STATISTICAL ANALYSIS PLAN

Study Title:	Combined Phase 2b/3, Double-Blinded, Randomized, Placebo-Controlled Studies Evaluating the Efficacy and Safety of Filgotinib in the Induction and Maintenance of Remission in Subjects with Moderately to Severely Active Ulcerative Colitis
Name of Test Drug:	Filgotinib
Study Number:	GS-US-418-3898
Protocol Version (Date):	Original: 15 July 2016 Amendment 1: 07 September 2016 Amendment 2: 27 October 2016 Amendment 3: 15 June 2017 Amendment 4: 05 March 2018 Amendment 5: 02 April 2019
Analysis Type:	Final Analysis
Analysis Plan Version:	3.0
Analysis Plan Date:	28 Apr 2020
Analysis Plan Author(s):	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF ABBREVIATIONS

AE	adverse event
AEIs	adverse events of interest
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical drug class
BLQ	below the limit of quantitation
BMI	body mass index
CI	confidence interval
CMH	Cochran-Mantel-Haenszel
CRF	case report form
CSR	clinical study report
CTCAE	Common Toxicity Criteria for Adverse Events
DMC	data monitoring committee
EBS	endoscopy/bleeding/stool
ECG	electrocardiogram
E-Day	endoscopy day
EQ-5D	EuroQoL (health-related quality of life questionnaire)
ET	early termination
FAS	Full Analysis Set
FDA	Food and Drug Administration
Gilead	Gilead Sciences, Inc.
HCRU	health care resource utilization
HLGT	high-level group term
HLT	high-level term
HRQoL	health-related quality of life
hs-CRP	high-sensitivity C-reactive protein
IBD	inflammatory bowel disease
IBDQ	Inflammatory bowel disease questionnaire
ID	identification
IVRS	interactive voice response system
IWRS	interactive web response system
LLOQ	lower limit of quantitation
LLT	lower-level term
LOCF	last observation carried forward
LOQ	limit of quantitation
LTE	long-term extension
MCS	Mayo Clinic Score composed of subscores from endoscopy, rectal bleeding, stool frequency, and physician's global assessment

M-Day	maintenance study day
MedDRA	Medical Dictionary for Regulatory Activities
MST	MedDRA search term
MMP-9	matrix metalloproteinase-9
OL	open label
Partial MCS	all components of MCS except for endoscopic subscore
PD	pharmacodynamics
PGA	Physician's Global Assessment
PK	pharmacokinetic
PP	Per Protocol
PT	preferred term
PTM	placebo-to-match
Q1, Q3	first quartile, third quartile
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SE	standard error
SF-36	short-form 36 health survey
SMQ	Standardised MedDRA Queries
SOC	system organ class
SOP	standard operation procesdure
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TNFα	tumor necrosis factor-alpha
UC	ulcerative colitis
UCEIS	ulcerative colitis endoscopic index of severity
ULN	upper limit of normal
US	United States
V-Day	visit day
WHO	World Health Organization
WPAI	Work Productivity and Activity Impairment

PHARMACOKINETIC ABBREVIATIONS

AUC_{last}	the area under the concentration versus time curve from time zero to the last quantifiable concentration
AUC_{tau}	area under the concentration versus time curve over the dosing interval
C_{max}	the maximum observed serum/plasma/peripheral blood mononuclear (PBMC) concentration of drug
C_{tau}	observed drug concentration at the end of the dosing interval
CL_{ss}/F	apparent oral clearance after administration of the drug: at steady state: $CL_{ss}/F = Dose/AUC_{tau}$, where "Dose" is the dose of the drug
T_{max}	the time (observed time point) of C_{max}

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) in the clinical study report (CSR) for Study Protocol GS-US-418-3898. This SAP is based on the study protocol amendment 5 dated 02 April 2019 and the electronic case report form (eCRF). The SAP will be approved before database finalization of the final analysis. Any changes made after the finalization of the SAP will be documented in the CSR.

1.1. Study Objectives

The overall objective of the study is to evaluate the effect of treatment with filgotinib on the induction and maintenance of remission in subjects with moderately to severely active ulcerative colitis (UC). Subject who are biologic-naïve and biologic experienced will be enrolled in Cohort A and Cohort B respectively. Treatment assignments will be randomized within each cohort.

1.1.1. Primary Objectives

The primary objectives are:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- To evaluate the efficacy of filgotinib as compared to placebo in establishing endoscopy/bleeding/stool (EBS) remission at Week 10

Maintenance Study

- To evaluate the efficacy of filgotinib as compared to placebo in establishing EBS remission at Week 58

1.1.2. Secondary Objectives

The key secondary objectives are:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- To evaluate the efficacy of filgotinib as compared to placebo in establishing Mayo Clinic Score (MCS) remission at Week 10
- To evaluate the efficacy of filgotinib as compared to placebo in establishing an endoscopic subscore of 0 at Week 10
- To evaluate the efficacy of filgotinib as compared to placebo in establishing Geboes histologic remission at Week 10
- To evaluate the efficacy of filgotinib as compared to placebo in establishing MCS remission (alternative definition) at Week 10

Maintenance Study

- To evaluate the efficacy of filgotinib as compared to placebo in establishing MCS remission at Week 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing sustained EBS remission at Week 58, defined as EBS remission at both Weeks 10 and 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing 6-month corticosteroid-free EBS remission at Week 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing an endoscopic subscore of 0 at Week 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing Geboes histologic remission at Week 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing MCS remission (alternative definition) at Week 58

The other secondary objectives are:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- To evaluate the safety and tolerability of filgotinib
- To assess the pharmacokinetic (PK) characteristics of filgotinib

Maintenance Study

- To evaluate the safety and tolerability of filgotinib
- To assess the PK characteristics of filgotinib

1.1.3. Exploratory Objectives

The exploratory objectives are:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- To evaluate the efficacy of filgotinib as compared to placebo in improving endoscopic appearance as determined by Ulcerative Colitis Endoscopic Index of Severity (UCEIS) scoring system at Week 10
- To evaluate the association of changes in systemic or localized inflammatory biomarkers (including but not limited to C-reactive protein [hs-CRP], fecal calprotectin, fecal lactoferrin, and fecal matrix metalloproteinase-9 [MMP-9]) with clinical outcomes

- To evaluate stool microbiome
- To characterize the association of host genetics and other markers with disease severity, disease progression, and treatment response to filgotinib
- To evaluate health-related quality of life (HRQoL)
- To evaluate the effect of filgotinib on health care resource utilization (HCRU)
- To evaluate the efficacy of filgotinib as compared to placebo in achieving novel histologic outcomes (eg, resolution of basal plasmacytosis)

