigator.

The study blind will be maintained during the LTE until the last participant in the maintenance study has completed the M-44 visit evaluations and the Week M-44 analyses have been completed. Therefore, to maintain the blind, participants will continue to receive study intervention at all visits until that time. After the Maintenance Study is unblinded to the investigative sites, participants receiving placebo will be terminated from study participation, and participants receiving guselkumab will continue to receive guselkumab, but will have their study visits scheduled to coincide with their dose regimen (either q4w or q8w).

The final DBL will occur when all participants have completed the final safety visit or terminated participation. Additional DBLs may occur during the LTE for publications or regulatory purposes.

For more information on the LTE, see the protocol (Sections 1.3.3, 4.1.3.2. and 4.1.3.2.1.).

1.5. Randomization and Blinding

Randomization will be used to minimize bias in the assignment of participants to treatment groups, to increase the likelihood that known and unknown participant attributes (e.g., demographic and baseline characteristics) are evenly balanced across treatment groups, and to enhance the validity of statistical comparisons across treatment groups. Blinded treatment will be used to reduce potential bias during data collection and evaluation of clinical endpoints.

Intervention Allocation

Central randomization will be implemented based on a computer-generated randomization schedule prepared before the study by or under the supervision of the sponsor.

In the Maintenance Study, participants in the randomized population will be randomly assigned to 1 of 3 treatment groups in a 1:1:1 ratio. The randomization will be balanced by using randomly permuted blocks and will be stratified by clinical remission status at maintenance baseline (Yes/No), concomitant use of corticosteroids at maintenance baseline (Yes/No), and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg).

An Interactive Web Response System (IWRS) will assign a unique treatment code, which will dictate the treatment assignment and matching study intervention kit(s) for the participant. The requestor must use his or her own user identification and personal identification number when contacting the IWRS and will then give the relevant participant details to uniquely identify the participant.

Blinding

To maintain the study blind, the study intervention container will have a label containing the study name, study intervention number, and reference number. The study intervention number will be entered in the electronic case report form (eCRF) when the study intervention is dispensed for on-site administration. Each active study intervention and its matching placebo will be identical in appearance.

Data that may potentially unblind the treatment assignment (i.e., study intervention serum concentrations, antibodies to study intervention) will be handled with special care to ensure that the integrity of the blind is maintained and the potential for bias is minimized. This can include making special provisions, such as segregating the data in question from view by the investigators, clinical team, or others as appropriate until the study unblinding.

The results of CRP and fecal calprotectin tests performed by the central laboratory will be blinded to the investigative sites. If an investigative site requests these data, it will be provided to them after the Week M-44 analyses of the Maintenance Study have been completed.

Full sponsor unblinding for CNTO1959UCO3001 will occur following the Week M-44 DBL. In accordance with the predefined Unblinding Plans for Induction Study 1 and Induction Study 2, limited sponsor personnel will be unblinded to the induction data at various DBLs of an induction study; all the sponsor personnel will remain blinded to the assigned maintenance treatment until after the Week M-44 DBL has occurred. Treatment assignment blinding for all 3 studies will be maintained for investigative sites, site monitors, and participants in this protocol until the Week M-44 analyses for the Maintenance Study have been completed.

Identification of sponsor personnel who will have access to the unblinded participant-level data at the time of each analysis will be documented before unblinding.

The investigator will not be provided with randomization codes. The codes will be maintained within the IWRS, which has the functionality to allow the investigator to break the blind for an individual participant, if specific emergency treatment/course of action would be dictated by knowing the treatment status of the participant. In such cases, the investigator may in an emergency determine the identity of the treatment via the IWRS. Participants who have had their treatment assignment unblinded by the investigator will not be eligible to receive further study intervention but should complete evaluations specified in the appropriate schedule of activities (SoA) (Protocol Section 1.3) for participants who discontinue study intervention.

Additionally, a given participant's treatment assignment may be unblinded to the sponsor, the Independent Ethics Committee/Institutional Review Board (IEC/IRB), and site personnel to fulfill regulatory reporting requirements for suspected unexpected serious adverse reactions (SUSARs).

A separate code break procedure will be available for use by the J&J Global Medical Safety group to allow for unblinding of individual participants to comply with specific requests from regulatory or health authorities.

2. STATISTICAL HYPOTHESES

Primary Hypothesis

The primary hypothesis is that guselkumab maintenance therapy is superior to placebo in achieving clinical remission at Week M-44 in participants with moderately to severely active UC who were induced into clinical response with IV guselkumab.

Secondary Hypotheses

Hypotheses for major secondary endpoints in participants with moderately to severely active UC who were induced into **clinical response** with IV guselkumab are listed below:

- Guselkumab maintenance therapy is superior to placebo in achieving symptomatic remission at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving endoscopic healing at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving corticosteroid-free clinical remission at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in maintaining clinical response at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving histologic-endoscopic mucosal healing at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving IBDQ remission at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving fatigue response at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in achieving endoscopic normalization at Week M-44.
- Guselkumab maintenance therapy is superior to placebo in maintaining clinical remission at Week M-44 among participants who were induced into clinical remission with guselkumab IV.

3. SAMPLE SIZE DETERMINATION

Key efficacy analyses in the Maintenance Study will be based on the **Randomized Full Analysis Set** (primary population**)**, which includes all participants with a modified Mayo score of 5 to 9 who are randomized (i.e., Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) in the Maintenance Study who received at least 1 dose of study intervention in this Maintenance Study (Section 4). It is expected that very few participants in the randomized population would not receive study intervention. Therefore, for the purpose of sample size consideration, we will assume the **Randomized Full Analysis Set** is the same as the randomized population with a modified Mayo score of 5 to 9. Unless otherwise stated, the sample size/power calculations in this section refer to the primary population.

A multiplicity-controlled testing procedure, starting with the high guselkumab dose group (200 mg SC q4w), will be used to control the overall Type-I error rate at the 0.05 level (2-sided) over the primary and major secondary endpoints. As such, sample size/power calculations were based on the chi-square test to detect a significant difference between participants receiving SC guselkumab 200 mg q4w and those receiving placebo.

The assumptions for the sample size calculations were based on data from the Phase 3 ustekinumab (anti IL 12/23 mAb) CNTO1275UCO3001 study, which was conducted by the sponsor in a very similar target population, i.e., participants with moderately to severely active UC who had failed or were intolerant to biologic or conventional therapies. In CNTO1275UCO3001, the proportions of participants in clinical remission at Week 44 were 26.3% and 44.9% for placebo and ustekinumab 90 mg SC q8w, respectively, for a treatment difference of 18.6%. Based on these data, the clinical remission at Week M-44 rates are assumed to be 25% for placebo and 45% for each of the guselkumab doses. Given these assumptions, 118 participants in each group (354 participants in total) will provide statistical power of 90% at a significance level of 0.05 (2 sided) for the primary endpoint. However, the actual power can vary depending on the proportions of participants in clinical remission. Examples are presented below in Table 1:

Table 1: Power for Detecting a Treatment Effect Based on Different Proportions of Participants in Clinical Remission at Week M-44 With a Fixed Sample Size of 354 Participants (118 in Each Treatment Group)

Proportion of Participants in Clinical Remission at Week M-44 (%)		Power [a] (%)
Placebo	Guselkumab	
25	50	98
	47	95
	45	90
	42	79
	40	69
a: Based on testing the guselkumab 200 mg SC q4w group versus placebo at α=0.05 (2-sided).		

The number of participants in the primary population of the Maintenance Study will depend on the number of participants from the following 2 groups in the induction studies:

- Group A: participants in clinical response to IV guselkumab at Week I-12 of either Induction Study 1 or Induction Study 2
- Group B: participants who were not in clinical response to IV placebo induction at Week I-12 of either Induction Study 1 or Induction Study 2 but were in clinical response at Week I-24 after receiving IV guselkumab at Weeks I-12, I-16, and I-20.

Based on the assumptions in Induction Study 1 as stated in the protocol, the clinical response rate to guselkumab IV induction is expected to be 60%, thus the 2 induction studies will result in approximately 484 participants with a modified Mayo score of 5 to 9 in the primary population of the Maintenance Study. However, the clinical response rate to guselkumab IV induction could range from 50% to 65%. With 596 participants with a modified Mayo score of 5 to 9 expected to receive guselkumab and 354 participants with a modified Mayo score of 5 to 9 expected to receive placebo at Week I-0 (across Induction Study 1 and Induction Study 2), the number of participants in the primary population of the Maintenance Study could range from 404 to 525 (Table 2). The expected enrollment in the primary population of the Maintenance Study is 484 participants, which is over the required sample size of 354. The targeted number was increased because the maintenance study is intended to power at least 90% for the majority of the major secondary endpoints based on the primary population.

Table 2: Projected Number of Participants in the Primary Population of the Maintenance Study and Associated Power for the Primary Endpoint

Clinical Response Rate to IV Guselkumab Induction	Participants in Group A Entering Maintenance	Participants in Group B Entering Maintenance [a]	Number of Participants in the Primary Population of the Maintenance Study	Power (%)
50%	298	106	404	93
55%	328	117	445	95
60%	357	127	484	97
65%	387	138	525	98

Group A=Participants in clinical response to IV guselkumab induction at Week I-12; Group B=Participants not in clinical response to IV placebo induction at Week I-12 but in clinical response at Week I-24 after receiving induction IV guselkumab at Weeks I-12, I-16, and I-20.
a: The proportion of participants not in clinical response to IV placebo induction at Week I-12 and did not discontinue study intervention is assumed to be 60%.

With 484 participants in the primary population, the power for detecting a treatment difference between the guselkumab 200 mg SC q4w group and the placebo group for the primary endpoint and for each of the major secondary endpoints is shown below in Table 3. The assumptions about the proportion of participants achieving the primary endpoint and each major secondary endpoint have been based on data from the CNTO1275UCO3001 maintenance study.

Table 3: Power for Detecting a Treatment Effect for the Primary Endpoint and Each of the Major Secondary Endpoints With 484 Participants in the primary population (about 161 in Each Treatment Group)

	Proportion of participants achieving the endpoint		
	Placebo	Guselkumab	Power [a] (%)
Primary endpoint			
Clinical remission at Week M-44	25	45	97
Major secondary endpoints			
Symptomatic remission at Week M-44	45	68	99
Endoscopic healing at Week M-44	30	50	96
Clinical response at Week M-44	51	77	99
Histologic-endoscopic mucosal healing at Week M-44	26	46	97
Corticosteroid-free (i.e., not requiring any treatment with corticosteroids for at least 8 weeks prior) clinical remission at Week M-44	25	42	90
IBDQ remission at Week M-44	40	60	95
Fatigue response (≥ 7-point improvement) at Week M-44	26	55	>99
Clinical remission at Week M-44 *among the participants who had achieved clinical remission at maintenance baseline* [b]	36	61	74
Endoscopic normalization at Week M-44.	18	29	65

a: Based on testing the SC guselkumab 200 mg q4w group versus placebo at α=0.05 (2-sided).
b: It is estimated that about 33% of participants in the primary population (53 participants per treatment group) will be in clinical remission at Week M-0.

4. POPULATIONS (ANALYSIS SETS) FOR ANALYSIS

Analysis Sets	Description
Participant Disposition Analysis Sets	
Enrolled Analysis Set	Includes all participants who are assigned to an intervention group in this maintenance study, including both randomized and nonrandomized participants.
Full Analysis Set (FAS)	Includes all participants **with a modified Mayo score of 5 to 9** who receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study, including both randomized and nonrandomized.
Randomized Analysis Set	Includes all participants who are randomized in the study (**regardless of modified Mayo score**).
Randomized (Modified Mayo 5-9) Analysis Set	Includes all participants who were randomized in the study **with a modified Mayo score of 5 to 9.**
Efficacy Analysis Sets: Participants in each efficacy analysis set will be analyzed according to their randomized or assigned study intervention regardless of the study intervention they actually received.	
Randomized Full Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who are randomized (i.e., Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) in this Maintenance Study and receive at least 1 (partial or complete) dose of study intervention in this maintenance study.

Analysis Sets	Description
Nonrandomized Full Analysis Set	Includes all the participants **with a modified Mayo score of 5 to 9** who enter this Maintenance Study but are not randomized (i.e., Guselkumab 24-week clinical responders, and Placebo responders at Week I-12 from Induction Study 1 or Induction Study 2) and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
Efficacy All Randomized and Treated Analysis Set	Includes all participants (**regardless of modified Mayo score**) who are randomized (i.e., Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) in this Maintenance Study and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
Efficacy All Nonrandomized and Treated Analysis Set	Includes all the participants (**regardless of modified Mayo score**) who enter this Maintenance Study but are not randomized (i.e., Guselkumab 24-week clinical responders, and Placebo responders at Week I-12 from Induction Study 1 or Induction Study 2) and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
Dose Adjustment Analysis Set	Includes all the participants in the **Randomized Full Analysis Set** who had a dose adjustment.
Safety Analysis Sets: In general, participants in each safety analysis set will be analyzed according to the study intervention assigned.	
Safety Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study, including both randomized and nonrandomized.
Randomized Safety Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who are randomized in this Maintenance Study (i.e., Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
Nonrandomized Safety Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who enter this Maintenance Study but are not randomized (i.e., Guselkumab 24-week clinical responders, and Placebo responders at Week I-12 from Induction Study 1 or Induction Study 2) and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
All Treated Analysis Set	Includes all participants (regardless of modified Mayo score) who receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study, including both randomized and nonrandomized.
Safety All Randomized and Treated Analysis Set	Includes all participants (regardless of modified Mayo score) who are randomized (i.e., Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) in this Maintenance Study and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
Safety All Nonrandomized and Treated Analysis Set	Includes all the participants (regardless of modified Mayo score) who enter this Maintenance Study but are not randomized (i.e., Guselkumab 24-week clinical responders, and Placebo responders at Week I-12 from Induction Study 1 or Induction Study 2) and receive at least 1 (partial or complete) dose of study intervention in this Maintenance Study.
PK Analysis Set: Participants in each PK analysis set will be analyzed according to the study intervention assigned.	
PK Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who receive at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo) and receive at least 1 dose of guselkumab in an induction study and have at least 1 valid blood sample for PK analysis after their first dose of study intervention in this Maintenance Study.

Analysis Sets	Description
Randomized PK Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who are randomized in the Maintenance Study (i.e. Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) and received at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo), and receive at least 1 dose of guselkumab in an induction study and have at least 1 valid blood sample for PK analysis after their first dose of study intervention in this Maintenance Study.
PK All Treated Analysis Set	Includes all participants (**regardless of modified Mayo score**) who receive at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo) and receive at least 1 dose of guselkumab in an induction study and have at least 1 valid blood sample for PK analysis after their first dose of study intervention in this Maintenance Study.
Immunogenicity Analysis Set: In general, participants in each immunogenicity analysis set will be analyzed according to the study intervention assigned.	
Immunogenicity Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who receive at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo) and receive at least 1 dose of guselkumab in an induction study and have appropriate samples for detection of antibodies to guselkumab (i.e., with at least 1 sample obtained after their first dose of guselkumab in an induction study).
Randomized Immunogenicity Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who are randomized in this Maintenance Study (i.e. Guselkumab clinical responders at Week I-12 and Placebo crossover responders at Week I-24 in an Induction Study) and who receive at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo), and receive at least 1 dose of guselkumab in an induction study, and have appropriate samples for detection of antibodies to guselkumab (i.e., participants with at least 1 sample obtained after their first dose of guselkumab in an induction study).
Continuous Guselkumab Immunogenicity Analysis Set	Includes all participants **with a modified Mayo score of 5 to 9** who receive guselkumab in an induction study and continue on guselkumab in maintenance (comprising guselkumab clinical responders at Week I-12 and placebo crossover responders at Week I-24 who are randomized to guselkumab in maintenance, and guselkumab 24-week clinical responders who continue to receive guselkumab in maintenance), and have appropriate samples for detection of antibodies to guselkumab (i.e., participants with at least 1 sample obtained after their first dose of guselkumab in an induction study).
Immunogenicity All Treated Analysis Set	Includes all participants (**regardless of modified Mayo score**) who receive at least 1 dose (partial or complete) of study intervention in this Maintenance Study (either guselkumab or placebo) and receive at least 1 dose of guselkumab in an induction study and have appropriate samples for detection of antibodies to guselkumab (i.e., with at least 1 sample obtained after their first dose of guselkumab in an induction study).

5. STATISTICAL ANALYSES

5.1. General Considerations

Descriptive statistics (i.e., N, mean, median, standard deviation (SD), interquartile (IQ) range, minimum, and maximum) will be used to summarize continuous variables. Counts and percentages will be used to summarize categorical variables. Graphical data displays (e.g., line plots) may also be used to summarize the data.

Analyses suitable for categorical data (e.g., chi-square tests, Cochran-Mantel-Haenszel tests, or logistic regression, as appropriate) will be used to compare the proportions of participants achieving selected endpoints (e.g., clinical response). In cases of rare events, Fisher's exact test will be used for treatment comparisons. Continuous response parameters measured at more than one postbaseline visit will be compared using Mixed-Effect Model Repeated Measure (MMRM) model (unless otherwise specified). If the normality assumption is in question, an appropriate transformation may be implemented before fitting MMRM model. Continuous response parameters measured at only one post-baseline visit will be compared using an ANOVA or ANCOVA, unless otherwise specified. The time to event endpoints will be compared using a stratified log-rank test or a log-rank test. The association between 2 categorical variables will be assessed based on a chi-square test and the association between a categorical variable and a continuous variable will be assessed based on a t-test.

A multiplicity-controlled testing procedure to control the Type-I error at a 2-sided 0.05 significance level over the primary and major secondary endpoints will be used (see Section 5.4.1.3.).

5.1.1. Visit Windows

Except for the early termination and unscheduled visits, actual scheduled visits will be used for over time summaries and listings with no visit windows applied.

5.1.2. Study Day and Relative Day

Study Day 1 refers to the day of the first study intervention administration of this Maintenance Study. All efficacy and safety assessments at all visits will be assigned a day relative to this day.

Study day for a visit is defined as:

- Visit date - (date of Study Day 1) +1, if visit date is ≥ date of Study Day 1
- Visit date - date of Study Day 1, if visit date < date of Study Day 1

There is no 'Study Day 0'.

5.1.3. Induction Baseline Definition

The induction baseline is defined as the closest non-missing value on or prior to the induction reference start date.

5.1.4. Maintenance Baseline Definition

The maintenance baseline is defined as the closest non-missing value on or prior to the maintenance reference start date which is also post-induction baseline.

5.2. Participant Disposition

The number of participants in the following disposition categories through M-44 will be summarized by treatment group and overall based on the **Full Analysis Set** (including both randomized and nonrandomized participants**)**, the **All Treated Analysis Set**, and the **Dose Adjustment Analysis Set**:

- Participants who received study intervention
- Participants who discontinued study intervention
    - Reasons for discontinuation of study intervention (including COVID-19 and the regional crisis)
- Participants who terminated study prematurely
    - Reasons for termination of study (including COVID-19 and the regional crisis)

Listings of participants based on the **All Treated Analysis Set** will be provided for the following categories:

- Participants who discontinued study intervention
- Participants who terminated study prematurely

In addition, the number and percentage of participants who have a dose adjustment over time will be summarized for participants in the **Randomized Full Analysis Set**.

5.3. Primary Endpoint Analysis

The primary endpoint is clinical remission at Week M-44.

5.3.1. Definition of Endpoint

Clinical remission based on modified Mayo score components: A Mayo stool frequency subscore of 0 or 1, a Mayo rectal bleeding subscore of 0, and a Mayo endoscopy subscore of 0 or 1 with no friability present on the endoscopy, where the stool frequency subscore has not increased from induction baseline.

5.3.1.1. Mayo Score, Partial Mayo Score, and Modified Mayo Score

The Mayo score (Protocol Section 8.1.1.) was developed from the criteria of Truelove and Witts[1] for mild, moderate, and severe UC and from the criteria of Baron et al[2] for grading endoscopic appearance.

The Mayo score consists of the following 4 subscores:

- Stool frequency
- Rectal bleeding
- Findings of endoscopy
- Physician's global assessment (PGA)

Each subscore is rated on a scale from 0 to 3, indicating normal to severe activity, as defined in Attachment 1.

The full **Mayo score** is calculated as the sum of 4 subscores (stool frequency, rectal bleeding, PGA, and endoscopy findings) and ranges from 0 to 12 points. A score of 3 to 5 points indicates mildly active disease, a score of 6 to 10 points indicates moderately active disease, and a score of 11 to 12 points indicates severely active disease.

The **partial Mayo score**, which is the Mayo score without taking into account the findings of endoscopy, is calculated as the sum of stool frequency, rectal bleeding, and PGA subscores, and may take on values from 0 to 9.

The **modified Mayo score**, which is the Mayo score without the PGA subscore, is calculated as the sum of the stool frequency, rectal bleeding, and endoscopy subscores, and may take on values from 0 to 9.

Due to the requirement for endoscopy findings for the evaluation of the Mayo score, it is not feasible to evaluate the Mayo score at each scheduled visit in the study. Therefore, the Mayo score will be evaluated at Week M-44 and at the time when loss of clinical response needs to be confirmed following a clinical flare, and the partial Mayo score will be evaluated at the other study visits. Note that the modified Mayo score is not a separate evaluation from the Mayo score as the latter contains all the components that are needed for the calculation of the modified Mayo score.

Mayo Stool Frequency and Rectal Bleeding Subscores

The eCRF captures seven days of rectal bleeding data and the number of stools per day prior to each visit at which the Mayo score or partial Mayo score is collected. Data from 3 of these 7 days are used to calculate the Mayo stool frequency and rectal bleeding subscores; sites are instructed to check the boxes next to the 3 days which are used (see below for information on what days the sites are instructed to use).

The Mayo rectal bleeding subscore is calculated using the average rectal bleeding number for the three days based on the criteria in Attachment 1.

The Mayo stool frequency subscore is calculated as follows: the absolute stool number is the average of the daily stool number over the three days; at the screening visit, each person indicates the number of stools he/she passed in a 24-hour period when in remission or before his/her UC diagnosis; the stool frequency subscore will be calculated based on the criteria in Attachment 1 by subtracting the number of stools when in remission or prior to UC from the absolute stool number.

Instructions on which 3 days to use in the calculation of the Mayo rectal bleeding and stool frequency subscores: Sites are directed to use the most recent 3 consecutive days within the 7 days prior to the visit and are directed to exclude the following:

- The day medications were taken for constipation, diarrhea or irregularity
- The day(s) of a procedure or preparation for procedure (e.g., enema, other laxatives, or clear liquid diet) that would affect stool frequency and/or blood content of the stool
- The 48 hours after the use of antimotility agents (i.e., diphenoxylate hydrochloride with atropine sulfate, loperamide or other opioids)

  Note: For participants maintained on a chronic stable dose of antimotility agents throughout the study, the days on which these agents are taken will not be excluded from consideration in calculating the Mayo score
- The 48 hours immediately following a colonoscopy

If three consecutive days are not available, the sites are instructed to choose two consecutive days and the closest nonconsecutive day. If two consecutive days are not available, then three nonconsecutive days closest to the visit should be chosen. If 3 days (within the 7 days prior to the indicated visit) that meet the criteria defined above are not available, then the absolute stool number, stool frequency subscore and rectal bleeding subscore cannot be calculated and will be missing in the eCRF.

Mayo Endoscopy Subscore

The endoscopic findings will be based on the criteria of the Mayo endoscopy subscore described in Attachment 1. The endoscopic findings will be assessed by the investigator (i.e., local endoscopist) during the endoscopy procedure and by a central reader reviewing a video of the endoscopy. The endoscopy may be either a colonoscopy or sigmoidoscopy. A full colonoscopy will replace a sigmoidoscopy if screening for polyps or dysplasia is required. The central reader will also perform a friability assessment (Yes/No) for each endoscopy.

Participant eligibility at baseline will be based on the final reported endoscopic subscore as determined by the following process:

- If the local endoscopist and the central reader agree on the endoscopic subscore, the agreed score will be the final reported endoscopic subscore.
- If there is a discrepancy between the local endoscopist and the central reader subscores, the video endoscopy will be submitted to a second central reader (designated for adjudication), who is blinded to the scores of the local and the first central reader. The median score of the 3 completed reads (i.e., local read, central read 1, and central read 2 designated for adjudication) will be the final reported endoscopic subscore.

Further details are provided in the Imaging Charter.

Unless otherwise specified, the analysis of endpoints that include the Mayo endoscopy subscore will be based on the final reported endoscopic subscore. If the final reported endoscopic subscore is not available, the corresponding central endoscopy score (central read #1) will be used, if available. If the central endoscopy score (central read #1) is also missing, then the

local endoscopy score will be used, if available. If the local endoscopy score is also not available, then the endoscopy subscore for the analysis will be left missing.

Mayo Physician's Global Assessment Subscore (PGA)

The PGA acknowledges the 3 other Mayo subscores, the patient's recall of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the patient's performance status.

5.3.2. Primary Estimand (Estimand 1)

The primary estimand, i.e., a precise definition of the primary targeted treatment effect is defined by the following 5 attributes:

Randomized Maintenance Treatment by Week M-44:

Experimental:

- Guselkumab 200 mg SC q4w
- Guselkumab 100 mg SC q8w

Control:

- Placebo SC

Population: Patients 18 years or older with moderately to severely active UC as reflected in the inclusion/exclusion criteria (Protocol Section 5) along with a requirement of a modified Mayo score of 5 to 9 who were induced into clinical response with guselkumab.

Variable: Clinical remission at Week M-44 (Section 5.3.1.) where participants are considered to have achieved clinical remission if they fulfill the clinical remission criteria based on modified Mayo score components (Section 5.3.1.1.), and do not experience intercurrent events in categories 1-4 and 6 (defined below) prior to the Week M-44 visit.

Intercurrent Events and Corresponding Strategies:

The following are the intercurrent events for this study:

1. An ostomy or colectomy (partial or total)
2. Have a dose adjustment (including a sham dose adjustment; Section 1.3.1)
3. Prohibited change in UC medication (described in Attachment 2)
4. Discontinuation of study intervention due to lack of efficacy or an AE of worsening of UC
5. Discontinuation of study intervention due to major disruption, including COVID-19 related reasons (excluding COVID-19 infection) or regional crisis in Russia and Ukraine
6. Discontinuation of study intervention due to reasons other than those in ICEs 4 and 5

Intercurrent events (ICEs) in categories 1-4 and 6 will be handled with the composite strategy as reflected in the variable definition. ICE category 5 will be handled by the Treatment Policy Strategy. Note that the application of ICE categories 1-3 overrides that of ICE 5. This means that a participant with ICE in categories 1-3 will be considered as not to have achieved the variable, regardless of whether the participant has had an ICE 5. For participants experiencing ICE 5, their observed clinical remission status (if available) at Week M-44 will be used.

Population-level summary:

The difference in the proportion of participants achieving the variable (as defined above for this estimand) between each guselkumab group and the placebo group.

Note: This estimand acknowledges that having an intercurrent event in categories 1-4 and 6 is an unfavorable outcome.

5.3.2.1. Analysis Methods

5.3.2.1.1. Main Estimator (Analysis) for the Primary Estimand

The primary endpoint is clinical remission at Week M-44 (as defined in Section 5.3.1. above).

The primary endpoint will be analyzed based on the Primary Estimand, Estimand 1 (Section 5.3.2.). After accounting for the ICE strategies, participants who are missing any or all of the Mayo subscores that comprise the primary endpoint at Week M-44 will be considered not to be in clinical remission at Week M-44 (i.e., nonresponder imputation). For participants experiencing ICE 5, their observed clinical remission status (if available) at Week M-44 will be used.

In the primary analysis, data from all participants in the **Randomized Full Analysis Set** (Section 4) will be analyzed according to the randomized study intervention regardless of the study intervention they actually received.

Summaries of the proportion of participants in clinical remission at Week M-44 (as well as the associated 95% confidence interval) by treatment group, the adjusted treatment difference (with Cochran-Mantel-Haenszel weight) between each guselkumab treatment group and the placebo group, as well as the associated 95% confidence interval will be presented.

For testing of the primary endpoint, the efficacy of each guselkumab group versus placebo will be compared. For all statistical comparisons of the primary endpoint, a Cochran-Mantel-Haenszel (CMH) test (2-sided) stratified by clinical remission status at maintenance baseline (Yes/No), and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg) will be used.

A multiplicity-controlled testing procedure, starting with the high guselkumab dose group (200 mg SC q4w), will be used to control the overall Type-I error rate at the 0.05 level (2-sided) over the primary and major secondary endpoints. The study will be considered positive if the test involving

the high maintenance dose group shows a statistically significant difference versus placebo for the primary endpoint of clinical remission at Week M-44 (see Section 5.4.1.3).

Testing Procedure:

The multiple testing procedure to control the Type 1 error will be different for the United States and the countries outside the United States, as described below.

Type I error control for Rest of World (i.e., countries outside the United States): A fixed-sequence testing procedure will be used to control the overall Type I error rate at the 0.05 level for the primary endpoint. Specifically, the high maintenance dose group (200 mg SC q4w) will be considered significant if its p-value vs placebo is < 0.05. The low maintenance dose group (100 mg q8w) will be significant if the p-value vs placebo for both high and low maintenance dose groups are < 0.05.

Type I error control in the United States (US-specific testing procedure): A fixed-sequence testing procedure will be employed for the United States to strongly control the overall Type 1 error rate at the 0.05 level across the primary and all the major secondary endpoints (except for IBDQ remission, which is not considered a major secondary endpoint in the US-specific testing procedure) and across the 2 guselkumab doses, starting with the high maintenance dose group (200 mg SC q4w) of the primary endpoint. The exact testing procedure is detailed in Section 5.4.1.3.

5.3.2.1.2. Subgroup Analyses

Subgroup analyses will be performed based on demographic and UC disease characteristics, and concomitant UC medication use and history of UC-related medications (including ADT-Failure status), all at Week 0 of the induction study, as well as maintenance stratification factors and UC clinical disease characteristics at Week 0 of the maintenance study, specified in Section 5.10.6. Note that, for subgroup analyses, the analysis sets are the individual subgroups of the **Randomized Full Analysis Set**. For each of these subgroups, the rate (risk) difference of each guselkumab group vs placebo and the associated 95% confidence interval will be provided. The rate (risk) difference and confidence intervals will be provided based on the Cochran-Mantel-Haenszel weight that includes factors for clinical remission status at maintenance baseline (Yes/No), and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg). For the subgroup analyses based on the maintenance stratification factors, the corresponding factor will not be included in the model. The primary estimand will be used for these subgroup analyses and the missing data rule (i.e., missing values will be imputed as non-responder) used for the primary estimand will be applied.

5.3.2.2. Sensitivity Analyses

5.3.2.2.1. Sensitivity Analyses 1: Tipping Point

A sensitivity analysis will be performed using a tipping point analysis with Bernoulli draws to impute missing clinical remission status at Week M-44 after the intercurrent event rules have been applied, when the number of participants with missing values (after accounting for the ICE

strategies) is ≥10 in any treatment group. This tipping point analysis involves the following distinct steps:

1. Some *p* will be assumed for each treatment group's response rate, which could vary by treatment group, to impute the response status (Yes/No) for participants with a missing response based on a Bernoulli distribution. This will be repeated N times (e.g., 200 times) to generate N multiple imputations.
2. Each of the resulting data sets will be analyzed based on the CMH test proposed for the primary analysis.
3. The results (with a Wilson-Hilferty transformation to the CMH statistic) from the imputed data sets will then be combined to produce inferential results based on Rubin's rules.

The analysis will be repeated for a range of values for *p* (for example, 0% to 100% in increments of 10% independently, for both the placebo and the guselkumab groups).

5.3.2.2.2. Sensitivity Analyses 2: Multiple Imputation

A multiple imputation method will be utilized to impute missing Mayo score components pertaining to the primary endpoint after the intercurrent event rules have been applied. This method involves the following distinct steps:

1. Any missing Mayo components pertaining to the primary endpoint at Week M-44 will be imputed N times (e.g., 200) to generate N complete data sets using the fully conditional specification (FCS) method, assuming missing at random (MAR). The following variables will be included in the imputation model: Mayo components pertaining to the primary endpoint at Week I-0 of an induction study and at Week M-0 and Week M-44 of the maintenance study, induction dose factor, and the maintenance treatment group.
2. Each of the N resulting data sets will be analyzed using a logistic regression model with treatment group, induction baseline modified Mayo score, and induction treatment as covariates. The modified Mayo score was used as the baseline disease activity measure as it contains the same Mayo components as those for the definition of clinical remission.
3. The results from the N data sets will be combined to produce inferential results.

5.3.2.2.3. Sensitivity Analyses 3: Exclusion of Participants Whose Data Cannot be Source Data Verified Due to Major Disruption

A sensitivity analysis will be performed based on the **Randomized Full Analysis Set** with the exclusion of participants whose data cannot be source data verified prior to the Week M-44 DBL due to major disruption, including COVID-19 and regional crisis in Russia and Ukraine. This analysis will use the same ICE strategies as those for the primary estimand, and missing data handling rule as those for the primary analysis. Data for participants that cannot be source data verified will be documented prior to the Week M-44 DBL and study unblinding.

5.3.2.3. Supplementary Estimands for the Primary Endpoint

Estimand 2 (Composite Strategy for all ICEs):

The attributes of this supplementary estimand are the same as those for the primary estimand with the exception of the Variable (Endpoint) and the strategies for the intercurrent events, which are described as follows:

Variable (Endpoint): Clinical remission at Week M-44, where participants are considered to have achieved clinical remission if they fulfill the clinical remission criteria based on modified Mayo score components (Section 5.3.1.). Participants who have intercurrent events in categories 1-6 prior to the Week M-44 visit will be considered to not have achieved clinical remission at Week M-44.

Intercurrent Events and corresponding strategies:

For this estimand, the intercurrent events (given below) are addressed with a Composite Strategy.

1. An ostomy or colectomy (partial or total)
2. Have a dose adjustment (including a sham dose adjustment; Section 1.3.1.)
3. Prohibited change in UC medication (described in Attachment 2)
4. Discontinuation of study intervention due to lack of efficacy or an AE of worsening of UC
5. Discontinuation of study intervention due to COVID-19 related reasons (excluding COVID-19 infection) or regional crisis in Russia and Ukraine
6. Discontinuation of study intervention due to reasons other than those in ICEs 4 and 5

Estimand 3 (Alternative Mayo Calculation 1):

The attributes of this supplementary estimand are the same as those for the primary estimand with the exception of the Variable (Endpoint), where clinical remission at Week M-44 based on modified Mayo score components will be derived using an alternative calculation for the Mayo rectal bleeding subscore and stool frequency subscore instead of the methodology based on 3-day diary as detailed in Section 5.3.1. In this estimand, the Mayo rectal bleeding and stool frequency subscores will be calculated either based on eligible diary data (data not excluded from the calculation per Section 5.3.1.1.) collected on 3 consecutive days or, when 3 consecutive days are not available, all eligible diary data collected in a 7-day window prior to a clinical visit. If neither 3 consecutive days nor 4 nonconsecutive days are available, then the Mayo rectal bleeding and stool frequency subscores will not be calculated.

Estimand 4 (Alternative Mayo Calculation 2):

The attributes of this supplementary estimand are the same as those for the primary estimand with the exception of the Variable (Endpoint), where clinical remission at Week M-44 based on modified Mayo score components will be derived using an alternative calculation for the Mayo rectal bleeding subscore and stool frequency subscore instead of the methodology based on 3-day

diary as detailed in Section 5.3.1. In this estimand, the Mayo rectal bleeding and stool frequency subscores will be calculated based on all available diary data collected in the 7-day window prior to a clinical visit, excluding the day of an endoscopy and the day prior to an endoscopy. If neither 3 consecutive days nor 4 non-consecutive days are available, then the Mayo rectal bleeding and stool frequency subscores will not be calculated.

Estimand 5 (Different Population):

The attributes of this supplementary estimand are the same as those for the primary estimand with the exception of Population as defined below, which will include all participants randomized under the inclusion/exclusion criteria:

Population: Patients 18 years or older with moderately to severely active UC as reflected in the inclusion/exclusion criteria (Protocol Section 5), which includes patients with a modified Mayo score of 4 to 9.

5.3.2.3.1. Estimator (Analysis) for the Supplementary Estimands of the Primary Endpoint

For testing of the primary endpoint using the supplementary estimands, the efficacy of each guselkumab group versus placebo will be compared, using the **Randomized Full Analysis Set** for Estimands 2 and 3, and the **Efficacy All Randomized and Treated Analysis Set** for Estimand 4. For all statistical comparisons, a Cochran-Mantel-Haenszel (CMH) test (2-sided) stratified by clinical remission status at maintenance baseline (Yes/No) and induction dose treatment will be used. Summaries of the proportion of participants in clinical remission at Week M-44 as well as the associated 95% confidence interval by treatment group, the adjusted treatment difference (with Cochran-Mantel-Haenszel weight) between each guselkumab treatment group and the placebo group, and the associated 95% confidence interval will be presented.