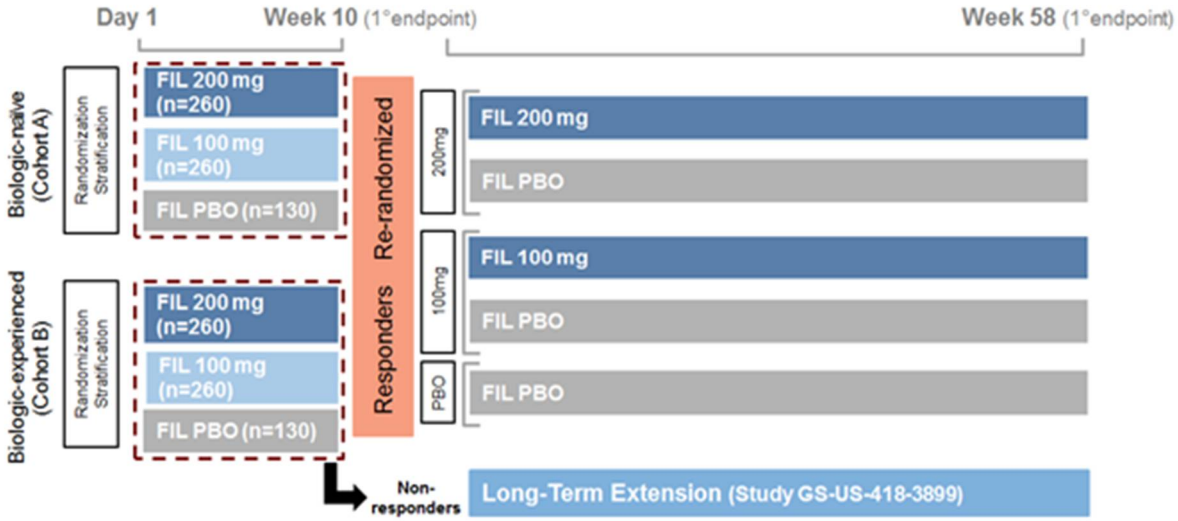
Maintenance Study

- To evaluate the efficacy of filgotinib as compared to placebo in improving endoscopic appearance as determined by UCEIS scoring system at Week 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing sustained MCS remission at Week 58, defined as MCS remission at both Weeks 10 and 58
- To evaluate the efficacy of filgotinib as compared to placebo in establishing 6-month corticosteroid-free MCS remission at Week 58
- To evaluate the association of changes in systemic or localized inflammatory biomarkers (including but not limited to hs-CRP, fecal calprotectin, fecal lactoferrin, and fecal MMP-9) with clinical outcomes
- To evaluate stool microbiome
- To characterize the association of host genetics and other markers with disease severity, disease progression, and treatment response to filgotinib
- To evaluate HRQoL
- To evaluate the effect of filgotinib on HCRU
- To evaluate the efficacy of filgotinib as compared to placebo in achieving novel histologic outcomes (eg, resolution of basal plasmacytosis)

1.2. Study Design

These are combined Phase 2b/3, double-blind, randomized, placebo-controlled studies evaluating the efficacy and safety of filgotinib in the induction and maintenance of remission in subjects with moderately to severely active ulcerative colitis (UC). A schema for this study is in Figure 1-1.

Figure 1-1. Study Schema



FIL filgotinib; PBO placebo; mg milligram.
Non responders are subjects who achieve neither EBS remission nor MCS response at Week 10
Subjects in the Maintenance Study who meet disease worsening criteria will be offered open label filgotinib.

1.2.1. Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

Based on protocol eligibility criteria, subjects will be screened for enrollment in either Cohort A or Cohort B. Subjects who meet protocol eligibility criteria will be assigned to the respective cohort and subsequently randomized in a blinded fashion in a 2:2:1 ratio to 1 of 3 treatments as follows:

Treatment Groups (Induction Studies)

Treatment Group 1 (n 260): Filgotinib 200 mg and placebo-to-match (PTM) filgotinib 100 mg, once daily

Treatment Group 2 (n 260): Filgotinib 100 mg and PTM filgotinib 200 mg, once daily

Treatment Group 3 (n 130): PTM filgotinib 200 mg and PTM filgotinib 100 mg, once daily

Note: US and Korea males who have not failed to at least 2 biologic regimens (any TNFα antagonist and vedolizumab) will be randomized in a 2:1 ratio to receive either filgotinib 100 mg or matching placebo.

Within each cohort, treatment assignments will be stratified according to the following factors:

Stratification Factors (Cohort A, Biologic Naïve Induction Study)

- Concomitant use of oral, systemically absorbed corticosteroids (eg, prednisone) at Day 1 (Yes or No)
- Concomitant use of immunomodulators (eg, 6-mercaptopurine [6-MP], azathioprine, methotrexate [MTX]) at Day 1 (Yes or No)

Stratification Factors (Cohort B, Biologic Experienced Induction Study)

- Exposure to 1 biologic agent versus more than 1 biologic agent
- Concomitant use of oral, systemically absorbed corticosteroids (eg, prednisone) at Day 1 (Yes or No)
- Concomitant use of immunomodulators (eg, 6-MP, azathioprine, MTX) at Day 1 (Yes or No)

1.2.2. Maintenance Study

Subjects in Cohorts A and B who complete the Induction Study and achieve either EBS remission or MCS response at Week 10 will be re-randomized into the Maintenance Study at Week 11.

Treatment Assignment in Induction Studies, Cohorts A and B	Maintenance Study Re-randomization:
Treatment 1, filgotinib 200 mg	Treatment 1, 200 mg
	Treatment 3, placebo
Treatment 2, filgotinib 100 mg	Treatment 2, 100 mg
	Treatment 3, placebo
Treatment 3, placebo	Continue Treatment 3, placebo

Note: Subjects receiving Treatment 1 or 2 in the Induction Study will be randomized in a 2:1 manner to either continue on the assigned filgotinib regimen or to placebo for the duration of the Maintenance Study.

Stratification Factors (Maintenance Study)

- Participation in Cohort A or Cohort B
- Concomitant use of oral, systemically absorbed corticosteroids (eg, prednisone) at Day 1 (Yes or No)
- Concomitant use of immunomodulators (eg, 6-MP, azathioprine, MTX) at Day 1 (Yes or No)

Subjects who achieve neither EBS remission nor MCS response at Week 10 will have the option to enter a separate, Long-Term Extension (LTE) study (GS-US-418-3899). The Maintenance Study will run from Weeks 11 to 58 with the primary efficacy assessment at Week 58. Subjects who opt out of the LTE study will return 30 days after the last dose of study drug for post-treatment safety assessments. Subjects who complete all procedures per protocol, including the endoscopy, of the 58-week study will be offered the option to continue into the LTE study. Subjects who are eligible and opt to participate in the LTE study can continue into the study without post-treatment safety assessments.

1.2.3. Pharmacokinetics Substudy

An optional PK substudy will be performed in a subset of subjects (approximately 30 subjects each in Cohort A and Cohort B) who provide separate informed consent. In the PK substudy, the daily dose of study drug should be administered under supervision in the clinic (at 1 visit between Week 2 and Week 10, inclusive), and additional PK samples should be collected predose and at 0.5, 1, 2, 3, 4, and 6 hours post dose.