Missing data rule for Estimands 2-5:

After accounting for the ICE strategies, participants who are missing any or all of the Mayo subscores that comprise the primary endpoint at Week M-44 will be considered not to be in clinical remission at Week M-44 (i.e., nonresponder imputation).

5.4. Major Secondary Endpoints Analysis

5.4.1. Confirmatory Major Secondary Endpoints

The following are the major secondary endpoints, presented in the order in which they will be tested for the global testing procedure (Section 5.4.1.3.):

1. Symptomatic remission at Week M-44.
2. Endoscopic healing at Week M-44.
3. Corticosteroid-free (i.e., not requiring any treatment with corticosteroids for at least 8 weeks prior) clinical remission at Week M-44.
4. Clinical response at Week M-44 (maintenance of clinical response at M-44).

5. Histologic-endoscopic mucosal healing at Week M-44.
6. IBDQ remission at Week M-44.
7. Fatigue response at Week M-44.
8. Clinical remission at Week M-44 (maintenance of clinical remission at M-44) among the participants who had achieved clinical remission at maintenance baseline.
9. Endoscopic normalization at Week M-44.

5.4.1.1. Definition of Endpoints

The following endpoints are defined based on the corresponding scales; certain ICEs will be incorporated in the variable definition of each estimand as appropriate:

Clinical response: A decrease from induction baseline in the modified Mayo score (Section 5.3.1.1) by ≥ 30% and ≥ 2 points, with either a ≥ 1-point decrease from baseline in the rectal bleeding subscore or a rectal bleeding subscore of 0 or 1.

Symptomatic remission: a stool frequency subscore of 0 or 1 and a rectal bleeding subscore of 0, where the stool frequency subscore has not increased from induction baseline.

Histologic healing: neutrophil infiltration in < 5% of crypts, no crypt destruction, and no erosions, ulcerations or granulation tissue according to the Geboes grading system[3] (i.e., Geboes score ≤3.1). (See Attachment 3).

Endoscopic healing: an endoscopy subscore of 0 or 1 with no friability present on the endoscopy.

Endoscopic normalization: An endoscopy subscore of 0 (which requires that no friability is present).

Histologic-endoscopic mucosal healing: achieving a combination of histologic healing and endoscopic healing with no friability present on the endoscopy.

IBDQ remission: total IBDQ score ≥ 170 (Irvine et al, 1994[13]; Higgins et al, 2005[14]).

Fatigue response: a ≥ 7-point improvement from induction baseline in PROMIS-Fatigue short form 7a.

Corticosteroid-free Clinical remission: not requiring any treatment with corticosteroids for at least 8 weeks prior to Week M-44 and also meeting the criteria for clinical remission at Week M-44.

5.4.1.2. Main Estimands for the Major Secondary Endpoints

The attributes and strategies for the ICEs that were used for the primary estimand (Estimand 1) for the primary endpoint analysis (Section 5.3.2.) will also be used for each of the major secondary endpoints.

5.4.1.3. Main Estimators (Analyses) for the Secondary Estimands

The major secondary endpoint analyses will be based on the **Randomized Full Analysis Set**. For all statistical comparisons of the major secondary endpoints (except for clinical remission at Week M-44 among the participants who had achieved clinical remission at maintenance baseline), a CMH test (2-sided) stratified by clinical remission status at maintenance baseline (Yes/No) and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg) will be used. In cases of rare events, Fisher's exact test will be used for treatment comparisons.

For clinical remission at Week M-44 among the participants who had achieved clinical remission at maintenance baseline, a CMH test (2-sided) stratified by induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg) will be used.

Summaries of the proportion of participants achieving each major secondary endpoint by treatment group as well as the associated 95% confidence interval, the adjusted treatment difference (with Cochran-Mantel-Haenszel weight) between each guselkumab treatment group and the placebo group, as well as the associated 95% confidence interval will be presented.

Missing Data Rules for the Secondary Estimands:

After accounting for the ICE strategies, any missing data for the major secondary endpoints will be handled with nonresponder imputation. In particular, the following rules will be used:

- Participants who are missing any or all of the 3 Mayo subscores that comprise the modified Mayo score will be considered not to have achieved corticosteroid-free clinical remission, maintenance of clinical response, or maintenance of clinical remission.

- Participants who have a missing endoscopy subscore at Week M-44 will be considered not to have achieved endoscopic healing or histologic-endoscopic mucosal healing or endoscopic normalization at Week M-44.

- Participants who are missing any or all of the components in the Geboes grading system pertaining to histologic healing endpoint will be considered not to have achieved histologic-endoscopic mucosal healing.

- Participants who are missing either stool frequency or rectal bleeding subscores at Week M-44 will be considered not to have achieved symptomatic remission at Week M-44.

- Participants who have a missing IBDQ total score at Week M-44 will be considered not to have achieved IBDQ remission at Week M-44.

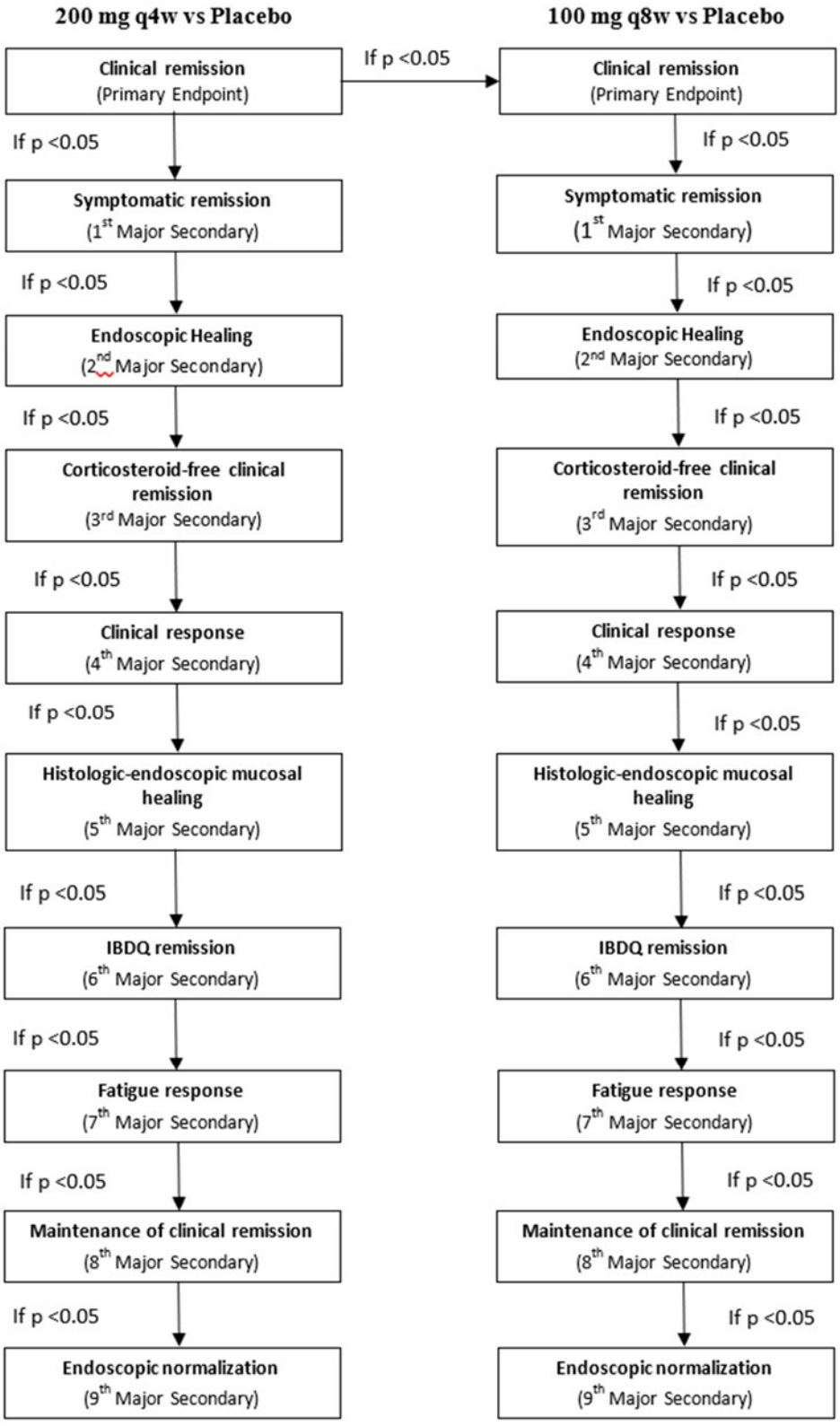
- Participants who have any missing PROMIS-Fatigue short form 7a item at Week M-44 will be considered not to have achieved fatigue response at Week M-44.

Testing Procedure:

The United States and the global regions will employ a different multiple testing strategy as described below to control the overall Type I error.

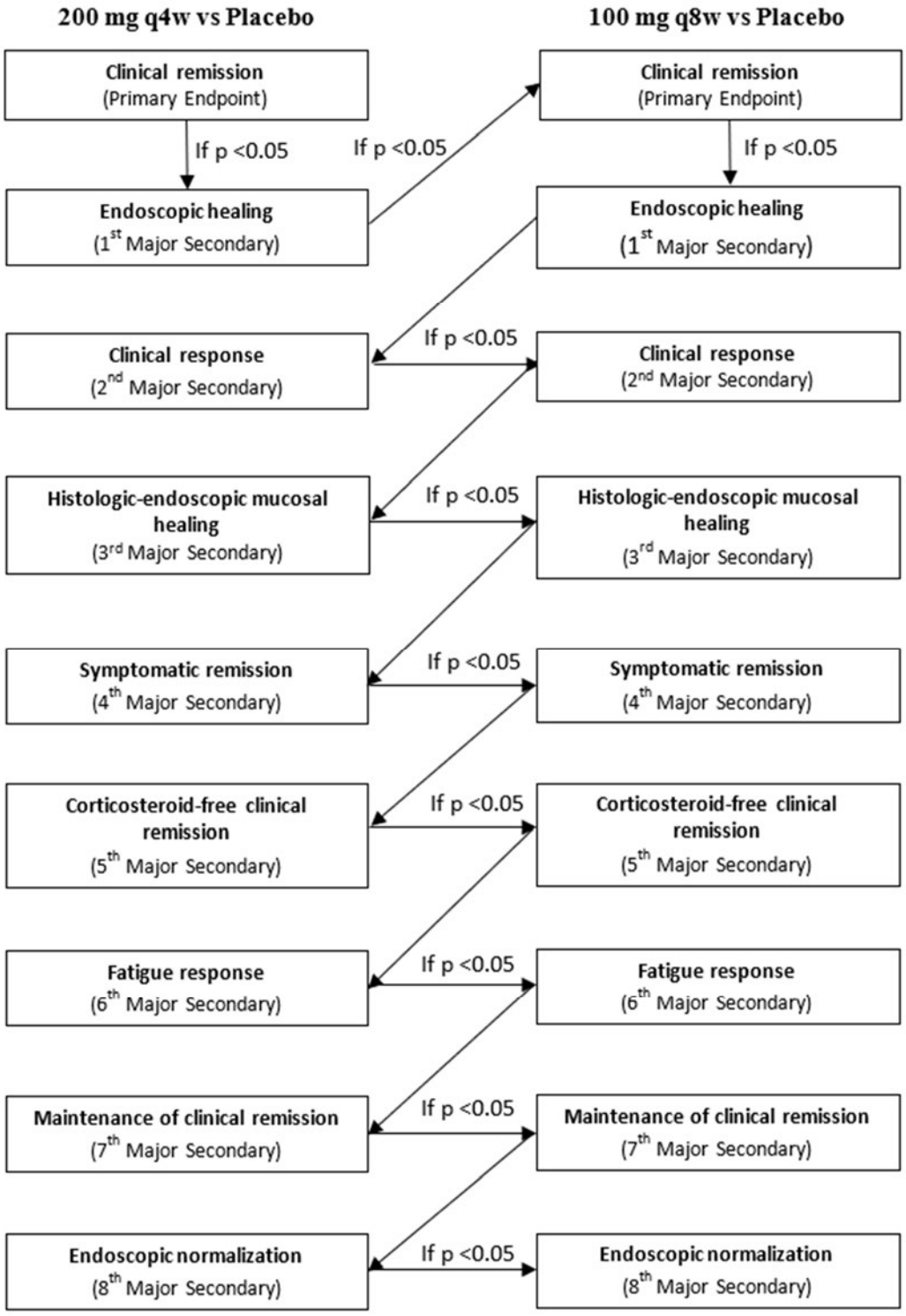
Type I error control in countries outside the United States (global testing procedure): A hierarchical testing procedure as shown in Figure 4a will be employed to control the overall Type 1 error rate over the 9 major secondary efficacy analyses at the (2-sided) 0.05 significance level **within** a guselkumab dose group, for the doses that test positive for the primary endpoint. A major secondary endpoint for a guselkumab dose group will be considered significant only if all the previous endpoints in the hierarchy and the current endpoint test positive at the 2-sided 0.05 level of significance. If an endpoint is not significant, all subsequent tests in the hierarchy will be considered not to be significant.

Figure 4a: Global Testing Procedure for the Primary and Major Secondary Endpoints at Week M-44



Type I error control in the United States (US-specific testing procedure): A hierarchical testing procedure as shown in Figure 4b will be employed for the United States to strongly control the overall Type 1 error rate at the 0.05 level across the primary and all major secondary endpoints and across the 2 guselkumab doses. An endpoint will be considered significant only if all the previous endpoints in the hierarchy and the current endpoint test positive at the 2-sided 0.05 level of significance. If an endpoint is not significant, all subsequent tests in the hierarchy will be considered not to be significant. Note that the rank order of the major secondary endpoints is different for the US-specific and global testing procedures due to regional preferences. In addition, the major secondary endpoint of IBDQ remission at Week M-44 will not be included in the US-specific testing procedure since the FDA does not accept the IBDQ endpoint to support future labeling claims due to consideration that the IBDQ is not a validated instrument.

Figure 4b: US Specific Testing Procedure for the Primary and Major Secondary Endpoints at Week M-44



5.4.1.4. Subgroup Analyses

The consistency of treatment effect for the major secondary endpoints will be evaluated for the subgroups defined in Section 5.10.6. for the major secondary endpoints. Note that, for subgroup analyses, the analysis sets are the individual subgroups of the **Randomized Full Analysis Set**. For each of these subgroups, the rate (risk) difference of each guselkumab group vs placebo and the associated 95% confidence interval will be provided. The rate (risk) difference and confidence intervals will be provided based on the Cochran-Mantel-Haenszel weight that includes factors for clinical remission status at maintenance baseline (Yes/No), and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg). For the subgroup analyses based on the maintenance stratification factors, the corresponding factor will not be included in the model. The main estimand for the major secondary endpoints will be used for these subgroup analyses and the missing data rule (i.e., missing values will be imputed as nonresponder) used for the secondary estimands will be applied.

5.4.1.5. Sensitivity Analysis

5.4.1.5.1. Sensitivity Analysis 4: Exclusion of participants Whose Data Cannot be Source Data Verified Due to Major Disruption

A sensitivity analysis will be performed for each major secondary endpoint based on the **Randomized Full Analysis Set** with the exclusion of participants whose data cannot be source data verified due to major disruption, including COVID-19 and regional crisis. This analysis will use the same ICE strategies as those for the main estimands (Section 5.4.1.2.), and missing data handling rule as those for the main analyses for the major secondary endpoints (Section 5.4.1.3.). Data for participants that cannot be source data verified will be documented prior to the Week M-44 DBL and study unblinding.

5.4.2. Supplementary Estimands for the Major Secondary Endpoints

Estimand 6 (Composite Strategy for all ICEs):

The attributes and strategies for the ICEs that were used for the Supplementary Estimand 2 for the primary endpoint (Section 5.3.2.2.2.) in which all ICEs are handled by the composite strategy, will also be used for each of the major secondary endpoints.

Estimand 7 (Alternative Mayo Calculation 1):

This estimand applies to symptomatic remission at Week M-44, corticosteroid-free clinical remission at Week M-44, clinical response at Week M-44 (maintenance of clinical response at M-44), and clinical remission at Week M-44 among the participants who had achieved clinical remission at maintenance baseline (maintenance of clinical remission at Week M-44).

The attributes of this supplementary estimand are the same as those for the main estimand with the exception of the Variable (Endpoint), where symptomatic remission, clinical response, and clinical remission criteria will be derived using an alternative calculation for the Mayo rectal bleeding subscore and stool frequency subscore instead of the methodology based on 3-day diary as detailed in Section 5.3.1.1. In this estimand, the Mayo rectal bleeding and stool frequency subscores will

be calculated either based on eligible diary data collected on 3 consecutive days or, when 3 consecutive days are not available, all eligible diary data collected in a 7-day window prior to a clinical visit. If neither 3 consecutive days nor 4 nonconsecutive days are available, then the Mayo rectal bleeding and stool frequency subscores will not be calculated.

Estimand 8 (Alternative Mayo Calculation 2):

This estimand applies to symptomatic remission at Week M-44, corticosteroid-free clinical remission at Week M-44, clinical response at Week M-44 (maintenance of clinical response at Week M-44), and clinical remission at Week M-44 among the participants who had achieved clinical remission at maintenance baseline (maintenance of clinical remission at M-44).

The attributes of this supplementary estimand are the same as those for the primary estimand with the exception of the Variable (Endpoint), where symptomatic remission, clinical response, and clinical remission criteria will be derived using an alternative calculation for the Mayo rectal bleeding subscore and stool frequency subscore instead of the methodology based on 3-day diary as detailed in Section 5.3.1.1. In this estimand, the Mayo rectal bleeding and stool frequency subscores will be calculated based on all available diary data collected in the 7-day window prior to a clinical visit, excluding the day of an endoscopy and the day prior to an endoscopy. If neither 3 consecutive days nor 4 non-consecutive days are available, then the Mayo rectal bleeding and stool frequency subscores will not be calculated.

Estimand 9 (Different Population):

The attributes of this supplementary estimand are the same as those for the main estimand with the exception of Population, as defined below, which will include all participants enrolled under the inclusion/exclusion criteria.

Population: Patients 18 years or older with moderately to severely active UC as reflected in the inclusion/exclusion criteria (Protocol Section 5), which includes patients with a modified Mayo score of 4 to 9.

5.4.2.1. Estimator (Analyses) for the Supplementary Estimands

The same analysis methods for the main estimands described in Section 5.4.1.3. will be used for the supplementary estimands of the major secondary endpoints, using the **Randomized Full Analysis Set** for Estimands 5 and 6, and the **Efficacy All Randomized and Treated Analysis Set** for Estimand 7.

Missing Data Rules for the Supplementary Estimands

The same missing data rules for the main estimands described in Section 5.4.1.3. will be used for the supplementary estimands of the major secondary endpoints.

5.5. Other Endpoints Analysis

In addition to the primary and major secondary endpoints, other endpoints related to disease status, HRQoL outcomes (including fatigue), inflammatory biomarkers, and health economics will be analyzed. This section lists the other endpoints, followed by their definitions and analysis methods.

These endpoints will be summarized and compared between each of the guselkumab groups and placebo.

The following endpoints are defined based on the corresponding scales; certain ICEs will be incorporated in the variable definition of each estimand as appropriate:

Clinical Endpoints

- Clinical response (Alternative Definition 1) at Week M-44, defined as a decrease from induction baseline in the full Mayo score by ≥ 30% and ≥ 3 points, with either a ≥ 1-point decrease from baseline in the rectal bleeding subscore or a rectal bleeding subscore of 0 or 1.
- Clinical response (Alternative Definition 2) at Week M-44, defined as a decrease from induction baseline in the modified Mayo score by ≥ 35% (instead of 30%) and ≥ 2 points, with either a ≥ 1-point decrease from baseline in the rectal bleeding subscore or a rectal bleeding subscore of 0 or 1.
- Clinical remission (Alternative Definition 1) at Week M-44, defined as a full Mayo score ≤ 2 points, with no individual subscore >1 point.
- Clinical remission (Alternative Definition 2) at Week M-44, defined as a full Mayo score ≤ 2 points, with a rectal bleeding subscore of 0 and no individual subscore > 1 point (i.e., clinical remission [Alternative Definition 1] with a rectal bleeding subscore of 0).
- Clinical remission (Alternative Definition 3) at Week M-44, based on a stool frequency subscore of 0, a rectal bleeding subscore of 0, and an endoscopy subscore of 0 or 1, with no friability present on the endoscopy.
- Clinical remission at Week M-44 by clinical remission status at maintenance baseline (yes/no).
- Symptomatic remission at each visit through Week M-44.
- Durable symptomatic remission through Week M-44.
- Symptomatic remission at each visit through Week M-44 among the participants who had achieved symptomatic remission at maintenance baseline.
- Symptomatic remission at Week M-44 by symptomatic remission status at maintenance baseline (yes/no).
- Deep symptomatic remission through Week M-44.
- Endoscopic healing at Week M-44 by endoscopic healing status at maintenance baseline (yes/no).
- Histologic healing (based on Geboes Grading System[3]) at Week M-44.
- Histologic healing (based on Geboes Grading System[3]) at Week M-44 among those without histologic healing at induction baseline.
- Histologic remission based on the Geboes score[3] at Week M-44.
- Histologic remission based on the Nancy Histologic Index[4] at Week M-44.

- Histologic remission based on the Robarts Histopathology Score[5] (equivalent to histologic remission based on Geboes score) at Week M-44.
- Histologic-endoscopic mucosal healing (Alternative definition 1) at Week M-44.
- Histologic-endoscopic mucosal healing (Alternative definition 2) at Week M-44.
- Deep histologic-endoscopic mucosal healing at Week M-44.
- A combination of symptomatic remission at Week M-44 and histologic-endoscopic mucosal healing at Week M-44 and fecal calprotectin concentration ≤ 250 mg/kg at Week M-44.
- A combination of symptomatic remission at Week M-44 and deep histologic-endoscopic mucosal healing at Week M-44 and fecal calprotectin concentration ≤ 250 mg/kg at Week M-44.
- Change from induction and maintenance baseline in Geboes total score at Week M-44.
- Change from induction and maintenance baseline in Geboes high activity subscore at Week M-44.
- Change from induction and maintenance baseline in Geboes low activity subscore at Week M-44.
- Change from induction and maintenance baseline in Robarts Histopathology Index at Week M-44.
- Change from induction and maintenance baseline in Nancy Histological Index at Week M-44.
- Change from induction baseline in the modified Mayo score at Week M-44.
- Change from maintenance baseline in the modified Mayo score at Week M-44.
- Change from induction baseline in the partial Mayo score at each visit through Week M-44.
- Change from maintenance baseline in the partial Mayo score at each visit through Week M-44.
- Partial Mayo response at each visit through Week M-44.
- Time to first loss of partial Mayo response through Week M-44 among the participants who had achieved partial Mayo response at maintenance baseline.
- Time to first loss of symptomatic remission through Week M-44 among the participants who had achieved symptomatic remission at maintenance baseline.
- Change from induction baseline in the stool frequency and rectal bleeding subscores at each visit through Week M-44.
- Change from maintenance baseline in the stool frequency and rectal bleeding subscores at each visit through Week M-44.
- Change from induction baseline in the full Mayo score at Week M-44.
- Change from maintenance baseline in the full Mayo score at Week M-44.
- Change from induction baseline in absolute stool number at each visit through Week M-44.

- Change from maintenance baseline in absolute stool number at each visit through Week M-44.
- Absolute stool number ≤ 3 at each visit through Week M-44.
- Rectal bleeding subscore of 0 at each visit through Week M-44.
- Stool frequency subscore of 0 or 1 at each visit through Week M-44.
- Change from induction baseline in the endoscopy subscore at Week M-0 and Week M-44.
- Change from maintenance baseline in the endoscopy subscore at Week M-44.
- Ulcerative Colitis Endoscopic Index of Severity (UCEIS) score ≤ 4 at Week M-44.
- Change from induction baseline in the UCEIS score at Week M-44.
- Change from maintenance baseline in UCEIS score at Week M-44.
- Extraintestinal manifestations over time through Week M-44 (i.e., arthritis/arthralgia, aphthous stomatitis, erythema nodosum, iritis/uveitis, pyoderma gangrenosum, and primary sclerosing cholangitis). The attributes and strategies for the ICEs will not be used for this endpoint. Analyses will include the following:
  - Presence of extraintestinal manifestations at maintenance baseline and Week M-44
  - Change from induction and maintenance baseline in arthritis/arthralgia pain level at Week M-44
  - Absence of extraintestinal manifestations among participants with extraintestinal manifestations at induction baseline and maintenance baseline
- Association of histologic-endoscopic mucosal healing, histologic healing, or endoscopic healing at Week M-0 with the following efficacy parameters at Week M-44 will be presented:
  - Clinical remission
  - Corticosteroid-free clinical remission
  - Symptomatic remission
  - IBDQ remission
  - Fatigue response
  - Mayo score
  - Partial mayo score
  - Modified mayo score
  - Stool frequency subscore
  - Rectal bleeding subscore
  - CRP and
  - Fecal calprotectin
  - Clinical flare

- Clinical remission at Week M-44 and not receiving corticosteroids for at least 12 weeks prior to Week M-44.
- Clinical remission at Week M-44 and not receiving corticosteroids for at least 4 weeks prior to Week M-44.
- Clinical remission at Week M-44 and not receiving corticosteroids for at least 8 weeks prior to Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Clinical remission at Week M-44 and not receiving corticosteroids for at least 12 weeks prior to Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Clinical remission at Week M-44 and not receiving corticosteroids for at least 4 weeks prior to Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Clinical response at Week M-44 and not receiving corticosteroids for at least 8 weeks prior to Week M-44.
- Clinical response at Week M-44 and not receiving corticosteroids for at least 8 weeks prior to Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Change from maintenance baseline in the average daily prednisone-equivalent corticosteroid dose (excluding budesonide and beclomethasone dipropionate) at each visit through Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Corticosteroid-free for at least 8 weeks prior to Week M-44
- Elimination of concomitant corticosteroids for at least 8 weeks prior to Week M-44 among the participants who were receiving concomitant corticosteroids at maintenance baseline.
- Elimination of corticosteroids for at least 12 weeks prior to Week M-44 among participants who were receiving corticosteroids at maintenance baseline.
- Elimination of corticosteroids for at least 4 weeks prior to Week M-44 among participants who were receiving corticosteroids at maintenance baseline.
- Elimination of corticosteroids over time among participants who were receiving corticosteroids at maintenance baseline.
- Clinical remission, clinical response, endoscopic healing, endoscopic normalization, histologic-endoscopic mucosal healing, histologic-endoscopic mucosal healing (Alternative definition 1-2), and deep histologic-endoscopic mucosal healing at Week M-44 based on the local endoscopy subscores.

- Primary and major secondary endpoints, histologic healing, histologic remission, histologic-endoscopic mucosal healing (Alternative definition 1-2), and deep histologic-endoscopic mucosal healing at Week M-44 by:
  - ADT-naïve, ADT-experienced but not failed, ADT-non failure, and ADT-Failure subgroups
  - Biologic-naïve, biologic-experienced without documented failure [i.e., discontinued biologic therapy for other reasons], biologic-non failure and biologic-failure subgroups
  - Colonic molecular predictive signature status (MPS) at induction baseline
  - Mayo endoscopy subscore at induction baseline (2 vs. 3)
  - Induction Study 1 versus Induction Study 2

- The following endpoints that include the criteria of endoscopic subscore of 0 or 1: clinical remission, endoscopic healing, histologic-endoscopic mucosal healing, and histologic-endoscopic mucosal healing (Alternative Definition 1-2) will also be assessed regardless of whether the endoscopic subscore of 1 includes friability.

Additionally, the following will also be summarized descriptively:

- Summary of completeness of modified Mayo score and full Mayo score at Week M-44.
- Summary of missing modified Mayo score and missing full Mayo score components at Week M-44 by intercurrent event occurrence.
- Number of participants who experienced intercurrent events prior to Week M-44.
- Summary of Mayo subscores through Week M-44.
- UCEIS score at Weeks M-0 and M-44 by Mayo endoscopy subscore at the corresponding visit.
- Corticosteroid-free at Week M-44 (i.e., not requiring any treatment with corticosteroids for at least 8 weeks prior) among the participants who had achieved clinical remission at Week M-44.
- Primary and major secondary endpoints, among participants with an induction baseline modified Mayo score of 4 in the Efficacy All Randomized and Treated Analysis Set.

Inflammatory Biomarkers (CRP and Fecal Calprotectin)

- Change and percentage change from induction baseline in CRP over time through Week M-44.
- Change and percentage change from maintenance baseline in CRP over time through Week M-44.
- Change and percentage change from maintenance baseline in CRP over time through Week M-44 among participants with abnormal CRP at induction baseline.

- Change and percentage change from induction baseline in fecal calprotectin over time through Week M-44.
- Change and percentage change from maintenance baseline in fecal calprotectin over time through Week M-44.
- Change and percentage change from maintenance baseline in fecal calprotectin over time through Week M-44 among participants with abnormal fecal calprotectin at induction baseline.
- Normalization of CRP from maintenance baseline over time through Week M-44 among participants with abnormal CRP at induction baseline.
- Normalization of fecal calprotectin from maintenance baseline over time through Week M-44 among participants with abnormal fecal calprotectin at induction baseline.
- Change and percentage change from maintenance baseline in fecal calprotectin over time through Week M-44 among participants with fecal calprotectin > 150 mg/kg at induction baseline.
- Fecal calprotectin ≤ 150 mg/kg over time through Week M-44 among participants with fecal calprotectin > 150 mg/kg at induction baseline.

Health-Related Quality of Life

- Change from induction baseline in the total score of Inflammatory Bowel Disease Questionnaire (IBDQ) through Week M-44.
- Change from maintenance baseline in the total score of IBDQ through Week M-44.
- A ≥ 16-point or > 20-point improvement from induction baseline in the IBDQ total score through Week M-44.
- Change from induction baseline in each of the 4 dimensions of the IBDQ through Week M-44.
- Change from maintenance baseline in each of the 4 dimensions of the IBDQ through Week M-44.
- Change from induction baseline and response distribution for each of the 10 items in the IBDQ bowel domain (Attachment 7) through Week M-44.
- Change from maintenance baseline and response distribution for each of the 10 items in the IBDQ bowel domain (Attachment 7) through Week M-44.
- Change from induction baseline and response distribution for Q16 in the IBDQ social domain through Week M-44 (Q16: How often during the last 2 weeks have you had to avoid attending events where there was no washroom close at hand?).
- Change from maintenance baseline and response distribution for Q16 in the IBDQ social domain through Week M-44 (Q16: How often during the last 2 weeks have you had to avoid attending events where there was no washroom close at hand?).

- Participants with ≥ 2-point improvement from induction baseline in the IBDQ Q13 through Week M-44 among participants with Q13 ≤ 5 at induction baseline (Q13: How often during the last 2 weeks have you been troubled by pain in the abdomen?).
- Participants with ≥ 2-point improvement from induction baseline in the IBDQ Q16 through Week M-44 among participants with Q16 ≤ 5 at induction baseline.
- Participants with ≥ 2-point improvement from induction baseline in the IBDQ Q24 through Week M-44 among participants with Q24 ≤ 5 at induction baseline (Q24: How much of the time during the last 2 weeks have you been troubled by a feeling of having to go to the bathroom even though your bowels were empty?).
- IBDQ remission defined as total score of IBDQ ≥ 170, through Week M-44.
- Maintenance of IBDQ remission (Section 5.5.1) among participants with IBDQ remission at the maintenance baseline.
- IBDQ remission through Week M-44 among participants who had achieved IBDQ remission at maintenance baseline.
- Maintenance of ≥ 16-point improvement in IBDQ total score through Week M-44 among participants achieving ≥ 16-point improvement from induction baseline in IBDQ total score at maintenance baseline.
- Maintenance of > 20-point improvement in IBDQ total score through Week M-44 among participants achieving > 20-point improvement from induction baseline in IBDQ total score at maintenance baseline.
- Maintenance of ≥ 2-point improvement from induction baseline in IBDQ Q13 through Week M-44 among participants achieving ≥ 2-point improvement from induction baseline at maintenance baseline.
- Maintenance of ≥ 2-point improvement from induction baseline in IBDQ Q16 through Week M-44 among participants achieving ≥ 2-point improvement from induction baseline at maintenance baseline.
- Maintenance of ≥ 2-point improvement from induction baseline in IBDQ Q24 through Week M-44 among participants achieving ≥ 2-point improvement from induction baseline at maintenance baseline.
- Participants with IBDQ Q13 = 7 (“None of the time”) at Week M-44.
- Participants with IBDQ Q13 = 7 at Week M-44 among participants with abdominal pain (Q13 ≤ 6) at induction baseline.
- Participants with IBDQ Q16 = 7 (“None of the time”) at Week M-44.
- Participants with IBDQ Q16 = 7 at Week M-44 among participants with Q16 ≤ 6 at induction baseline.
- Participants with IBDQ Q24 = 7 (“None of the time”) at Week M-44.

- Participants with IBDQ Q24 = 7 at Week M-44 among participants with urgency Q24 ≤ 6 at induction baseline.
- Participants with IBDQ Q16 = 7 ("None of the time") and IBDQ Q24 = 7 ("None of the time") at Week M-44.
- Participants with IBDQ Q16 = 7 and IBDQ Q24 = 7 at Week M-44 among participants with Q16 ≤ 6 or Q24 ≤ 6 at induction baseline.
- Change from induction baseline in each of the 7 domain T-scores of Patient-Reported Outcomes Measurement Information System (PROMIS-29) and the pain intensity through Week M-44.
- Change from maintenance baseline in each of the 7 domain T-scores of PROMIS-29 and the pain intensity through Week M-44.
- A ≥ 3-point, ≥ 5-point, ≥ 7-point, or ≥ 9-point improvement from induction baseline in each of the 7 domain T-scores of PROMIS-29 through Week M-44.
- A ≥ 3-point improvement from induction baseline in the PROMIS-29 pain intensity through Week M-44 among the participants with pain intensity ≥ 3 at induction baseline.
- A ≥ 5-point or a ≥ 9-improvement from induction baseline in PROMIS Fatigue short form 7a through Week M-44.
- Change from induction baseline in PROMIS Fatigue Short Form 7a Score through Week M-44.
- Change from maintenance baseline in PROMIS Fatigue Short Form 7a Score through Week M-44.
- Change from induction baseline in the health state VAS score and EQ-5D dimensions through Week M-44.
- Change from maintenance baseline in the health state VAS score and EQ-5D dimensions through Week M-44.
- Summary of Patient's Global Impression of Severity (PGIS) of UC at induction baseline and through Week M-44.
- Change from induction baseline in PGIS of UC through Week M-44.
- Change from maintenance baseline in PGIS of UC through Week M-44.
- A ≥ 1-point or a ≥ 2-point improvement from induction baseline in PGIS of UC through Week M-44.
- Change from induction baseline in PROMIS-29 PCS and MCS through Week M-44.
- Change from maintenance baseline in PROMIS-29 PCS and MCS through Week M-44.
- A ≥ 5-point, ≥ 7-point, or ≥ 9-point improvement from induction baseline in PROMIS-29 PCS and MCS through Week M-44.

- Maintenance of ≥ 7-point improvement in PROMIS-29 PCS and MCS through Week M-44 among participants achieving ≥ 7-point improvement from induction baseline in PROMIS-29 PCS and MCS at maintenance baseline.
- Cumulative percent of participants by improvement from induction baseline in PROMIS Fatigue Short Form 5a T-score and 7a T-score at Week M-44.
- Probability density plot of the change from induction baseline in PROMIS Fatigue Short Form 5a and 7a Score at Week M-44.
- Change from induction baseline in PROMIS Fatigue Short Form 5a T-score through Week M-44.
- Change from maintenance baseline in PROMIS Fatigue Short Form 5a T-score through Week M-44.
- Change from induction baseline in PROMIS Fatigue Short Form 5a raw score through Week M-44.
- Change from maintenance baseline in PROMIS Fatigue Short Form 5a raw score through Week M-44.
- Change from induction baseline in PROMIS Fatigue Short Form 7a raw score through Week M-44.
- Change from maintenance baseline in PROMIS Fatigue Short Form 7a raw score through Week M-44.
- Change from induction baseline in PROMIS-29 Fatigue Domain raw score through Week M-44.
- Change from maintenance baseline in PROMIS-29 Fatigue Domain raw score through Week M-44.
- A ≥ 5-point, ≥ 7-point, or ≥ 9-point improvement from induction baseline in PROMIS Fatigue Short Form 5a T-score through Week M-44.
- Change from maintenance baseline in each of the 7 individual raw score items of the PROMIS Fatigue Short Form 7a.
- Participants in fatigue response at Week M-44 among participants with PROMIS Fatigue Short Form 7a T-score ≥ 36.9 at induction baseline.
- Maintenance of fatigue response through Week M-44 among participants in fatigue response at maintenance baseline.