1.2.4. Schedule of Assessments

Efficacy will be assessed primarily through MCS and partial MCS (all components of MCS except for endoscopic subscore). MCS will be assessed at Screening, Week 10, and Week 58. Partial MCS will be assessed at all scheduled visits.

For additional details, please see the Schedule of Assessments in Appendix 1.

1.3. Sample Size and Power

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

The sample size was chosen to ensure that a clinically meaningful difference in EBS remission rate at Week 10 could be detected when comparing filgotinib to placebo within each Induction Study. A sample size of 130 subjects in the placebo group and 260 subjects in each filgotinib dose (200 mg or 100 mg) group (N 650 per cohort) will provide 90% power for each filgotinib dose group comparison to placebo at a 2-sided 0.025 significance level to detect a treatment difference in EBS remission rate of 15% (25% on filgotinib and 10% on placebo).

Maintenance Study

Assuming an induction response rate (ie, proportion of subjects achieving either EBS remission or MCS response at Week 10) of 55% among subjects receiving filgotinib 200 mg or 100 mg treatment, approximately 285 subjects from each filgotinib dose group from Cohorts A and B Induction Studies combined would be eligible to be re-randomized into the Maintenance Study.

The sample size was chosen to ensure that a clinically meaningful difference in EBS remission rate at Week 58 could be detected when comparing each filgotinib dose group (200 mg or 100 mg) to placebo in the Maintenance Study. A sample size of 95 subjects in the placebo group and 190 subjects in the filgotinib group at the same dose level as the induction dose will provide more than 85% power for each filgotinib dose group comparison to placebo at a 2-sided 0.025 significance level to detect a treatment difference in maintenance EBS remission rate of 20% (40% on filgotinib and 20% on placebo).

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee (DMC) Analyses

The DMC's role and responsibilities and the scope of analysis to be provided to the DMC are described in a mutually agreed upon charter, which defines the DMC membership, meeting logistics, and meeting frequency.

2.1.1. Safety Analyses

An external multidisciplinary DMC will review the progress of the study and perform interim reviews of safety data in order to protect subject welfare and preserve study integrity. The DMC is to recommend to the sponsor whether the nature, frequency, and severity of adverse events (AEs) associated with study treatment warrant the early termination of the study in the best interests of study participants, whether the study should continue as planned, or whether the study should continue with modifications.

The initial meeting for each Induction Study will occur after approximately 100 subjects complete Week 10 or discontinue from the study.

After the futility analysis described in Section 2.1.2, routine DMC meetings for safety review will be held approximately once every 4 months or at a frequency determined by the DMC.

2.1.2. Futility Analysis

The second DMC meeting for each Induction Study will include an interim futility analysis and occur after approximately 175 subjects complete Week 10 or discontinue from the study. The futility analysis will be conducted to evaluate endoscopic efficacy. Overall safety will be evaluated at this timepoint as well. Please refer to Section 6.5.1 for details on the futility rule.

2.1.3. Cohorts A and B End-of-Induction Analysis

Efficacy and safety analyses will be performed after *all* subjects in both cohorts complete Week 10 or discontinue from the study. Please refer to Section 6.5.2 for details.

2.2. Final Analysis

After all subjects have completed the study, outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned and finalized, the study blind will be broken and the final analysis of the data will be performed.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

Analysis results will be presented using descriptive statistics. For categorical variables, the number and percentage of subjects in each category will be presented; for continuous variables, the number of subjects (n), mean, standard deviation (SD) or standard error (SE), median, first quartile (Q1), third quartile (Q3), minimum, and maximum will be presented.

By-subject listings will be presented for all subjects in the All Randomized Analysis Set. Data collected on log forms, such as AEs, will be presented in chronological order within subject. The treatment group to which subjects were randomized or initially assigned will be used in the listings.

3.1. Analysis Sets

Analysis sets define the subjects to be included in an analysis. Analysis sets and their definitions are provided in this section. The analysis set will be identified and included as a subtitle of each table, figure, and listing.

For each analysis set, the number and percentage of subjects eligible for inclusion, as well as the number and percentage of subjects who were excluded and the reasons for their exclusion, will be summarized by treatment group.

A listing of reasons for exclusion from analysis sets will be provided by subject.

3.1.1. All Randomized Analysis Sets

The All Randomized Analysis Set for each Induction Study (Cohort A Induction Study and Cohort B Induction Study) includes all subjects who were randomized on Day 1 into each corresponding study.

The All Randomized Analysis Set for the Maintenance Study includes all subjects who were re-randomized into the Maintenance Study (including subjects randomized to the placebo treatment group in the Induction Studies who continued on placebo treatment in the Maintenance Study).

The All Randomized Analysis Sets are the primary analysis sets for by-subject listings.

3.1.2. Full Analysis Set

The Full Analysis Set (FAS) for each Induction Study (Cohort A Induction Study and Cohort B Induction Study) includes all randomized subjects who took at least 1 dose of study drug in the corresponding Induction Study.

The FAS for the Maintenance Study includes all subjects randomized to either the filgotinib 200 mg or filgotinib 100 mg treatment groups in the Induction Studies who were re-randomized in the Maintenance Study and:

- Took at least 1 dose of study drug in the Maintenance Study, and
- Achieved EBS remission or MCS response at Week 10 as specified in the SAP

The Full Analysis Sets are the primary analysis sets for efficacy analyses.

3.1.3. Per-Protocol Analysis Sets

The Per-Protocol (PP) Analysis Set for each Induction Study (Cohort A Induction Study and Cohort B Induction Study) includes subjects in the respective FAS who meet the following criteria:

- Met the following key eligibility criteria:

  Documented diagnosis of UC of at least 6 months AND with a minimum disease extent of 15 cm from the anal verge. Documentation should include endoscopic and histopathologic evidence of UC as follows:

  - The criteria for documentation of UC based on endoscopy will be medical record documentation, or an ileocolonoscopy (full colonoscopy with the intubation of the terminal ileum) report dated ≥ 6 months before enrollment, which shows features consistent with UC, determined by the procedure performing physician
  - The criteria for documentation of UC based on histopathology will be medical record documentation of or a histopathology report indicating features consistent with UC as determined by the pathologist

  Moderately to severely active UC as determined by a centrally read endoscopy score ≥ 2, a rectal bleeding score ≥ 1, a stool frequency score ≥ 1, and PGA of ≥ 2 as determined by the Mayo clinic scoring system with endoscopy occurring during screening; total score must be between 6 and 12, inclusive.

- Dosed at least 80% of both study drugs (filgotinib and PTM) while on-treatment (see Section 4.2.2) for the Induction Study
- Either had sufficient data to evaluate EBS remission at Week 10, or met treatment failure criteria for Week 10 EBS remission outcome
- Never exposed to any biologics listed in Appendix 2 for Cohort A Induction Study, or exposed to at least 1 of the biologics listed in Appendix 2 for Cohort B Induction Study

The PP Analysis Set for the Maintenance Study includes subjects in the FAS who meet the following criteria:

- Met the key eligibility criteria from Induction Study as stated above
- Dosed at least 80% of each study drug type while on-treatment (see Section 4.2.2) for Maintenance Study
- Met one of following 3 categories:

  Had sufficient data to evaluate EBS remission outcome at Week 58, or

  Met treatment failure criteria for Week 58 EBS remission outcome, or

  Discontinued study drug due to protocol-specified disease worsening criterion

The PP Analysis Set is the secondary analysis set for efficacy analyses.