Health Economics

For the endpoints listed below, analysis will be based on the data as observed and the attributes and strategies for the ICEs will only be applied to Work Productivity and Activity Impairment Questionnaire-General Health (WPAI-GH) endpoints:

- UC-related emergency department visits through Week M-44.

- UC-related hospitalizations through Week M-44.
- UC-related surgeries through Week M-44.
- UC-related hospitalizations or surgeries through Week M-44.
- The time to the first UC-related hospitalization through Week M-44.
- The time to the first UC-related surgery through Week M-44.
- The time to the first UC-related hospitalization or surgery through Week M-44.
- Change from induction baseline in WPAI-GH through Week M-44.
- Change from maintenance baseline in WPAI-GH through Week M-44.

5.5.1. Definitions

Clinical Endpoints

- The definition for clinical remission is provided in Section 5.3.1. Definitions of clinical response, symptomatic remission, endoscopic normalization, endoscopic healing, IBDQ remission, fatigue response, histologic-endoscopic mucosal healing, corticosteroid-free clinical remission and histologic healing are provided in Section 5.4.1.1. and for modified Mayo score, partial Mayo score, full Mayo score, see Section 5.3.1.1.
- **Histologic remission:** Absence of neutrophils from the mucosa (both lamina propria and epithelium), no crypt destruction, and no erosions, ulcerations or granulation tissue according to the Geboes grading system[3] (i.e., Geboes score ≤2B.0). This is equivalent to RHI-based histologic remission.
- RHI-based histologic remission: RHI ≤3 with sub-scores of 0 for lamina propria neutrophils and neutrophils in the epithelium and without ulcers or erosion according to the Robarts Histopathology Index[9] (See Attachment 6). This is equivalent to histologic remission defined above according to the Geboes grading system (i.e., Geboes score ≤2B.0).
- NHI-based histologic remission: NHI ≤1 according to the Nancy Histological Index[10] (See Attachment 5).
- **Histologic-endoscopic mucosal healing (Alternative definition 1)**: Achieving a combination of histologic remission (as defined above) and endoscopic healing (as defined in Section 5.4.1.1).
- **Histologic-endoscopic mucosal healing (Alternative definition 2):** achieving a combination of histologic remission (based on Nancy Index[4]) and endoscopic healing (as defined in Section 5.4.1.1).
- **Deep histologic-endoscopic mucosal healing:** Achieving a combination of endoscopic normalization (as defined in Section 5.4.1.1.) and histologic remission (as defined above).
- **Deep symptomatic remission**: a Mayo rectal bleeding subscore of 0 and a Mayo stool frequency subscore of 0.
- **Partial Mayo response**: a decrease from induction baseline of ≥2 in the partial Mayo score.

- **Durable symptomatic remission:** Symptomatic remission for ≥80% of all visits between Week M-0 and Week M-44 [i.e. at least 10 of 12 visits], which must include Week M-44.
- **Colonic molecular predictive signature (MPS):** this is a predictive gene expression signature, initially discovered from colon biopsies collected at baseline from a subset of participants in the ACT 1 infliximab UC study[6] refined in the PURSUIT golimumab UC study[7] and prospectively evaluated for prediction of mucosal healing in an open-label study of 103 UC participants treated with golimumab (PROgECT)[8] (see protocol 8.8.3 and 8.8.6).
- **UCEIS:** an index that provides an overall assessment of endoscopic severity of UC, based on mucosal vascular pattern, bleeding, and ulceration[9] The score ranges from 3 to 11, with a higher score indicating more severe disease by endoscopy. The UCEIS score will be assessed only by the central video readers for all endoscopies. See Attachment 4 for more details.
- Geboes total score: the continuous histology score is derived as the sum of all Geboes Grades and may take on values from 0 to 22.
- Geboes high activity subscore: the continuous histology score is derived as the sum of Geboes Grades 3, 4, and 5 that define histologic healing and may take on values from 0 to 10.
- Geboes low activity subscore: the continuous histology score is derived as the sum of Geboes Grades 0, 1, 2A and 2B and may take on values from 0 to 12.
- Elimination of corticosteroids is defined as achieving corticosteroid-free status through Week M-44.
- **Clinical flare**: Participants who meet the following criteria will be considered to be in clinical flare:
  - an increase from maintenance baseline in the partial Mayo score (i.e., the Mayo score without the endoscopy subscore) of ≥2 points and an absolute partial Mayo score ≥4;

  OR

  - an absolute partial Mayo score ≥7 points

CRP

C-reactive protein (CRP) has been demonstrated to be useful as a marker of inflammation in participants with IBD. In participants with UC, elevated CRP has been associated with severe clinical activity, an elevated sedimentation rate, and active disease as detected by colonoscopy[10,11]. C-reactive protein will be assayed using a validated, high sensitivity CRP assay. CRP normalization is defined as CRP concentration ≤3 mg/L.

Fecal Calprotectin

Fecal calprotectin has been demonstrated to be a sensitive and specific marker in identifying colonic inflammation and response to treatment in participants with IBD, especially in UC[12].

Assays for fecal calprotectin will be performed by the central laboratory using a validated method. Fecal calprotectin normalization is defined as fecal calprotectin concentration ≤250 mg/kg.

Patient-Reported Outcomes

IBDQ

The IBDQ[13] is a validated, 32-item, self-reported questionnaire for participants with IBD that will be used to evaluate the disease-specific HRQoL across 4 dimensional scores: bowel symptoms (loose stools, abdominal pain), systemic functions (fatigue, altered sleep pattern), social function (work attendance, need to cancel social events), and emotional function (anger, depression, irritability). Scores range from 32 to 224, with higher scores indicating better outcomes.

The individual IBDQ dimensions will be calculated when no more than 1 item is missing in the dimension. If an item is missing, it will be estimated using the average value across the non-missing items. If any of the 4 dimensions of the IBDQ cannot be calculated, then the total IBDQ score cannot be calculated and will be missing for that visit.

- **IBDQ remission:** total IBDQ score ≥170 (Irvine et al, 1994[13]; Higgins et al, 2005[14])
- **Maintenance of IBDQ remission:** in IBDQ remission at both Week M-28 and Week M-44 among participants with IBDQ remission at the maintenance baseline.

PROMIS-29

The Patient-Reported Outcomes Measurement Information System (PROMIS)-29 is a validated general health profile instrument that is not disease-specific. It is a collection of short forms containing 4 items for each of 7 domains (depression, anxiety, physical function, pain interference, fatigue, sleep disturbance, and ability to participate in social roles and activities). PROMIS-29 also includes an overall average pain intensity 0-10 numeric rating scale. Norm-based scores have been calculated for each domain on the PROMIS measures, with a score of 50 representing the mean or average of the reference population. On symptom-oriented domains of PROMIS-29 (anxiety, depression, fatigue, pain interference, and sleep disturbance), higher scores represent worse symptomatology. On the function-oriented domains (physical functioning and social role), higher scores represent better functioning. Additionally, the physical component summary score (PCS) and mental component summary score (MCS) will each be derived from all 7 domain scores of PROMIS-29 (Hays et al, 2018[17]) as measures for general health related quality (HRQOL). Higher PCS and MCS scores indicate better HRQOL.

PROMIS Fatigue 7-item Short Form

The PROMIS Fatigue 7-item Short Form (PROMIS Fatigue 7a) contains 7 items evaluating fatigue-related symptoms (i.e., tiredness, exhaustion, mental tiredness, and lack of energy) and associated impacts on daily activities (i.e., activity limitations related to work, self-care, and exercise) using a 5-point Likert scale ranging from 1 (never) to 5 (always) over the past 7 day. The raw total score (ranging from 5-35) is converted into a standardized score with a mean of 50 and a standard deviation (SD) of 10 (T-Score). Compared to the fatigue domain of PROMIS-29, PROMIS Fatigue SF 7a provides additional information to evaluate frequency of fatigue.

- **Fatigue response**: ≥ 7-point improvement in PROMIS Fatigue short form 7a.

PROMIS Fatigue 5-item Short Form

PROMIS Fatigue 5-item Short Form (PROMIS Fatigue SF5a): PROMIS Fatigue SF5a is a customized short form to evaluate fatigue frequency. It consists of items 1-5 of SF7a (the name of each item is labeled as FATEXP20, FATEXP5, FATEXP18, FATIMP33, FATIMP30 from of the instrument, which are the names used in fatigue item bank from the developer). Similarly, the raw score of SF4a is the sum of score of 5 items (5-25). The raw total score is converted into a standardized T-score calibrated from the general population using a service provided by the developer.

Patient's Global Impression of Severity of Ulcerative Colitis

Participants will rate their UC disease activity at the baseline and planned visit using a 5-point scale ("None", "Mild", "Moderate", "Severe" and "Very Severe").

EQ-5D-5L

The EQ-5D-5L is a validated instrument consisting of the EuroQol 5 dimensions descriptive system (EQ-5D) and the EuroQol visual analog scale (EQ-VAS). The descriptive system comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The respondent is asked to indicate his/her health state by checking the most appropriate statement in each of the 5 dimensions. The EQ-VAS records the respondent's self-rated health on a 20-cm vertical, visual analog scale with endpoints labeled 'the best health you can imagine' and 'the worst health you can imagine'. The respondents mark an "X" on the scale to indicate their health TODAY and then write the number marked on the scale in the box.

Health Economics

Work Productivity and Activity Impairment Questionnaire-General Health

The WPAI-GH will also be utilized to evaluate work productivity. The WPAI-GH is a validated instrument created as a patient-reported quantitative assessment of the amount of absenteeism, presenteeism, and daily activity impairment attributable to general health. The WPAI-GH consists of 6 questions to determine employment status, hours missed from work due to health problems, hours missed from work for other reasons, hours actually worked, the degree to which general health affected work productivity while at work, and the degree to which general health affected activities outside of work. Four scores are derived: percentage of absenteeism, percentage of presenteeism (reduced productivity while at work), an overall work impairment score that combines absenteeism and presenteeism, and percentage of impairment in activities performed outside of work. Higher scores indicate greater impairment.

5.5.2. Estimands for the Other Endpoints

The attributes and strategies for the ICEs that were used for the primary estimand for the primary endpoint analysis (Section 5.3.2.) will also be used for other endpoints (Section 5.5.) except for the endpoints related to extraintestinal manifestations and UC-related hospitalization or surgery,

which will be based on data as observed. The methods to be used to analyze the other endpoints (i.e., binary, continuous, ordinal, and time to event) are defined in the Section 5.5.3. (below).

5.5.3. Analysis Methods for the Estimands for the Other Endpoints

Unless otherwise specified, other efficacy endpoints listed and defined in Section 5.5. above will be analyzed based on **Randomized Full Analysis Set** according to randomized treatment group regardless of the treatment actually received.

All statistical testing will be performed at the 2-sided 0.05 significance level. No adjustments for multiple comparisons will be made for these other endpoints and nominal p-values will be presented.

Binary Endpoints

The intercurrent events captured for binary endpoints specified in Section 5.3.2. will be applied to each of the above binary endpoints except extraintestinal manifestations. Participants with ICEs in categories 1-4 and 6 will be considered not to have achieved the binary endpoints. For participants experiencing an ICE 5, their observed values (if available) will be used. Note that the application of ICE categories 1-3 overrides that of ICE 5.

Participants with any missing data for an endpoint after application of ICE strategies will be imputed as not achieving the associated binary endpoints. Binary endpoints will be summarized with the number and frequency of participants who achieve the endpoint by treatment group, as well as the associated 95% confidence interval. Treatment comparisons (each guselkumab group versus placebo) will be performed using CMH test stratified by clinical remission status at maintenance baseline and induction dose treatment to compare the proportion of participants achieving the endpoints. In case of rare events, the Fisher's exact test will be used for treatment comparisons.

Continuous Endpoints

The ICEs specified in Section 5.3.2. will be applied to each of the above continuous endpoints, i.e., if a participant has an ICE in categories 1-4 and 6, induction baseline values will be assigned from the point of ICE onward (i.e., no change from baseline). For participants experiencing an ICE 5, their observed values (if available) will be used. Note that the application of ICE categories 1-3 overrides that of ICE 5.

To account for the missing data (after applying the ICE strategies) for continuous endpoints of change from baseline measured at more than one post-baseline visit through Week M-44, a Mixed-Effect Model Repeated Measures (MMRM) will be used, under the assumption of missing at random (MAR), to test the difference between each guselkumab group and placebo. In MMRM, missing data will not be imputed, but rather missing data will be accounted for through correlation of repeated measures in the model. Additionally, if the MMRM normality assumption is in question, an appropriate transformation may be implemented before fitting the MMRM model.

The explanatory variables of the MMRM model will include treatment group, clinical remission status at maintenance baseline (Yes/No), and induction dose treatment (guselkumab 400 mg, guselkumab 200 mg, placebo crossover to guselkumab 200 mg), respective baseline score, and an interaction term of visit with treatment group.

An unstructured covariance matrix for repeated measures within a participant will be used. The F-test will use Kenward-Roger's approximation for degree of freedom. In case of lack of convergence, empirical structured covariances will be used in the following order until convergence is reached: 1) Toeplitz 2) first order Autoregressive Moving Average.

The treatment difference between each guselkumab group and placebo will be estimated by the difference in the least squares means (LSmeans). The 95% 2-sided CI for the differences in LSmeans and p-values will be calculated based on the MMRM.

Endpoints that are measured at only one post-baseline visit through Week M-44 (e.g., the full Mayo score or modified Mayo score) will be compared between each guselkumab group and placebo using an analysis of covariance (ANCOVA) with treatment group, clinical remission status at maintenance baseline (Yes/No), induction dose treatment, and corresponding baseline value as covariates.

For analyses using ANCOVA, unless otherwise specified, multiple imputation (same as defined in Section 5.3.2.2.2.) will be used for missing data (after applying ICE strategies), under the assumption that the data are missing at random. The analysis will impute the missing scores at Week M-44 using the corresponding value at baseline and related component score at pre-Week M-44 visits if applicable, clinical remission status at maintenance baseline, induction dose treatment, and treatment group.

Ordinal Endpoints

The ICEs specified in Section 5.3.2. will be applied to each of the above ordinal endpoints, i.e., if a participant has an ICE in categories 1-4 and 6, induction baseline values will be assigned from the point of ICE onward if applicable. For participants experiencing an ICE 5, their observed values (if available) will be used. Note that the application of an ICE in categories 1-3 overrides that of ICE 5.

For the ordinal endpoints, the treatment comparisons (each guselkumab group versus placebo) will be performed using CMH (Row Mean Scores) stratified by clinical remission status at maintenance baseline and induction dose treatment. Missing data (after applying the ICEs) will not be imputed for ordinal endpoints.

Time to Event Endpoints

The time to first loss of partial Mayo response (not meeting the partial Mayo response criteria) and time to first loss of symptomatic remission (not meeting the symptomatic remission criteria) through Week M-44 will be compared between the treatment groups using the stratified log-rank test with clinical remission status at maintenance baseline and induction treatment as the

stratification factors. The Kaplan-Meier curve by treatment group through Week M-44 will be provided. The time to first loss of partial Mayo response or first loss of symptomatic remission is defined as the number of days elapsed from the date of the Week M-0 study intervention administration to the date of the first loss of partial Mayo response or first loss of symptomatic remission prior to or at Week M-44. Participants with an ICE in categories 1-4 and 6 will be considered as having lost partial Mayo response or symptomatic remission from the point of first ICE onward. Participants who have not lost partial Mayo response or symptomatic remission prior to or at Week M-44 will be censored at the last Mayo subscore assessment at or prior to Week M-44.

The time to the first UC-related hospitalization or surgery through Week M-44 will be compared using the stratified log-rank test with clinical remission status at maintenance baseline and induction treatment as the stratification factors. The Kaplan-Meier curve by treatment group will be provided. The time to the first UC-related hospitalization or surgery is defined as the number of days elapsed from the date of the Week M-0 study intervention administration in this maintenance study to the date of the first hospitalization or surgery prior to or at Week M-44. Participants who are not hospitalized nor had a surgery prior to Week M-44 will be censored at Week M-44, date of last available visit (including the safety follow-up visit), or early termination, whichever happens earlier.

5.6. Efficacy Endpoints in the Nonrandomized Full Analysis Set

The proportion of participants in clinical remission; symptomatic remission; endoscopic healing; Corticosteroid-free clinical remission; maintenance of clinical response; histologic-endoscopic mucosal healing; IBDQ remission; fatigue response; maintenance of clinical remission among the participants who had achieved clinical remission at maintenance baseline; and endoscopic normalization, all at Week M-44, and symptomatic remission at each visit through Week M-44, will be summarized descriptively based on the **Nonrandomized Full Analysis Set** (Section 4). The ICEs specified in Section 5.3.2. will be applied with ICE 2 (have a dose adjustment [including a sham dose adjustment]) excluded, and missing data rule (i.e., missing value will be imputed as non-responder) used for primary estimand will also be applied.

5.7. Efficacy in Participants Who Had a Dose Adjustment

The attributes and strategies for the ICEs that were used for the primary estimand for the primary endpoint analysis (Section 5.3.2.) will not be used for endpoints described in this section. The missing data rule used for primary estimand will be applied to symptomatic remission (i.e., missing value will be imputed as non-responder); however, for continuous data there will be no imputation for the missing data.

For the participants in the **Randomized Full Analysis Set** who had a dose adjustment (**Dose Adjustment Analysis Set**), the following endpoints at the time of dose adjustment and through 12 weeks after dose adjustment will be summarized and presented using descriptive statistics only (i.e., no statistical tests will be performed).

- Symptomatic remission

- Symptomatic response. Symptomatic response was defined as a decrease from induction baseline in the symptomatic Mayo Score by ≥30% and ≥1 point, with either a ≥1 point decrease from induction baseline in the rectal bleeding subscore or a rectal bleeding subscore of 0 or 1.
- Partial Mayo response
- CRP

In addition, the following summaries will be provided:

- The change in the Partial Mayo score from the time of dose adjustment through 12 weeks after dose adjustment.
- The change in CRP at or after 12 weeks from the time of dose adjustment.