3.1.4. Safety Analysis Set

The Safety Analysis Set for each Induction Study (Cohort A Induction Study and Cohort B Induction Study) includes all subjects who took at least 1 dose of study drug in the corresponding Induction Study.

The Safety Analysis Set for the Maintenance Study includes all subjects who took at least one dose of study drug in the Maintenance Study.

The Overall Safety Analysis Set for the study includes all subjects who took at least one dose of study drug from either Induction or Maintenance.

The Safety Analysis Sets are the primary analysis sets for safety analyses.

3.1.5. Pharmacokinetic Analysis Sets

The primary analysis set for general PK analyses will be defined separately for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study). For each study, the PK analysis set includes all randomized subjects who took at least 1 dose of filgotinib and have at least 1 nonmissing concentration value for filgotinib and/or its metabolite GS-829845 by the PK laboratory.

3.1.6. Pharmacokinetic Substudy Analysis Set

The primary analysis set for intensive PK analyses will be the PK substudy analysis set for each Induction Study (Cohort A Induction Study and Cohort B Induction Study), which includes all randomized subjects from the corresponding Induction Study who took at least 1 dose of filgotinib, participated in the PK Substudy, and have at least 1 nonmissing intensive concentration value for filgotinib and/or its metabolite GS-829845 reported by the PK laboratory. This is the primary analysis set for detailed PK analysis of intensive PK sampling.

3.1.7. Biomarker Analysis Set

The Biomarker Analysis Set will be defined separately for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study). For each study, the Biomarker Analysis Set includes all subjects in the respective FAS who have nonmissing baseline and at least 1 postbaseline measurement from at least 1 of the 3 biomarkers (hs-CRP, fecal calprotectin, and fecal lactoferrin).

3.2. Subject Grouping

For analyses based on the All Randomized Analysis Sets and FAS, subjects will be grouped according to the treatment to which they were randomized. For analyses based on the Safety Analysis Sets, PP Analysis Sets, and Biomarker Analysis Set, subjects will be grouped according to actual treatment received. The actual treatment received will differ from the randomized treatment only when their actual treatment differs from randomized treatment for the entire treatment duration.

For the PK Analysis Sets, subjects will be grouped according to the actual treatment they received.

For the summary of AEs and laboratory abnormalities using the Overall Safety Analysis Set, subjects will be grouped according to the actual treatment received across Induction and Maintenance studies, if applicable.

In general, all data will be summarized by treatment group for:

- Cohort A Induction Study
- Cohort B Induction Study
- Cohort A and B Induction Studies combined (not applicable to efficacy analysis)
- Maintenance Study

3.3. Strata and Covariates

Subjects will be randomly assigned to treatment groups via the interactive web response system (IWRS) using a stratified randomization schedule into each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study). Details of the randomization ratio and stratification variables for each study are provided in Section 1.2.

If there are discrepancies in stratification factor values between the IWRS and the clinical database, the baseline values recorded in the clinical database will be used for analyses. For deriving concomitant medication use at Day 1, the start date of such medication should be before or on the same date of the first dose of study drug and with either the stop date of such medication being on or after the first dose of study drug or with 'ongoing' status.

Stratification factors will be used as covariates in evaluating efficacy endpoints, as specified in Section 6.

3.4. Examination of Subject Subgroups

Subgroup analyses for primary efficacy endpoint and adverse events are specified in Section 6.2.5 and Section 7.1.8, respectively.

3.5. Multiple Comparisons

The graphical approach {Bretz 2009} to sequentially rejective multiple test procedures will be used to control a family-wise type I error rate (FWER) at 5% (ie, α 0.05) for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study). This procedure strongly protects the FWER on all the primary and key secondary endpoints.

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

For each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study), a Bonferroni approach with equal alpha allocation of 0.025 (2-sided) to each filgotinib dose group comparison with placebo will be used to control the overall study-wide type I error rate at 0.05 within each study. To protect the integrity of the study due to the unblinded interim futility analysis planned for each induction study (Cohort A Induction Study and Cohort B Induction Study), an alpha of 0.00001 will be spent for each filgotinib dose group comparison to placebo within each induction study. As a result, a nominal p-value of less than 0.02499 (2-sided) will be needed to declare statistical significance for the final primary analysis of each filgotinib dose group when compared to placebo to control the study-wise type I error rate at 0.05. If one filgotinib dose group is discontinued for futility at the interim futility analysis, the overall alpha error rate stays unchanged at a 2-sided 0.02499 level for the continuing filgotinib dose group at the primary analysis.

The hypotheses to be tested for the induction studies are outlined below.

The primary null hypotheses to be tested:

- H1: The EBS remission rate in the filgotinib 200 mg group is equal to the EBS remission rate in the placebo group at Week 10
- H2: The EBS remission rate in the filgotinib 100 mg group is equal to the EBS remission rate in the placebo group at Week 10

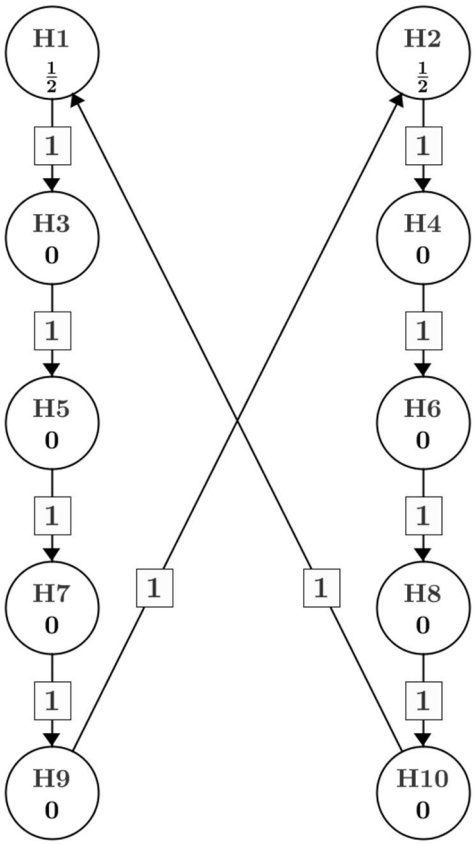
The key secondary null hypotheses to be tested:

- H3: The MCS remission rate in the filgotinib 200 mg group is equal to the MCS remission rate in the placebo group at Week 10
- H4: The MCS remission rate in the filgotinib 100 mg group is equal to the MCS remission rate in the placebo group at Week 10
- H5: The endoscopic subscore of 0 rate in the filgotinib 200 mg group is equal to the endoscopic subscore of 0 rate in the placebo group at Week 10
- H6: The endoscopic subscore of 0 rate in the filgotinib 100 mg group is equal to the endoscopic subscore of 0 rate in the placebo group at Week 10
- H7: The Geboes histologic remission rate in the filgotinib 200 mg group is equal to the Geboes histologic remission rate in the placebo group at Week 10
- H8: The Geboes histologic remission rate in the filgotinib 100 mg group is equal to the Geboes histologic remission rate in the placebo group at Week 10
- H9: The MCS remission (alternative definition) rate in the filgotinib 200 mg group is equal to the MCS remission (alternative definition) rate in the placebo group at Week 10
- H10: The MCS remission (alternative definition) rate in the filgotinib 100 mg group is equal to the MCS remission (alternative definition) rate in the placebo group at Week 10

If the primary null hypothesis is rejected, then the next key secondary hypothesis in the same filgotinib dosing regimen will be tested at the same alpha level. Testing of the hypotheses will be performed in the order outlined in Figure 3-1 (Induction Studies) in the same filgotinib dosing regimen. Once all hypotheses within the same filgotinib dosing regimen are rejected, then the respective 0.02499 alpha can be passed on to the other regimen's hypotheses, that is, all hypotheses in the other filgotinib dosing regimen will be tested at 0.04998 (2-sided) for Induction Studies. If an endpoint within a filgotinib dosing regimen fails to reach statistical significance, then formal sequential testing will stop and only nominal significance will be reported for the remaining endpoints within that filgotinib dosing regimen. If not all primary and key secondary hypotheses within the same filgotinib dosing regimen can be rejected, all hypotheses in the other filgotinib dosing regimen will still be tested at 0.02499 (2-sided) for Induction Studies.

A graphical illustration of the testing strategy for the primary and key secondary hypotheses in the Induction Studies is shown in Figure 3-1.

Figure 3-1. Induction Studies (Cohort A Induction Study and Cohort B Induction Study): Testing Strategy for the Primary and Key Secondary Hypotheses



Maintenance Study

Since there is no interim analysis planned for the maintenance study, the significance level for the final primary analysis will be at 0.025 (2-sided) level for each filgotinib dose group when compared to placebo for the maintenance study.

The hypotheses to be tested for the maintenance study are outlined below.

The primary null hypotheses to be tested:

- H1: The EBS remission rate in the filgotinib 200 mg group is equal to the EBS remission rate in the placebo group at Week 58
- H2: The EBS remission rate in the filgotinib 100 mg group is equal to the EBS remission rate in the placebo group at Week 58

The key secondary null hypotheses to be tested:

- H3: The 6-month corticosteroid-free EBS remission rate in the filgotinib 200 mg group is equal to the 6-month corticosteroid-free EBS remission rate in the placebo group at Week 58
- H4: The 6-month corticosteroid-free EBS remission rate in the filgotinib 100 mg group is equal to the 6-month corticosteroid-free EBS remission rate in the placebo group at Week 58
- H5: The sustained EBS remission rate in the filgotinib 200 mg group is equal to the sustained EBS remission rate in the placebo group at Week 58
- H6: The sustained EBS remission rate in the filgotinib 100 mg group is equal to the sustained EBS remission rate in the placebo group at Week 58
- H7: The MCS remission rate in the filgotinib 200 mg group is equal to the MCS remission rate in the placebo group at Week 58
- H8: The MCS remission rate in the filgotinib 100 mg group is equal to the MCS remission rate in the placebo group at Week 58
- H9: The endoscopic subscore of 0 rate in the filgotinib 200 mg group is equal to the endoscopic subscore of 0 rate in the placebo group at Week 58
- H10: The endoscopic subscore of 0 rate in the filgotinib 100 mg group is equal to the endoscopic subscore of 0 rate in the placebo group at Week 58
- H11: The Geboes histologic remission rate in the filgotinib 200 mg group is equal to the Geboes histologic remission rate in the placebo group at Week 58
- H12: The Geboes histologic remission rate in the filgotinib 100 mg group is equal to the Geboes histologic remission rate in the placebo group at Week 58

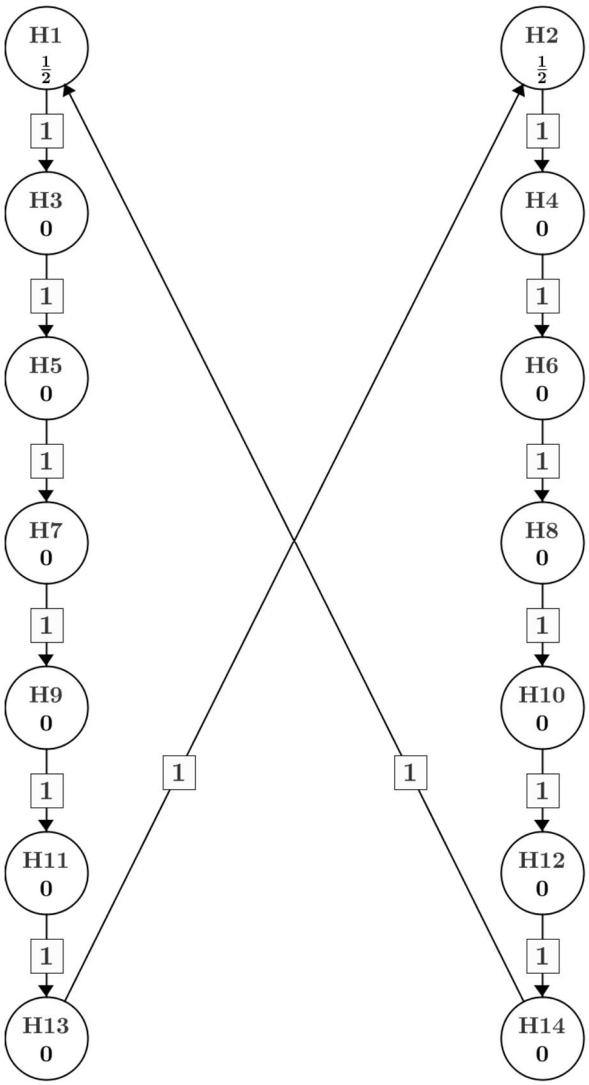
- H13: The MCS remission (alternative definition) rate in the filgotinib 200 mg group is equal to the MCS remission (alternative definition) rate in the placebo group at Week 58

- H14: The MCS remission (alternative definition) rate in the filgotinib 100 mg group is equal to the MCS remission (alternative definition) rate in the placebo group at Week 58

The same approach described for Induction Studies will be applied to the Maintenance Study at the alpha level of 0.025 (2-sided) for each filgotinib dose group when compared to placebo.

Graphical illustrations for the testing strategies for the primary and key secondary hypotheses in the Maintenance Study are shown in Figure 3-2.

Figure 3-2. Maintenance Study: Testing Strategy for the Primary and Key Secondary Hypotheses



3.6. Missing Data and Outliers

3.6.1. Missing Data

In general, missing data will not be imputed unless methods for handling missing data are specified. Exceptions are presented in this document.