5.8. Dose Adjustment as a Treatment Strategy

The attributes and strategies for the ICEs that were used for the primary estimand for the primary endpoint analysis (Section 5.3.2.) will also be used except that application of ICE 2 (i.e., have a dose adjustment) will be suspended for endpoints described in this section. The same analysis methods for the main estimands described in Section 5.3.2.1.1. and Section 5.4.1.3. will be used for the corresponding endpoints listed below.

The following endpoints will be presented by treatment group based on the **Randomized Full Analysis Set**:

- Clinical remission at Week M-44
- Symptomatic remission at Week M-44
- Symptomatic remission at each visit through Week M-44
- Endoscopic healing at Week M-44
- Corticosteroid-free clinical remission at Week M-44
- Clinical response at Week M-44
- Histologic-endoscopic mucosal healing at Week M-44
- IBDQ remission at Week M-44
- Fatigue response at Week M-44
- Clinical remission at Week M-44 among the participants who had achieved clinical remission at maintenance baseline.
- Endoscopic normalization at Week M-44
- Histologic healing at Week M-44
- Histologic remission at Week M-44

5.9. Safety Analyses

Safety data, including but not limited to, AEs and changes in laboratory assessments, will be summarized. Treatment-emergent AEs will be summarized by treatment group and Medical Dictionary for Regulatory Activities (MedDRA) system organ class and preferred terms.

Unless otherwise mentioned, safety summaries will be provided for the **Randomized Safety Analysis Set**, the **Nonrandomized Safety Analysis Set, and** the **Safety Analysis Set**. In general, participants will be analyzed according to their assigned treatment. However, participants assigned to placebo who incorrectly received guselkumab at any time during the maintenance phase (i.e., up to Week M-44) will be analyzed in the guselkumab group; participants assigned to guselkumab who received **only** placebo during the maintenance phase (i.e., up to Week M-44) will be analyzed in the placebo group.

For all continuous safety variables, descriptive statistics by treatment group will include the N, mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized by treatment group using frequency counts and percentages.

5.9.1. Extent of Exposure

The number and percentage of participants who receive study intervention through M-44 will be summarized based on the **Enrolled Analysis Set**. The number of administrations of study intervention received, the cumulative dose of study intervention, and the average duration of follow-up in weeks will be summarized by treatment group through M-44 based on **Randomized Safety Analysis Set**, the **Nonrandomized Safety Analysis Set, and** the **Safety Analysis Set**.

The distribution of participants by study intervention lot through Week M-44 will also be provided based on the **All Treated Analysis Set**.

5.9.2. Adverse Events

The verbatim terms used in the CRF by investigators to identify adverse events will be coded using the Medical Dictionary for Regulatory Activities (MedDRA). Any AE occurring at or after the initial administration of study intervention of this maintenance study is considered to be treatment emergent. All reported treatment-emergent adverse events will be included in the analysis. For each adverse event, the number and percentage of participants who experience at least 1 occurrence of the given event will be summarized by treatment group. Since safety should be assessed relative to exposure and follow-up, all AE summary tables will summarize the average weeks of follow-up and average exposure (number of administrations) for each treatment group.

Type of AEs to assess the safety of participants:

Summary tables will be provided for the following treatment-emergent adverse events (TEAEs):

- AEs (including infections)
- SAEs (SAEs including serious infections)
- Reasonably related AEs as assessed by the investigator

- AEs of severe intensity
- AEs leading to discontinuation of study intervention
- Injection-site reactions

Overall TEAE summary table and number of TEAEs per hundred subject-years of follow-up through Week M-44 for the events above (except for injection-site reactions) will also be provided.

An overall TEAE summary table of events through Week M-44 and frequency and type of AEs will also be provided for the **All Treated Analysis Set, Safety All Randomized and Treated Analysis Set, and Safety All Nonrandomized and Treated Analysis Set.**

In addition to the summary tables, listings will be provided for participants in the **All Treated Analysis Set** who:

- Had SAEs
- Had AEs of severe intensity
- Had AEs leading to discontinuation of study intervention
- Died
- AE of special interest (malignancy and tuberculosis)
- Other AEs of interest (e.g., opportunistic infections, MACE [CV death, nonfatal MI, nonfatal stroke], drug-related hepatic disorders, venous thromboembolism (VTE), anaphylactic reactions, and serum-sickness)

The above AEs will either be presented in a listing or described in the clinical study report. A list of AEs of special interest and other AEs of interest are provided in Appendix 7.

Definitions

- A reasonably related AE is defined as any event with a relationship to study agent of 'Very likely', 'Probable', or 'Possible' on the AE eCRF page or if the relationship to study agent is missing.
- An infection is any AE that was recorded based on the MedDRA system organ class "Infections and Infestations".
- A study intervention injection-site reaction is any reaction at an SC study intervention injection site that was recorded as an injection-site reaction by the investigator on the eCRF.

5.9.3. Additional Safety Assessments

5.9.3.1. Clinical Laboratory Tests

Laboratory assessments include, but are not limited to, the assessments listed below:

- Hematology: hemoglobin, hematocrit, platelet count, total and differential WBC count.

- Blood chemistry: total and direct bilirubin, alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, albumin, total protein, calcium, phosphate, sodium, potassium, chloride, blood urea nitrogen (BUN), and creatinine.

The maintenance baseline value for a participant is the value closest to but prior to the first dose of study agent of the maintenance study unless specified otherwise. In addition, change from maintenance baseline is defined to be the assessment at the post-maintenance baseline visit minus the assessment at maintenance baseline. There will be no imputation for missing laboratory values.

Laboratory parameters and change from maintenance baseline in laboratory parameters (hematology and chemistry) through Week M-44 will be summarized and displayed by treatment group based on the **Randomized Safety Analysis Set** and **the Nonrandomized Safety Analysis Set**.

The following summaries will also be presented through Week M-44 for the **Randomized Safety Analysis Set**, the **Nonrandomized Safety Analysis Set,** and the **Safety Analysis Set**:

- Summary of maximum modified National Cancer Institute Common Terminology Criteria for Adverse Events (NCI-CTCAE) toxicity grade for post-maintenance baseline laboratory values through Week M-44 for the predefined hematology and chemistry lab parameters except for liver tests (i.e., ALT, AST, alkaline phosphatase, and total bilirubin). This summary will also be presented for the **All Treated Analysis Set**. Shift tables for maximum modified NCI-CTCAE toxicity grade from maintenance baseline through Week M-44 will be summarized for the predefined hematology and chemistry lab parameters except for liver tests (i.e., ALT, AST, alkaline phosphatase, and total bilirubin).
- Summary of maximum post-maintenance baseline measurement through Week M-44 for liver tests (i.e., ALT, AST, alkaline phosphatase, and total bilirubin) relative to ULN threshold. This summary will also be presented for the **All Treated Analysis Set**, the **Safety All Randomized**, the **Treated Analysis Set,** and the **Safety All Nonrandomized and Treated Analysis Set**.

Line graphs will also be provided for ALT, AST, total bilirubin, and alkaline phosphatase.

Select clinical laboratory test analyses will utilize a modified NCI-CTCAE toxicity grade (Appendix 9), and liver tests (i.e., ALT, AST, total bilirubin, and alkaline phosphatase) will use the predefined upper limit normal (ULN) thresholds (Appendix 10).

Listings based on the **All Treated Analysis Set** will be provided for participants with any of the following:

- Abnormal post-maintenance baseline laboratory values of toxicity grade ≥ 2 except liver tests

- Post-maintenance baseline elevated liver tests of AST or ALT ≥ 5xULN, or total bilirubin ≥ 2xULN, or alkaline phosphatase ≥ 2xULN
- Post-maintenance baseline elevated liver tests with combined ALT or AST ≥ 3xULN and total bilirubin ≥ 2xULN.

5.9.4. Other Safety Parameters

5.9.4.1. Suicidal Ideation and Behavior

The Columbia-Suicide Severity Rating Scale (C-SSRS) will be used as a screening tool to prospectively evaluate suicidal ideation and behavior in this study, as part of a comprehensive evaluation of safety. The C-SSRS is an investigator-administered questionnaire[15, 16] that defines five subtypes of suicidal ideation and 4 possible suicidal behaviors, as well as non-suicidal self-injurious behavior and completed suicide.

The maintenance baseline is defined as the most severe/maximum score at Week M-0. Suicidal ideation and behavior will be analyzed by the most severe/maximum post-maintenance baseline C-SSRS outcome or AE of suicidal ideation and behavior. Listings based on the **All Treated Analysis Set** will be provided for participants with positive (i.e., score >0) ideation and behavior.

5.10. Other Analyses

5.10.1. Pharmacokinetics

5.10.1.1. Serum Guselkumab Concentrations

PK analyses will be performed based on the **PK Analysis Set** and **Randomized PK Analysis Set** (Section 4), unless otherwise specified.

Descriptive statistics (N, mean, SD, median, range, coefficient of variation (%) and interquartile range) will be used to summarize concentrations at each sampling time point by treatment group (unless otherwise specified, the treatment group mentioned in this section refers to the maintenance treatment group). PK data may be displayed graphically over time by treatment group.

The proportion of participants without detectable serum guselkumab concentration (below the lower limit of quantification) at each visit by treatment group through Week M-44 will also be presented.

In addition, serum guselkumab concentrations at each visit based on **Randomized PK Analysis Set** will be presented:

- by induction and maintenance treatment groups
- by induction baseline body weight quartiles and maintenance treatment groups.

- by induction baseline use of concomitant immunomodulators and maintenance treatment groups
- by induction baseline ADT failure status (Yes/No) and maintenance treatment groups

PK data will be displayed graphically, i.e.:

- Plot of median serum guselkumab concentrations by treatment group through Week M-44
- Plot of median serum guselkumab concentrations through Week M-44 by treatment group and Week M-44 clinical response status for the **Randomized PK Analysis Set**

Serum guselkumab concentrations will also be summarized for participants who had a dose adjustment and receive an increased dose (i.e., participants randomized to placebo and increased to guselkumab 200 mg, and participants randomized to guselkumab 100 mg and increased to 200 mg). Serum guselkumab concentrations over time will be summarized (when the number of participants permitted) starting from the time of dose adjustment. Summary of PK concentrations (micrograms/mL) through Week M-44; treated subjects who were randomized and did not have a dose increase will also be provided.

Data Handling Rules:

Unless otherwise specified, the following data handling rules will apply to PK analyses:

- Participants will be analyzed according to their assigned treatment.
- All serum concentration summaries for a particular time point will include data obtained from treated participants at the timepoint of interest without imputing any missing data.
- A concentration not quantifiable (below the lower limit of quantification) will be treated as 0 in the summary statistics and shown as the lower limit of quantification (< LLOQ) in the data listings.
- The data from a participant who meets any of the following dosing deviation criteria will be excluded from the by-visit data analyses from that point onwards:
  - Discontinued guselkumab administrations.
  - Skipped a guselkumab administration.
  - Received an incomplete / incorrect dose.
  - Received an incorrect study intervention.
  - Received an additional guselkumab dose.

In addition, if a participant has an administration outside of dosing windows (± 10 days), the concentration data collected at and after that visit will be excluded from the by-visit data analyses. For participants who received an increased dose at the time of dose adjustment, data after dose adjustment were excluded (this does not apply to the summary of serum guselkumab concentrations for participants who have a dose adjustment and receive an increased dose).

5.10.1.2. PK vs Efficacy

The relationship between serum guselkumab concentrations and efficacy endpoints will be explored for the **Randomized PK Analysis Set**, e.g.:

- The relationship between serum guselkumab concentrations (quartiles) at Week M-44 and clinical response, clinical remission, symptomatic remission, endoscopic healing, histologic-endoscopic mucosal healing, and endoscopic normalization, and the change in the modified Mayo score at Week M-44 will be explored. Similar PK vs efficacy analysis by ADT failure status (yes, no) may also be provided.
- Summary of change from maintenance baseline in CRP concentration (mg/L) and Fecal Calprotectin Concentration (mg/kg) at Week M-44 by serum guselkumab concentration quartiles at Week M-44 will be presented.

5.10.1.3. Population PK Analysis

If sufficient data are available, then population PK analysis using serum concentration-time data of guselkumab will be performed using nonlinear mixed-effects modeling (NONMEM). Details will be given in a population PK analysis plan and the results of the analysis will be presented in a separate report.

5.10.2. Immunogenicity

5.10.2.1. Antibodies to Guselkumab

"Sample ADA status" and sample titer as well as the cumulative "participant ADA status" and peak titer through the visit will be coded and provided by the bioanalytical group.

Participants evaluable for immunogenicity are defined as having at least one post-dose ADA time point collected for antibodies to guselkumab detection.

The antibodies to guselkumab status (positive at any time, negative) and titers will be summarized by treatment group through Week M-44 for all participants in the following analysis sets:

- Randomized Immunogenicity Analysis Set
- Randomized Immunogenicity Analysis Set who increased their dose through Week M-44
- Randomized Immunogenicity Analysis Set who did not increase their dose through Week M-44
- Immunogenicity Analysis Set
- Continuous Guselkumab Immunogenicity Analysis Set
- Immunogenicity All Treated Analysis Set

The maximum titers of antibodies to guselkumab will be provided for participants who are positive for antibodies to guselkumab. The antibodies to guselkumab summary and analysis will be based on the observed data; therefore, no imputation of missing data will be performed.

In addition, a listing of participants who are positive for antibodies to guselkumab will be provided. The sample antibodies status, the titer, and the neutralizing antibodies status to guselkumab will be listed by visit. This listing will also provide information regarding immunomodulator status at induction baseline, dose administered, injection site reactions, adverse events temporally associated with an infusion, guselkumab serum concentration, and modified Mayo score (at applicable visits) for all visits. In addition, a list of antibodies to guselkumab status in participants who discontinued study intervention early will be provided.

5.10.2.2. Neutralized Antibodies to Guselkumab

The incidence of neutralizing antibodies (NAbs) to guselkumab will be summarized for participants who are positive for antibodies to guselkumab for all participants in the **Immunogenicity Analysis Set**, **Randomized Immunogenicity Analysis Set**, **Guselkumab Immunogenicity Analysis Set**, and **Immunogenicity Analysis Set** and have samples evaluable for Nabs to guselkumab through Week M-44.

5.10.2.3. Antibodies vs PK/Efficacy/Safety

To explore the relationship between antibodies to guselkumab status and serum guselkumab concentrations, efficacy and safety, the following analyses may be performed, if sufficient numbers of participants are positive for antibodies.

- Summary of serum guselkumab concentrations over time through Week M-44 (data following dose increase will be excluded) by antibodies to guselkumab status through Week M-44 (**Randomized Immunogenicity Analysis Set**).
- Plots of median trough serum guselkumab concentrations over time through Week M-44 (data following dose increase will be excluded) by antibodies to guselkumab status through Week M-44 (**Randomized Immunogenicity Analysis Set**).
- Summary of clinical response, clinical remission, symptomatic remission, change from maintenance baseline in modified Mayo score, endoscopic healing, histologic-endoscopic mucosal healing, and endoscopic normalization at Week M-44 by antibodies to guselkumab status through Week M-44 for the **Randomized Immunogenicity Analysis Set.**
- Summary of injection-site reactions through Week M-44 by antibodies to guselkumab status through Week M-44 for the **Randomized Immunogenicity Analysis Set.**

5.10.3. Pharmacokinetic/Pharmacodynamic Relationships

If data permit, the relationships between serum guselkumab concentration and efficacy may be analyzed graphically. If deemed feasible and necessary, exposure-response analyses may be performed. The analysis methods will be summarized in a separate analysis plan. Results of such analyses may be presented in a separate technical report.

5.10.4. Biomarkers

Changes in serum protein analytes, biopsy and whole blood RNA obtained over time will be summarized by treatment group. Associations between baseline levels and changes from baseline in select biomarkers and response to treatment will be explored. Biomarker analyses will be summarized in a separate technical report.

The biomarker analyses will characterize the effects of guselkumab concentration to identify biomarkers relevant to treatment, and to determine if these biomarkers can predict response to guselkumab. Results of serum, stool, whole blood, and mucosal biopsy analyses will be reported in separate technical reports.

5.10.5. Health Economics

Medical resource utilization, including but not limited to UC-related emergency department visits, hospitalizations, and surgeries, will be collected in this study. The WPAI-GH will also be utilized to evaluate work productivity. Analyses for medical resource utilization and health economics data are described in Section 5.5 (Other Endpoints).

5.10.6. Definition of Subgroups

The primary endpoint will be evaluated for subgroups based on demographics and baseline UC clinical disease characteristics, UC-related concomitant medication usage, and UC-related medication history all at Week 0 of the induction study, as well as maintenance stratification factors and UC clinical disease characteristics at Week 0 of the maintenance study, when the number of participants within each level of the subgroup permits.