For missing last dose date of study drug, imputation rules are described in Section 4.2.1. Imputation and calculation rules for missing patient diary data are described in Appendix 4. The handling of missing or incomplete dates for AE onset is described in Section 7.1.5.2 and for prior and concomitant medications in Section 7.4.

3.6.2. Outliers

Outliers will be identified during the data management and data analysis process. All data, including outliers, will be included in the data analyses, unless otherwise specified.

3.7. Data Handling Conventions and Transformations

In general, age (in years) on the date of the first dose of study drug will be used for analyses and presentation in listings. If an enrolled subject was not dosed with any study drug, the randomization date will be used instead of the first dosing date of study drug. For screen failures, the date the informed consent was signed will be used for age calculation. If only birth year is collected on the CRF, "01 July" will be used for the unknown birth day and month for the purpose of age calculation. If only birth year and month are collected, "15" will be used for the unknown birth day.

Duration of UC (years) is the number of years between the diagnosis date of UC and date of first dose of Induction study drug. The partial diagnosis date of UC (if any) will be imputed for calculation as follows:

- If day and month are missing but year is available, then the imputed day and month will be 01 Jan.

- If day is missing but the month and year are available, then the imputed day will be the first day of the month. Partial date will not be imputed if the year is missing.

Non-PK Data that are continuous in nature but are less than the lower limit of quantitation (LOQ) or above the upper LOQ will be imputed as follows:

- A value that is 1 unit less than the LOQ will be used for calculation of descriptive statistics if the datum is reported in the form of "< x" (where x is considered the LOQ). For example, if the values are reported as < 50 and < 5.0, values of 49 and 4.9, respectively, will be used for calculation of summary statistics. An exception to this rule is any value reported as < 1 or < 0.1, etc. For values reported as < 1 or < 0.1, a value of 0.9 or 0.09, respectively, will be used for calculation of summary statistics.

- A value that is 1 unit above the LOQ will be used for calculation of descriptive statistics if the datum is reported in the form of "> x" (where x is considered the LOQ). Values with decimal points will follow the same logic as above.

- The LOQ will be used for calculation of descriptive statistics if the datum is reported in the form of "≤ x" or "≥ x" (where x is considered the LOQ).

Natural logarithm transformation will be used for plasma concentrations and analysis of PK parameters. Plasma concentration values that are below the limit of quantitation (BLQ) will be presented as "BLQ" in the concentration data listing. Values that are BLQ will be treated as 0 at predose time points, and one-half the value of the LOQ at postbaseline time points.

The following conventions will be used for the presentation of summary and order statistics:

- If at least 1 subject has a concentration value of BLQ for the time point, the minimum value will be displayed as "BLQ."

- If more than 25% of the subjects have a concentration data value of BLQ for a given time point, the minimum and Q1 values will be displayed as "BLQ."

- If more than 50% of the subjects have a concentration data value of BLQ for a given time point, the minimum, Q1, and median values will be displayed as "BLQ."

- If more than 75% of the subjects have a concentration data value of BLQ for a given time point, the minimum, Q1, median, and Q3 values will be displayed as "BLQ."

- If all subjects have concentration data values of BLQ for a given time point, all order statistics (minimum, Q1, median, Q3, and maximum) will be displayed as "BLQ."

Pharmacokinetic parameters that are BLQ will be imputed as one-half LOQ before log transformation or statistical model fitting.

3.8. Analysis Visit Windows

3.8.1. Definitions

3.8.1.1. Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

The First Dosing Date of each Induction Study is defined as the date when subjects take the first dose of Induction study drug, as recorded in the Study Drug Administration eCRF.

The Last Dosing Date of each Induction Study is defined as the date when subjects take the last dose of Induction study drug as recorded in the Study Drug Administration eCRF.

Study Day will be calculated from the first dosing date of study drug and derived as follows:

- For days on or after first dosing date: Assessment Date First Dosing Date + 1
- For days prior to the first dosing date: Assessment Date First Dosing Date

Therefore, Study Day 1 is the day of the first dose of study drug administration.

Baseline is defined as the last available observation on or prior to the first dosing date, unless otherwise specified.

3.8.1.2. Maintenance Study

The First Dosing Date of Maintenance Study is defined as the date when subjects take the first dose of Maintenance Study drug, as recorded in the Study Drug Administration eCRF.

The Last Dosing Date of Maintenance Study is defined as the date when subjects take the last dose of Maintenance study drug as recorded in the Study Drug Administration eCRF.

Maintenance Study Day (M-Day) will be calculated from the First Dosing Date of Maintenance Study and derived as:

- For days on or after First Dosing Date of Maintenance Study: Assessment Date First Dosing Date + 1
- For days prior to First Dosing Date of Maintenance Study: Assessment Date First Dosing Date

Therefore, M-Day 1 is the first dosing date of the Maintenance Study.

Re-randomization baseline (henceforth referred to as re-baseline) is defined as the last available observation on or prior to the first dose date of the Maintenance Study. Re-baseline will be considered the baseline of Maintenance Study, unless otherwise specified.

3.8.1.3. Overall (Induction and Maintenance)

The First Dosing Date is defined as the date when subjects take the first dose of study drug.

The Last Dosing Date is defined as the date when subjects take the last dose of study drug considering both induction and maintenance study drugs.

Study Day will be calculated from the first dose date of study drug and derived as follows:

- For postdose study days: Assessment Date First Dosing Date + 1
- For days prior to the first dosing date: Assessment Date First Dosing Date

Therefore, Study Day 1 is the day of first dose of study drug administration.

Baseline is defined as the last available observation on or prior to the first dosing date.

3.8.2. Analysis Visit Windows

Subject visits may not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows.

For the Induction Studies (Cohort A Induction Study and Cohort B Induction Study), the analysis windows for endoscopic subscore and Geboes histologic score are provided in Table 3-1. The analysis windows for stool frequency, rectal bleeding, and PGA subscores are provided in Table 3-3. The analysis windows for HRQoLs are provided in Table 3-5. The analysis windows for weight, vital signs, electrocardiogram (ECG), safety laboratory parameters (hematology, chemistry, lipid profile, and serum immunoglobulin), and key inflammatory biomarkers (including hs-CRP, fecal calprotectin, and fecal lactoferrin) are based on study day and provided in Table 3-7.

For the Maintenance Study, the analysis windows for endoscopic subscore and Geboes histologic score are based on M-Day and provided in Table 3-2. The analysis windows for stool frequency, rectal bleeding, and PGA subscores are provided in Table 3-4. The analysis windows for HRQoLs are provided in Table 3-6. The analysis windows for weight, vital signs, ECG, safety laboratory parameters (hematology, chemistry, lipid profile, serum immunoglobulin), and key inflammatory biomarkers (including hs-CRP, fecal calprotectin, and fecal lactoferrin) are (defined in Section 3.8.1) provided in Table 3-8.