Subgroup	Definition
Demographics at Induction Baseline	
Region	• Eastern Europe: Bulgaria, Czech Republic, Hungary, Poland, Russia, Serbia, Slovakia, Ukraine, Turkey, Latvia • Asia: China, Japan, Malaysia, South Korea, Taiwan • Rest of World: Argentina, Australia, Austria, Belgium, Brazil, Canada, France, Germany, Ireland, Israel, Italy, Jordan, Netherlands, New Zealand, Portugal, Spain, Sweden, United Kingdom, United States
Baseline Age	• ≤ median Age • > median Age
Gender	• male • female
Race	• Caucasian • non-Caucasian
Baseline body weight	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile
Tobacco or nicotine use status	• non-user • prior user • current user

Subgroup	Definition
UC Clinical Disease Characteristics at Induction Baseline	
UC disease duration	• ≤ 5 years • > 5 years to ≤ 15 years • > 15 years
UC disease duration	• < 2 years • ≥ 2 years to ≤ 5 years • > 5 years to ≤ 10 years • > 10 years
Extent of disease	• limited • extensive
Severity of UC disease	• moderate: 6 ≤ Mayo score ≤ 10 • severe: Mayo score > 10
Severity of UC disease	• Modified Mayo score of 5 – 6 (moderate) • Modified Mayo score of 7 – 9 (severe)
Baseline endoscopy subscore	• moderate: subscore of 2 • severe: subscore of 3
Extraintestinal manifestations	• absent • present
CRP	• ≤ 3 mg/L • > 3 mg/L
CRP	• ≤ median • > median
CRP	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile
Fecal calprotectin	• ≤ 250 mg/kg • > 250 mg/kg
Fecal calprotectin	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile
Albumin	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile
UC-related Concomitant Medications at Induction Baseline	
Oral 5-ASA compounds	• receiving • not receiving
Oral corticosteroids including budesonide and beclomethasone dipropionate	• receiving • not receiving
6-MP/AZA/MTX	• receiving • not receiving
Oral corticosteroids and 6-MP/AZA/MTX	• receiving • not receiving
Oral corticosteroids or 6-MP/AZA/MTX	• receiving • not receiving

Subgroup	Definition
UC-related Medication History at Induction Baseline	
ADT failure: inadequate response or failure to tolerate advanced therapy (ADT; i.e., tumor necrosis factor alpha [TNFα] antagonists [golimumab, infliximab, adalimumab and biosimilars to these], vedolizumab, or tofacitinib)	• yes • no
Participants with ADT failure	For participants with ADT failure to: • One ADT class ▪ Anti-TNF only ▪ Vedolizumab only ▪ Tofacitinib only • Two ADT classes ▪ Anti-TNF and tofacitinib ▪ Anti-TNF and vedolizumab ▪ Vedolizumab and tofacitinib • Three ADT classes ▪ Any anti-TNF and vedolizumab and tofacitinib • Two or more ADT classes • Other ADT-failure combinations ▪ At least one anti-TNF (regardless of vedolizumab or tofacitinib) ▪ Vedolizumab (regardless of anti-TNF or tofacitinib) ▪ Tofacitinib (regardless of anti-TNF or vedolizumab) ▪ Any anti-TNF and vedolizumab (regardless of tofacitinib) ▪ Any anti-TNF and tofacitinib (regardless of vedolizumab) ▪ Vedolizumab and tofacitinib (regardless of anti-TNF) For participants with ADT failure • primary nonresponse (yes) • secondary nonresponse (yes) • intolerance (yes) For participants with failure to at least one anti-TNF • primary nonresponse (yes) • secondary nonresponse (yes) • intolerance (yes) For participants with failure to vedolizumab • primary nonresponse (yes) • secondary nonresponse (yes) • intolerance (yes)

Subgroup	Definition
	For participants with failure to tofacitinib (when the number of participants within the subgroup permits) • primary nonresponse (yes) • secondary nonresponse (yes) • intolerance (yes)
Participants without ADT failure	• naïve • ADT-experienced [but not documented failure]
Participants with biologic (i.e., tumor necrosis factor alpha [TNFα] antagonists and vedolizumab) failure	• yes • no
Participants without biologic failure	• naïve • bio-experienced [but not documented failure])
Colonic molecular predictive signature (MPS) status at induction baseline	• Predicted endoscopic healing • Predicted endoscopic non-healing
Refractory, dependent, or intolerant to oral or IV corticosteroids	• yes • no
Refractory or intolerant to 6-MP/AZA	• yes • no
Refractory, dependent, or intolerant to oral or IV corticosteroids, but not refractory or intolerant to 6-MP/AZA	• yes • no
Refractory, dependent or intolerant to oral or IV corticosteroids, and refractory or intolerant to 6-MP/AZA	• yes • no
UC clinical disease characteristics at Week M-0	
Endoscopic healing status	• yes • no
CRP	• ≤ 3 mg/L • > 3 mg/L
	• ≤ median • > median
Fecal calprotectin	• ≤ 250 mg/kg • > 250 mg/kg
IBDQ remission	• yes • no
Fatigue response	• yes • no
Induction treatment	• guselkumab 400 mg IV • guselkumab 200 mg IV • placebo crossover to guselkumab 200 mg IV

In addition, the consistency of treatment effect for the primary and major secondary endpoints, histologic healing, histologic remission, histologic-endoscopic mucosal healing (Alternative definition 1-2), and deep histologic-endoscopic mucosal healing, will be evaluated for the following subgroups in a table summary:

Subgroup	Definition
Induction baseline body weight	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile

Subgroup	Definition
CRP (at Week 0 of the induction study and the maintenance study)	• ≤ 3 mg/L • > 3 mg/L
	• ≤ median • > median
Fecal calprotectin (at Week 0 of the induction study and the maintenance study)	• ≤ 250 mg/kg • > 250 mg/kg
Fecal calprotectin (at Week 0 of the induction study and the maintenance study)	• ≤ 1st quartile • > 1st quartile and ≤ 2nd quartile • > 2nd quartile and ≤ 3rd quartile • > 3rd quartile
ADT failure profile	• ADT-naïve • ADT-experienced but not failed • ADT-non failure • ADT-Failure
Biologic failure profile	• Biologic-naïve • Biologic-experienced without documented failure [i.e., discontinued biologic therapy for other reasons] • Biologic-non failure • Biologic-failure
Colonic molecular predictive signature (MPS) status at induction baseline	• Predicted endoscopic healing • Predicted endoscopic non-healing
Induction treatment	• guselkumab 400 mg IV • guselkumab 200 mg IV • placebo crossover to guselkumab 200 mg IV

5.11. Interim Analyses

No interim analysis is planned for this Maintenance Study.

5.11.1. Data Monitoring Committee (DMC) or Other Review Board

An external independent DMC has been established and will meet periodically to review interim unblinded safety data to ensure the continuing safety of the participants enrolled in the study. The DMC consists of 2 physicians and a statistician. The DMC responsibilities, authorities, and procedures are documented in a separate DMC charter.

5.12. Analyses in Long-term Extension

The objective of the LTE is to enable participants reaching Week M-44 of the maintenance study to continue to receive study agent without interruption and collect long-term safety and efficacy data. It is important to note that participants enter in the LTE is based on investigator determination as to whether the participants would benefit from continuation of treatment.

The summary of efficacy endpoints such as symptomatic remission, corticosteroid-free symptomatic remission, partial Mayo score, extraintestinal manifestations, etc., inflammatory biomarkers, HRQoL, health economics, safety, PK, and immunogenicity over time for the participants in LTE will be provided.

To provide more details about the analyses to be included for LTE, a separate document will be prepared.

6. SUPPORTING DOCUMENTATION

6.1. Appendix 1 List of Abbreviations

5-ASA	5-aminosalicylic
6-MP	6-mercaptopurine
ADT	advanced therapy
AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
AZA	azathioprine
BUN	blood urea nitrogen
CDT	Clinical Development Team
CDTL	Clinical Development Team Leader
CIOMS	Council for International Organizations of Medical Sciences
CMH	Cochran-Mantel-Haenszel
CRF	case report form
CRO	Contract Research Organization
CRP	C-reactive protein
C-SSRS	Columbia-Suicide Severity Rating Scale
DBL	database lock
DMC	Data Monitoring Committee
eCRF	electronic case report form
ECG	electrocardiogram
EP	erythrodermic psoriasis
EQ-5D-5L	5-level EuroQol five dimensions instrument
EU	European Union
FAS	full analysis set
GMS	Global Medical Safety
GPP	generalized pustular psoriasis
HRQoL	health-related quality of life
I	induction (as I-4 [induction week 4])
IB	investigator's brochure
IBDQ	Inflammatory Bowel Disease Questionnaire
ICE	Intercurrent event
IEC/IRB	Independent Ethics Committee/Institutional Review Board
IL-23	interlukin-23
IL-23R	interlukin-23 receptor
IPPI	Investigational Product Preparation Instructions
IPPM	Investigational Product Procedures Manual
IRB	Institutional Review Board
IV	intravenous
IVRS/IWRS	interactive voice/web response system
J&J	Johnson & Johnson
LLOQ	lower limit of quantification
LTE	long-term extension
M	Maintenance (as in M-44 [maintenance week 44]
MedDRA	Medical Dictionary for Regulatory Activities
mAb	monoclonal antibody
mAbs	monoclonal antibodies
MAR	missing at random
MMRM	mixed-effect model repeated measures
MTX	methotrexate
NAbs	neutralizing antibodies
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NONMEM	nonlinear mixed-effects modeling
PD	pharmacodynamics

PFS	prefilled syringe
PFS-U	UltraSafe Plus™ Passive Needle Guard
PGA	Physician's Global Assessment
PGIC	Patient's Global Impression of Change
PGIS	Patient's Global Impression of Severity
PK	pharmacokinetics
PPP	palmoplantar pustulosis
PRO	Patient-Reported Outcome
PROMIS-29	Patient-Reported Outcomes Measurement Information System
PsA	psoriatic arthritis
q4w	every 4 weeks
q8w	every 8 weeks
QC	quality control
R&D	Research and Development
RBS	rectal bleeding subscore
SAE	serious adverse event
SAP	Statistical analysis plan
SAS	Statistical Analysis System
SC	subcutaneous
SEC	US Securities and Exchange Commission
SMQ	Standardised MedDRA Queries
SMT	Safety Management Team
SoA	schedule of activities
SSG	Statistical Support Group
SUSAR	suspected unexpected serious adverse reaction
TEAE	treatment-emergent adverse event
TF	treatment failure
TNF	tumor necrosis factor
UC	ulcerative colitis
UCEIS	Ulcerative Colitis Endoscopic Index of Severity
US	United States
USB	Universal Serial Bus
VAS	Visual analog scale
WPAI-GH	Work Productivity and Activity Impairment Questionnaire-General Health

6.2. Appendix 2 Changes to Protocol-Planned Analyses

There were no changes to protocol-planned analyses, however, a few additional analyses which are not stated in the protocol are added in Section 5.5.

6.3. Appendix 3 Demographics and Baseline Characteristics

Table 6 presents a list of the demographic variables that will be summarized by treatment group, combined active treatment group, and overall, for the **FAS** and the **Enrolled Analysis Set**. A similar summary will be provided for the **Dose Adjustment Analysis Set**. Summaries of UC disease characteristics at induction baseline and maintenance baseline by ADT-failure status will be provided based on the **FAS**. In addition, the distribution of participants by region, country, and site ID will be presented based on the **FAS** and the **Enrolled Analysis Set** unless otherwise noted.

Table 6: Demographic Variables and Disease Characteristics

Continuous Variables:	**Summary Type**
Demographic Variables at Week I-0	Descriptive statistics (N, mean, SD, median and range [minimum and maximum], and IQ range).
Age (years)	
Weight (kg)	
Height (cm)	
Disease Characteristics at Week I-0	
UC disease duration (years)	
Mayo score	
Partial Mayo score	
Modified Mayo score	
CRP	
Fecal Calprotectin	
Albumin	
Disease Characteristics at Week M-0	
Mayo score	
Partial Mayo score	
Modified Mayo score	
CRP	
Fecal Calprotectin	
Categorical Variables	**Summary Type**
Demographic Variables at Week I-0	
Sex (male, female, unknown, undifferentiated)	Frequency distribution with the number and percentage of participants in each category.
Race [a] (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, White, Not Reported, Multiple)	
Ethnicity (Hispanic or Latino, not Hispanic or Latino, Not Reported)	
Region (Eastern Europe, Asia-Pacific, Rest of World)	
Age (< 65, ≥ 65)	
Disease Characteristics at Week I-0	
Severity of UC disease (moderate: 6 ≤ Mayo score ≤ 10, severe: Mayo score > 10)	
Mayo endoscopy subscore at baseline (Moderate: Subscore of 2, Severe: Subscore of 3)	
Modified Mayo score of 5 – 6 (moderate)	
Modified Mayo score of 7 – 9 (severe)	
Extraintestinal manifestations (absent, present)	
Abnormal CRP (> 3 mg/L)	
Abnormal fecal calprotectin (> 250 mg/kg)	
Fecal calprotectin (> 150 mg/kg)	
Extent of disease (limited, extensive)	

Tobacco or nicotine use status (non-user, prior user, current user)	
Disease Characteristics at Week M-0	
Abnormal CRP (> 3 mg/L)	
Abnormal fecal calprotectin (> 250 mg/kg)	
Fecal calprotectin (> 150 mg/kg)	
Endoscopic healing	
Endoscopic normalization	
Clinical remission status (Yes/No)	
IBDQ remission (Yes/No)	

[a]If multiple race categories are indicated, the Race is recorded as 'Multiple'

6.4. Appendix 4 Protocol Deviations

In general, the following list of major protocol deviations may have the potential to impact participants' rights, safety or well-being, or the integrity and/or result of the clinical study. Participants with major protocol deviations will be identified prior to database lock and the participants with major protocol deviations through M-44 will be summarized by category based on FAS.

- Study intervention administration deviations
- Prohibited concomitant medications deviations
- Withdrawal criteria met but not withdrawn
- Other

Participants having study intervention administration deviations will be summarized in more detail using sub-categories identified prior to unblinding (e.g., participant receives the incorrect study intervention or dose) using the **FAS**. Separate listings of participants in the **All Treated Analysis Set** who have any major protocol deviation, or who have study intervention administration deviations will be provided.

6.5. Appendix 5 Prior and Concomitant Medications

Prior and Concomitant medications will be coded using the [World Health Organization Drug Dictionary (WHO-DD)]. Prior medications are defined as any therapy used before the day of first dose (partial or complete) of study intervention. Concomitant medications are defined as any therapy used on or after the same day as the first dose of study intervention, including those that started before and continue on after the first dose of study intervention.

The proportion of participants who receive UC-specific concomitant medications (5-aminosalicylic acids [5-ASAs], oral corticosteroids, and immunomodulators i.e., 6-mercaptopurine [6-MP], azathioprine [AZA] or methotrexate [MTX]) will be summarized as well as the proportion of participants who receive at least 1 of these UC-specific concomitant medication based on the **FAS**, **All Treated Analysis Set**, and the **Dose Adjustment Analysis Set**.

History of response to or intolerance of Corticosteroids and Immunomodulators (i.e., 6-mercaptopurine [6-MP], azathioprine [AZA]) or history of response to advanced therapy (ADT; i.e., tumor necrosis factor alpha [TNFα] antagonists, vedolizumab, or tofacitinib) will be

summarized by treatment group based on the **FAS**, **All Treated Analysis Set**, and the **Dose Adjustment Analysis Set**. Prior medications (i.e., participants who took medications for UC and their length of exposure prior to the study) will be summarized by treatment group based on the **FAS**.

6.6. Appendix 6 Intervention Compliance

Compliance will be summarized descriptively by treatment group based on the **FAS** (including both randomized and nonrandomized participants) and the **Enrolled Analysis Set**. In addition, a listing of participants who were assigned treatment but were never treated and a listing of participants who were unblinded prior to Week M-44 during the maintenance study, based on the **Enrolled Analysis Set**, will be provided. A listing of participants, based on the **All Treated Analysis Set**, who received the wrong treatment will also be provided.

6.7. Appendix 7 Adverse Events of Special Interest and Other Adverse Events of Interest

AEs of Special Interest (AESI) and other events of interest are as follows:

- Active Tuberculosis (AESI)
- Malignancies (AESI)
- Opportunistic Infections
- Major Adverse Cardiovascular Events (MACE)
    - Cardiovascular death
    - Nonfatal myocardial infarction (MI)
    - Nonfatal stroke
- Hepatic Disorders
- Venous thromboembolism (VTE)
- Anaphylactic reactions
- Serum-sickness

6.8. Appendix 8 Medications of Special Interest

Not applicable.

6.9. Appendix 9 Laboratory Toxicity Grading

The toxicity grading scale used for assessment of clinical laboratory tests of interest is a modified Common Terminology Criteria for Adverse Events (CTCAE) v5.0', where toxicity grades are based on the laboratory result and do not take into account the clinical component, if applicable.

Hematology Tests		**Criteria**			
Test	**Direction**	**1**	**2**	**3**	**4**
Hemoglobin (g/dL)	Increase	>0 - 2 g/dL	>2 - 4 g/dL	>4 g/dL	
Hemoglobin (g/dL)	Decrease	<LLN - 10.0	<10.0 - 8.0	<8.0	
Lymphocytes (/mm3)	Increase		>4000 - 20,000	>20,000	
Lymphocytes (/mm3)	Decrease	<LLN - 800	<800 - 500	<500 - 200	<200
Neutrophils (/mm3)	Decrease	<LLN - 1500	<1500 - 1000	<1000 - 500	<500
Platelets (/mm3)	Decrease	<LLN - 75,000	<75,000 - 50,000	<50,000 - 25,000	<25,000
Total WBC count (/mm3)	Increase			>100,000	
Total WBC count (/mm3)	Decrease	<LLN - 3000	<3000 - 2000	<2000 - 1000	<1000
Chemistry Tests		**Criteria**			
Test	**Direction**	**1**	**2**	**3**	**4**
Albumin (g/L)	Decrease	≥30 - <LLN	≥20 - <30	<20	
Creatinine	Increase	>ULN - ≤1.5 xULN	>1.5 - 3.0 x baseline; >1.5 - 3.0 x ULN	>3.0 x baseline; >3.0 - 6.0 x ULN	>6.0 xULN
Potassium (mmol/L)	Increase	>ULN - ≤5.5	>5.5 - 6.0	>6.0 - 7.0	>7.0
Potassium (mmol/L)	Decrease		<LLN - 3.0	<3.0 - 2.5	<2.5
Sodium (mmol/L)	Increase	>ULN - 150	>150 - 155	>155 - 160	>160
Sodium (mmol/L)	Decrease	<LLN - 130		–<130 - 120	<120

6.10. Appendix 10 Liver Test Threshold

Analyte	**ULN Thresholds**
ALT or AST	> 1 x to < 3 x ULN ≥ 3 x to < 5 x ULN ≥ 5 x ULN to < 8 x ULN ≥ 8 x ULN
Total Bilirubin	> 1 to < 2 x ULN ≥ 2 x ULN
Alkaline Phosphatase	> 1 to < 2 x ULN ≥ 2 to < 4 x ULN ≥ 4 x ULN

6.11. Appendix 11 SAP Amendment History

The SAP Amendment Summary of Changes Table for the current amendment is located directly before the Table of Contents.

Amendment 1 (04 October 2022)
The main reasons for this amendment are:

- modification of the primary analysis population to include only randomized and treated participants with a modified Mayo score of 5 to 9 at induction baseline, per health authority request
- addition of supplementary analyses for the primary and major secondary endpoints as appropriate, utilizing an alternative calculation of the Mayo stool frequency subscore and the rectal bleeding subscore
- modification of the definition of ICE 5 to include the regional crisis in Russia and Ukraine intercurrent events. This ICE is now specified as discontinuation of study intervention due to COVID-19 related reasons (excluding COVID-19 infection) or regional crisis in Russia and Ukraine
- modification of ICE strategy for handling ICE 5 from hypothetical to treatment policy
- addition of sensitivity analyses which exclude the participants whose data cannot be Source Data Verified, due to major disruption, including COVID-19 and the regional crisis in Russia and Ukraine, to the primary and major secondary endpoints.
- modification of the US-specific testing procedure for the primary and major secondary endpoints

The table below includes all changes made in this amendment.