Table 3-1. Analysis Visit Windows for Induction Study Endoscopic Subscore, and Geboes Histologic Score

Nominal Visit	Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Screening	Baseline	1	(none)	1
Week 10	Week 10	71	58	minimum of 85 and Study Day of [Maintenance first dose date (if applicable)]

Note: Anchor day defined in Appendix 4 will be used for analysis visit window calculation of MCS and all efficacy endpoints based on MCS; endoscopic day will be used for analysis visit window calculation of endoscopic subscore.

Table 3-2. Analysis Visit Windows for Maintenance Study Endoscopic Subscore, and Geboes Histologic Score

Nominal Visit	Analysis Visit	Nominal Day (M-Day)	Lower Limit (M-Day)	Upper Limit (M-Day)
Week 10	Re-baseline	1	(none)	1
Week 58	Maintenance Week 47	330	303	358

Note: Anchor day defined in Appendix 4 will be used for analysis visit window calculation of MCS and all efficacy endpoints based on MCS; endoscopic day will be used for analysis visit window calculation of endoscopic subscore of 0 endpoint.

Table 3-3. Analysis Visit Windows for Induction Study Stool Frequency, Rectal Bleeding, and PGA subscores

Nominal Visit	Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Screening/Day 1	Baseline	1	(none)	1
Week 2	Week 2	15	2	22
Week 4	Week 4	29	23	36
Week 6	Week 6	43	37	57
Week 10	Week 10	71	58	minimum of 85 and Study Day of [Maintenance first dose date (if applicable)]

Note: The anchor date for stool frequency and rectal bleeding defined in Appendix 4 will be used to evaluate the analysis visit window. The nominal visit date will be used to evaluate the analysis visit window for PGA.

Table 3-4. Analysis Visit Windows for Maintenance Study Stool Frequency, Rectal Bleeding and PGA Subscores

Nominal Visit	Analysis Visit	Nominal Day (M-Day)	Lower Limit (M-Day)	Upper Limit (M-Day)
Week 10/Week 11	Re-Baseline	1	None	1
Week 14	Maintenance Week 3	22	2	43
Week 20	Maintenance Week 9	64	44	85
Week 26	Maintenance Week 15	106	86	134
Week 34	Maintenance Week 23	162	135	190
Week 42	Maintenance Week 31	218	191	246
Week 50	Maintenance Week 39	274	247	302
Week 58	Maintenance Week 47	330	303	358

Note: The anchor date for stool frequency and rectal bleeding defined in Appendix 4 will be used to evaluate the analysis visit window. The nominal visit date will be used to evaluate the analysis visit window for PGA.

Table 3-5. Analysis Visit Windows for Induction Study QoLs

Nominal Visit	Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Screening/Day 1	Baseline	1	(none)	1
Week 10	Week 10	71	58	minimum of 85 and Study Day of [Maintenance first dose date (if applicable)]

Table 3-6. Analysis Visit Windows for Maintenance Study QoLs

Nominal Visit	Analysis Visit	Nominal Day (M-Day)	Lower Limit (M-Day)	Upper Limit (M-Day)
Week 10	Re baseline	1	(none)	1
Week 26	Maintenance Week 15	106	2	218
Week 58	Maintenance Week 47	330	219	358

Table 3-7. Analysis Visit Windows for Induction Study Weight, Vital Signs, ECG, Hematology, Chemistry, Lipid Profile, Serum Immunoglobulin, hs-CRP, Fecal Lactoferrin and Fecal Calprotectin

Nominal Visit	Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Screening/Day 1	Baseline	1	(none)	1
Week 2	Week 2	15	2	22
Week 4	Week 4	29	23	36
Week 6	Week 6	43	37	57
Week 10	Week 10	71	58	For subjects who entered Maintenance Study: minimum of 85 and Study Day of [Maintenance first dose date]; for other subjects: ≥ 71

Note: ECG, lipid profile, fecal lactoferrin, and fecal calprotectin are collected postbaseline at Weeks 10, 26, and 58 only and the corresponding analysis windows will be applied; serum immunoglobulin is collected postbaseline at Weeks 4, 10, 26, and 58 only and the corresponding analysis windows will be applied.

Table 3-8. Analysis Visit Windows for Maintenance Study Weight, Vital Signs, ECG, Hematology, Chemistry, Lipid Profile, Serum Immunoglobulin, hs-CRP, Fecal Calprotectin and Fecal Lactoferrin

Nominal Visit	Analysis Visit	Nominal Day (M-Day)	Lower Limit (M-Day)	Upper Limit (M-Day)
Week 10/11	Re-Baseline	1	none	1
Week 14	Maintenance Week 3	22	2	43
Week 20	Maintenance Week 9	64	44	85
Week 26	Maintenance Week 15	106	86	134
Week 34	Maintenance Week 23	162	135	190
Week 42	Maintenance Week 31	218	191	246
Week 50	Maintenance Week 39	274	247	302
Week 58	Maintenance Week 47	330	303	≥ 330

Note: ECG, lipid profile, fecal lactoferrin and fecal calprotectin are collected postbaseline at Weeks 10, 26, and 58 only and the corresponding analysis windows will be applied; serum immunoglobulin is collected postbaseline at Weeks 4, 10, 26, and 58 only and the corresponding analysis windows will be applied.

3.8.3. Selection of Data in the Event of Multiple Records in an Analysis Visit Window

Depending on the statistical analysis method, single values may be required for each analysis window. For example, change from baseline by visit usually requires a single value, whereas a time-to-event analysis would not require 1 value per analysis window.

If multiple, valid, nonmissing, continuous measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- In general, the baseline/re-baseline value will be the last nonmissing value on or prior to the first dose date of study drug for each individual Induction study, and Maintenance Study, respectively, unless otherwise specified. If multiple measurements occur on the same day, the last nonmissing value prior to the first dose of study drug will be considered as the baseline value. If these multiple measurements occur at the same time or the time is not available, the average of these measurements (for continuous data) will be considered the baseline/re-baseline value.

- For postbaseline visits:

  The record closest to the nominal day for that visit will be selected.

  If there are 2 records that are equidistant from the nominal day, the later record will be selected.

  If there is more than 1 record on the selected day, the average will be taken, unless otherwise specified.

If multiple, valid, nonmissing, categorical measurements exist in an analysis window, records will be chosen based on the following rules if a single value is needed:

- For baseline/re-baseline, the last available record on or prior to the date of the first dose of study drug for each individual study (Cohort A Induction Study, Cohort B Induction Study and Maintenance Study) will be selected. If there are multiple records with the same time or no time recorded on the same day, the value with the lowest severity will be selected (eg, normal will be selected over abnormal for safety ECG findings).

- For postbaseline visits:

  The record closest to the nominal day for that visit will be selected.

  If there are 2 records that are equidistant from the nominal day, the later record will be selected.

  If there is more than 1 record on the selected day, the worst severity will be taken (eg, abnormal will be selected over normal for safety ECG findings), unless otherwise specified.

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

For each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study) and Cohorts A and B Induction Studies combined, a summary of subject enrollment will be provided by treatment group for each country, investigator within a country, and overall. The summary will present the number and percentage of subjects enrolled. For each column, the denominator for the percentage calculation will be the total number of subjects analyzed for that column.