Section Number and Name	Description of Change	Brief Rationale
Throughout the SAP	Changed the duration of treatment during the LTE of the Maintenance Study from 2 years to approximately 4 years.	To extend the LTE of the Maintenance Study to approximately 4 years.
Section 1.2. Study Design Section 3. Sample Size Determination Section 4. Populations (Analysis Sets) for Analysis	The primary analysis population was updated to include only randomized and treated participants with a modified Mayo score of 5 to 9 at induction baseline. The targeted enrollment for the program was increased from approximately 950 participants to 1000 participants because the program will also enroll participants with a modified Mayo score of 4 (capped at ≤5%).	Per health authority request.
Section 4. Populations (Analysis Sets) for Analysis Section 5.2. Participant Disposition Section 5.5. Other Endpoint Analysis Section 5.9. Safety Analyses Section 5.10.1. Pharmacokinetics Section 5.10.2. Immunogenicity	Selected efficacy, safety, pharmacokinetics, and immunogenicity analyses for all randomized and treated participants (regardless of modified Mayo score) at induction baseline were added.	Analyses for all randomized and treated participants were added for completeness.
Section 5.3. Primary Endpoint Analysis Section 5.4. Major Secondary Endpoints Analysis Section 5.5. Other Endpoints Analysis	Supplementary analyses for the primary and major secondary endpoints, as appropriate, utilizing an alternative calculation of the Mayo stool frequency subscore and the rectal bleeding subscore were added.	Per health authority request.
Section 5.5. Other Endpoints Analysis	Additional histologic endpoints related to the Geboes scores, NHI and RHI were added. PRO endpoints related to the IBDQ urgency items were added.	Assess efficacy based on the Geboes scores, NHI, RHI and IBDQ urgency related item scores.
Section 5.10.6. Definition of Subgroups	Additional subgroups for UC-related ADT Medication History were added.	To add additional UC-related ADT Medication History subgroups.
Section 5.1. General Considerations Section 5.1.1. Visit Windows Section 5.3. Primary Endpoint Analysis Section 5.5. Other Endpoints Analysis Section 5.9. Safety Analyses Section 5.10.2. Immunogenicity Section 5.10.6. Definition of Subgroups	Some minor updates were made to clarify endpoints as well as related hypotheses and statistical analyses, and to correct editorial mistakes.	To add further clarification.

Section 5.3. Primary Endpoint Analysis Section 5.4. Major Secondary Endpoints Analyses Section 5.5.2. Estimands for the Other Endpoints	Modified the definition of ICE5 to include the regional crisis intercurrent events. This ICE is now specified as discontinuation of study intervention due to COVID-19 reasons (excluding COVID-19 infection) or regional crisis. Modified the ICE strategy for handling ICE5 from hypothetical to treatment policy.	To address health authority's comment and specify the ICE strategy for discontinuation of study intervention due to regional crisis.
Section 5.3. Primary Endpoint Analysis Section 5.4. Major Secondary Endpoints Analyses	Added sensitivity analyses to exclude the participants whose data cannot be Source Data Verified.	To assess impact of major disruption, including COVID-19 and regional crisis.
Section 5.3. Primary Endpoint Analysis Section 5.4. Major Secondary Endpoints Analyses	The missing data rule for alternative Mayo calculation was updated to consider the Mayo rectal bleeding and stool frequency subscores as missing if neither 3 consecutive days nor 4 nonconsecutive days are available.	Per health authority request.
Section 5.3.2.1.2. Subgroup Analyses Section 5.4.1.4. Subgroup Analyses	The subgroup analyses were updated to use rate (risk) difference instead of odds ratio as the summary statistics.	Per health authority request for other IBD programs.

ATTACHMENTS

Attachment 1 MAYO SCORE

Attachment Table 1: Mayo Scoring System for Assessment of Ulcerative Colitis Activity

Stool frequency [a] 0 = Normal number of stools for this patient 1 = 1-2 stools more than normal 2 = 3-4 stools more than normal 3 = 5 or more stools more than normal
Rectal bleeding [b] 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 passed
Findings of endoscopy 0 = Normal or inactive disease 1 = Mild disease (erythema, decreased vascular pattern, mild friability) 2 = Moderate disease (marked erythema, absent vascular pattern, friability, erosions) 3 = Severe disease (spontaneous bleeding, ulceration)
Physician's global assessment [c] 0 = Normal 1 = Mild disease 2 = Moderate disease 3 = Severe disease

[a] At the screening visit, each person indicates the number of stools he/she passed in a 24-hour period when in remission or before his/her UC diagnosis, thereby serving as his/her own control to establish the degree of abnormality of stool frequency.

[b] The daily bleeding score represents the most severe bleeding of the day.

[c] The physician's global assessment acknowledges the 3 other criteria, the patient's recall of abdominal discomfort and general sense of well-being, and other observations, such as physical findings and the patient's performance status.

Attachment 2 DETAILED PROHIBITED CHANGES IN UC MEDICATIONS RULES

Participants who had a prohibited change in UC medication described below are considered to have ICE 3:

- Initiation of prohibited medications or therapies as defined in the protocol (see protocol Section 6.5.2).
- Initiation of restricted medications (rectal 5-ASA compounds; parenteral or rectal corticosteroids, including budesonide and beclomethasone dipropionate).
- Increase in the dose of oral corticosteroids (excluding budesonide and beclomethasone dipropionate) > 5 mg/day (prednisone equivalent) above the maintenance baseline dose for more than 7 days after the maintenance baseline visit due to worsening of disease. This includes initiation of oral corticosteroids due to worsening of disease that lasts for more than 7 days after the maintenance baseline visit for participants who were not receiving oral corticosteroids at maintenance baseline.
- Increase in the dose of oral budesonide > 3 mg/day above the maintenance baseline dose for more than 7 days after the maintenance baseline visit due to worsening of disease. This includes initiation of oral budesonide due to worsening of disease that lasts for more than 7 days after the maintenance baseline visit for participants who were not receiving oral budesonide at maintenance baseline.
- Increase in the dose of oral beclomethasone dipropionate > 5 mg/day above the maintenance baseline dose for more than 7 days after the maintenance baseline visit due to worsening of disease. This includes initiation of oral beclomethasone dipropionate due to worsening of disease that lasts for more than 7 days after the maintenance baseline visit for participants who were not receiving oral beclomethasone dipropionate at maintenance baseline.
- Increase in the dose of oral corticosteroids (excluding budesonide and beclomethasone dipropionate) > 5 mg/day (prednisone equivalent) above the maintenance baseline dose for more than 28 days after the maintenance baseline visit due to reasons other than worsening of disease. This includes initiation of oral corticosteroids due to reasons other than worsening of disease that lasts for more than 28 days after the maintenance baseline visit for participants who were not receiving oral corticosteroids at maintenance baseline.
- Increase in the dose of oral budesonide > 3 mg/day above the maintenance baseline dose for more than 28 days after the maintenance baseline visit due to reasons other than worsening of disease. This includes initiation of oral budesonide due to reasons other than worsening of disease that lasts for more than 28 days after the maintenance baseline visit for participants who were not receiving oral budesonide at maintenance baseline.
- Increase in the dose of oral beclomethasone > 5 mg/day above the maintenance baseline dose for more than 28 days after the maintenance baseline visit due to reasons other than worsening of disease. This includes initiation of oral beclomethasone dipropionate due to reasons other than worsening of disease that lasts for more than 28 days after the maintenance baseline visit for participants who were not receiving oral beclomethasone dipropionate at maintenance baseline.
- Any switch among oral budesonide, oral beclomethasone dipropionate or other oral corticosteroids (excluding prednisone equivalent changes) due to worsening of disease.

- Initiation of oral 5-ASA compounds due to worsening of disease.
- Increase above maintenance baseline in the dosage of oral 5-ASA compounds due to worsening of disease.
- Change from one oral 5-ASA compound to another 5-ASA compound due to worsening of disease.
- Initiation of 6-MP/AZA/MTX due to worsening of disease.
- Increase above maintenance baseline in the dosage of 6-MP/AZA/MTX due to worsening of disease.
- Any switch between 6-MP/AZA and MTX due to worsening of disease.

Attachment 3 Grading Criteria for the histological evaluation of Disease Activity in Ulcerative Colitis (Geboes grading system3)

Grade 0	Structural (architectural change)
Subgrades	
0.0	No abnormality
0.1	Mild abnormality
0.2	Mild or moderate diffuse or multifocal abnormalities
0.3	Severe diffuse or multifocal abnormalities
Grade 1	Chronic inflammatory infiltrate
Subgrades	
1.0	No increase
1.1	Mild but unequivocal increase
1.2	Moderate increase
1.3	Marked increase
Grade 2	Lamina propria neutrophils and eosinophils
2A Eosinophils	
2A. 0	No increase
2A.1	Mild but unequivocal increase
2A.2	Moderate increase
2A.3	Marked increase
2B Neutrophils	
2B. 0	None
2B.1	Mild but unequivocal increase
2B.2	Moderate increase
2B.3	Marked increase
Grade 3	Neutrophils in epithelium
3.0	None
3.1	< 5% crypts involved
3.2	< 50% crypts involved
3.3	> 50% crypts involved
Grade 4	Crypt destruction
4.0	None
4.1	Probable—local excess of neutrophils in part of crypt
4.2	Probable—marked attenuation
4.3	Unequivocal crypt destruction
Grade 5	Erosion or ulceration
5.0	No erosion, ulceration, or granulation tissue
5.1	Recovering epithelium+adjacent inflammation
5.2	Probable erosion—focally stripped
5.3	Unequivocal erosion
5.4	Ulcer or granulation tissue

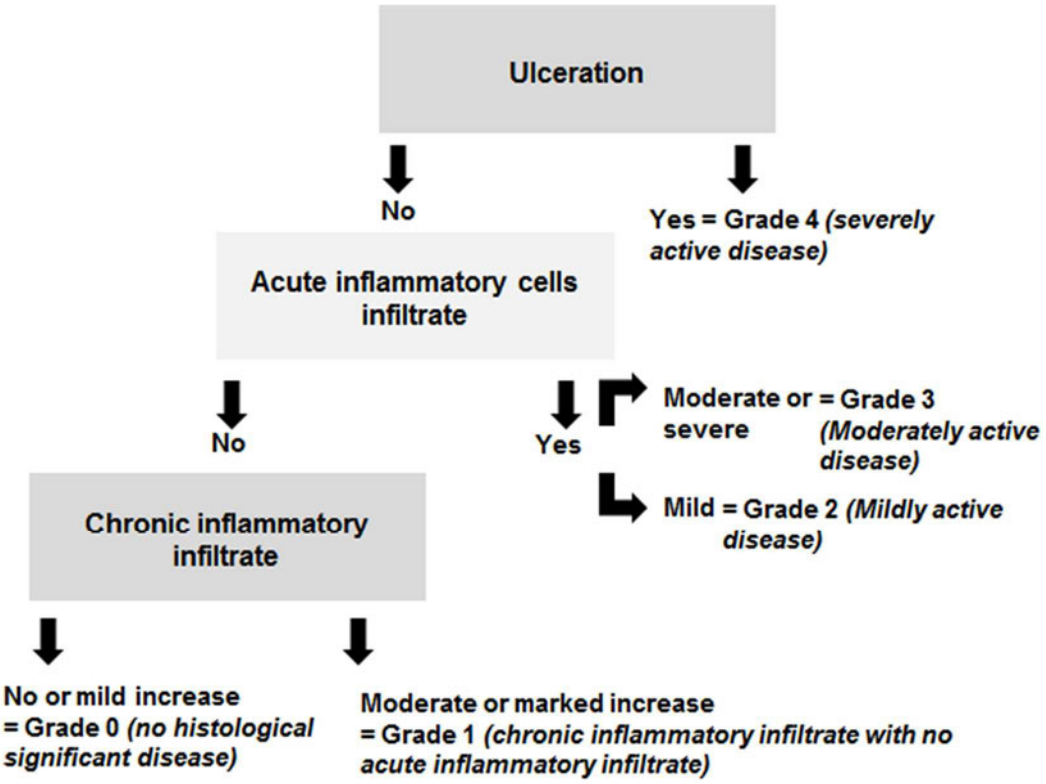
Attachment 4 UCEIS

The UCEIS is an index that provides an overall assessment of endoscopic severity of UC based upon mucosal vascular pattern, bleeding, and ulceration. The score ranges from 3 to 11. The UCEIS score will be assessed only by the central readers for all endoscopies received.

Attachment Table 2: UCEIS Descriptors and Definitions

Descriptor (score most severe lesions)	Likert scale anchor points	Definition
Vascular pattern	Normal (1)	Normal vascular pattern with arborization of capillaries clearly defined, or with blurring or patchy loss of capillary margins
	Patchy obliteration (2)	Patchy obliteration of vascular Pattern
	Obliterated (3)	Complete obliteration of vascular pattern
Bleeding	None (1)	No visible blood
	Mucosal (2)	Some spots or streaks of coagulated blood on the surface of the mucosa ahead of the scope, which can be washed away
	Luminal mild (3)	Some free liquid blood in the Lumen
	Luminal moderate or severe (4)	Frank blood in the lumen ahead of endoscope or visible oozing from mucosa after washing intraluminal blood, or visible oozing from a haemorrhagic mucosa
Erosions and ulcers	None (1)	Normal mucosa, no visible erosions or ulcers
	Erosions (2)	Tiny (# 5mm) defects in the mucosa, of a white or yellow color with a flat edge
	Superficial ulcer (3)	Larger (>5 mm) defects in the mucosa, which are discrete fibrin-covered ulcers in comparison with erosions, but remain superficial
	Deep ulcer (4)	Deeper excavated defects in the mucosa, with a slightly raised edge

Attachment 5 NANCY HISTOLOGIC INDEX



Attachment 6 ROBARTS HISTOLOGIC INDEX

Component

Chronic inflammatory infiltrate

0=No increase
1=Mild but unequivocal increase
2=Moderate increase
3=Marked increase

Lamina propria neutrophils

0=None
1=Mild but unequivocal increase
2=Moderate increase
3=Marked increase

Neutrophils in epithelium

0=None
1=<5% crypts involved
2=<50% crypts involved
3=>50% crypts involved

Erosion or ulceration

0=No erosion, ulceration or granulation tissue
1=Recovering epithelium+adjacent inflammation
1=Probable erosion—focally stripped
2=Unequivocal erosion
3=Ulcer or granulation tissue

RHI = 1 x chronic inflammatory infiltrate level (4 levels)
+ 2 x lamina propria neutrophils (4 levels)
+ 3 x neutrophils in epithelium (4 levels)
+ 5 x erosion or ulceration (4 levels)

Attachment 7 10 ITEMS IN THE IBDQ BOWEL DOMAIN

Question Number	Question
01	How frequent have your bowel movements been during the last two weeks?
05	How much of the time during the last two weeks have your bowel movements been loose?
09	How often during the last two weeks have you been troubled by cramps in your abdomen?
13	How often during the last two weeks have you been troubled by pain in the abdomen?
17	Overall, in the last two weeks, how much of a problem have you had with passing large amounts of gas?
20	How much of the time during the last two weeks have you been troubled by a feeling of abdominal bloating?
22	How much of the time during the last two weeks have you had a problem with rectal bleeding with your bowel movements?
24	How much of the time during the last two weeks have you been troubled by a feeling of having to go to the bathroom even though your bowels were empty?
26	How much of the time during the last two weeks have you been troubled by accidental soiling of your underpants?
29	How much of the time during the last two weeks have you been troubled by nausea or feeling sick to your stomach?

7. REFERENCE

1 Truelove SC, Witts LJ. Cortisone in ulcerative colitis: final report on a therapeutic trial. Brit Med J. 1955;2:1041-1048.
2 Baron JH, Connell AM, Lennard-Jones JE. Variation between observers in describing mucosal appearances in proctocolitis. Brit Med J. 1964;1:89-92.
3 Geboes K, Riddell R, Öst A, et al. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut. 2000;47:404-409.
4 Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, Bastien C, Cahn V, Cadiot G, et al. Development and validation of the Nancy histological index for UC. Gut 2017;66(1):43-9
5 Mosli MH, Feagan BG, Zou G, Sandborn WJ, D'Haens G, Khanna R, et al. Development and validation of a histological index for UC. Gut 2017;66(1):50-8
6 Rutgeerts P, Sandborn WJ, Feagan BG et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med. 2005;353(23):2462-2476
7 Sandborn WJ, Feagan B, Marano C, et al. Subcutaneous golimumab induces clinical response and remission in patients with moderate-to-severe ulcerative colitis. Gastroenterology. 2014;146(1):85-95.
8 Telesco SE, Brodmerkel C, Zhang H, et al. Gene Expression Signature for Prediction of Golimumab Response in a Phase 2a Open-Label Trial of Patients With Ulcerative Colitis. Gastroenterology. 2018;155(4):1008-1011.
9 Travis SP, Schnell D, Krzeski P, et al. Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS). Gut. 2012;61:535-542
10 Solem CA, Loftus EV, Tremaine WJ, et al. Correlation of C-reactive protein with clinical, endoscopic, histologic, and radiographic activity in inflammatory bowel disease. Inflamm Bowel Dis. 2005;11(8):707-712.
11 Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker of inflammatory bowel disease. Inflamm Bowel Dis. 2004;10(5):661-665.
12 Abraham BP and Kane S. Fecal Markers: Calprotectin and Lactoferrin. Gastroenterol Clin N Am. 2012;41:483-495.
13 Irvine EJ, Feagan B, Rochon J, et al. Quality of life: a valid and reliable measure of therapeutic efficacy in the treatment of inflammatory bowel disease. Canadian Crohn's Relapse Prevention Trial Study Group. Gastroenterology. 1994;106(2):287-296.
14 Higgins PD, Schwartz M, Mapili J, Krokos I, Leung J, Zimmermann EM. Patient defined dichotomous end points for remission and clinical improvement in ulcerative colitis. Gut. 2005;54(6):782-788.
15 Mundt, JC, Greist, JH, Jefferson, JW, et al. Prediction of suicidal behavior in clinical research by lifetime suicidal ideation and behavior ascertained by the electronic Columbia-Suicide Severity Rating Scale. J Clin Psych. 2013;74(9):887-893.
16 Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: Initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry. 2011;168(12):1266-1277
17 Hays RD, Spritzer KL, Schalet B & Celia D (2018). PROMIS®-29 v2.0 physical and mental health summary scores. Quality of Life Research, 27 (7): 1885-1891.