For each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study), a similar enrollment table will be provided by randomization stratum. The denominator for the percentage of subjects in the stratum will be the total number of enrolled subjects. If there are discrepancies in the value used for stratification assignment between the IWRS and the clinical database, the value collected in the clinical database will be used for the summary. Subjects with discrepancies in the value used for stratification assignment between the IWRS and the clinical database at the time of data finalization will be summarized by treatment group. A corresponding listing will also be provided.

The randomization schedule used for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study) will be provided as an appendix to the CSR.

Cohort A Induction and Cohort B Induction Studies

A summary of subject disposition will be provided by treatment group for each individual study (Cohort A Induction Study, Cohort B Induction Study) and Cohorts A and B Induction Studies combined. This summary will present the number of subjects screened, the number of subjects who met all eligibility criteria but were not randomized with reasons subjects were not randomized, the number of subjects randomized, and the number of subjects in each of the categories listed below:

- Full Analysis Set
- Safety Analysis Set
- Per-Protocol Analysis Set
- PK Analysis Set
- PK Substudy Analysis Set
- Biomarker Analysis Set
- Completed study drug dosing through Week 10 (indicated in Induction Study Drug Completion CRF)

- Continuing study drug (for analysis other than the final analysis)
- Did not complete study drug up to Week 10 with reasons for premature discontinuation of study drug
- Completed Induction Study through Week 11
- Continuing Induction Study (for analysis other than the final analysis)
- Did not complete Induction Study with reasons for premature discontinuation from the study

Maintenance Study

A summary of subject disposition will be provided by treatment group. This summary will present the number of subjects who completed the Induction Studies, the number of subjects re-randomized, and the number of subjects in each of the categories listed below:

- Full Analysis Set
- Safety Analysis Set
- Per-Protocol Analysis Set
- PK Analysis Set
- Biomarker Analysis Set
- Completed study drug dosing through Week 58
- Continuing study drug prior to Week 58 (for analysis other than the final analysis)
- Did not complete study drug with reasons for premature discontinuation of study drug
- Completed Maintenance Study (defined as completion of protocol-planned duration of the study through Week 58)
- Continuing Maintenance Study (for analysis other than the final analysis)
- Did not complete Maintenance Study with reasons for premature discontinuation from the study

For the status of study drug and study completion and reasons for premature discontinuation, the number and percentage of subjects in each category will be provided. The denominator for the percentage calculation will be the total number of subjects in the Safety Analysis Set corresponding to that column in that study.

The following by-subject listings will be provided by subject identification (ID) number in ascending order to support the above summary tables:

- Reasons for premature study drug or study discontinuation
- Reasons for screen failure (will be provided by screening ID number in ascending order)
- Lot number and kit ID of assigned study medication

4.2. Extent of Study Drug Exposure and Adherence

Extent of exposure to study drug will be examined by assessing the total duration of exposure to study drug and the level of adherence to the study drug specified in the protocol.

Summaries of extent of study drug exposure and adherence will be provided by:

- Cohort A Induction Study
- Cohort B Induction Study
- Cohort A and B Induction Studies Combined
- Maintenance Study

4.2.1. Duration of Exposure to Study Drug

Total duration of exposure to study drug will be defined as last dosing date minus first dosing date plus 1, regardless of any temporary interruptions in study drug administration, and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks). If the last study drug dosing date is missing, the latest date among the study drug end date, clinical visit date, and laboratory sample collection date that occurred during the on-treatment period will be used.

The total duration of exposure to study drug will be summarized for the following using descriptive statistics and using the number (ie, cumulative counts) and percentage of subjects exposed through the following time periods:

- Cohort A Induction Study: 1 day, 2 weeks, 4 weeks, 6 weeks, 10 weeks, and 11 weeks
- Cohort B Induction Study: 1 day, 2 weeks, 4 weeks, 6 weeks, 10 weeks, and 11 weeks
- Cohort A and B Induction Studies Combined: 1 day, 2 weeks, 4 weeks, 6 weeks, 10 weeks, and 11 weeks
- Maintenance Study: 1 day, 3 weeks, 9 weeks, 15 weeks, 23 weeks, 31 weeks, 39 weeks, and 47 weeks

Summaries will be provided by treatment group for the Safety Analysis Set.

No formal statistical testing is planned.

4.2.2. Adherence to Study Drug

The total number of tablets administered will be summarized using descriptive statistics.

The presumed total number of tablets administered to a subject will be determined by the data collected on the drug accountability CRF using the following formula:

$$\text{Total Number of tablets Administered} = (\sum \text{No. of Tablets Dispensed}) - (\sum \text{No. of Tablets returned})$$

4.2.2.1. On-Treatment Adherence

The level of on-treatment adherence to the study drug regimen will be determined by the total amount of study drug administered relative to the total amount of study drug expected to be administered during a subject's actual on-treatment period based on the study drug regimen.

The level of on-treatment adherence will be expressed as a percentage using the following formula:

$$\text{On-Treatment Adherence (\%)} = \left(\frac{\text{Total Amount of Study Drug Administered}}{\text{Study Drug Expected to be Administered on Treatment}}\right) \text{x } 100$$

Descriptive statistics for the level of on-treatment adherence with the number and percentage of subjects belonging to adherence categories (ie, $< 80\%$, ≥ 80 to $< 90\%$, $\geq 90\%$) will be provided by treatment group for the Safety Analysis Set.

No formal statistical testing is planned.

A by-subject listing of study drug administration and drug accountability will be provided separately by subject ID number (in ascending order) and visit (in chronological order).

4.3. Protocol Deviations

Subjects who did not meet the eligibility criteria for study entry, but enrolled in the study will be summarized. The summary will present the number and percentage of subjects who did not meet at least 1 eligibility criterion and the number of subjects who did not meet specific criteria by treatment group based on the All Randomized Analysis Set. A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected.

Protocol deviations occurring after subjects entered the study are documented during routine monitoring. The number and percentage of subjects with important protocol deviations by deviation reason (eg, nonadherence to study drug, violation of select inclusion/exclusion criteria) will be summarized by treatment group for the All Randomized Analysis Set for each individual study. A by-subject listing will be provided for those subjects with any important protocol deviation.

5. BASELINE CHARACTERISTICS

5.1. Demographics and Other Baseline Characteristics

For the Induction Studies (Cohort A Induction Study and Cohort B Induction Study) and Cohorts A and B Induction Studies combined, subject demographic variables will be summarized by treatment group and overall using descriptive statistics for continuous variables, and using number and percentage of subjects for categorical variables. The summary of demographic data will be provided for the Safety Analysis Set.

- Age (years, on the date of first dose of study drug)
- Age group (< 65 years, ≥ 65 years)
- Sex at birth (female, male)
- Race
- Ethnicity (Hispanic or Latino, Not Hispanic or Latino)
- Geographic region (United States [US], non-US)
- Weight
- Height
- Body mass index (BMI; in kg/m^2)
- Smoking status (former, current, never)

For the Maintenance Study, the same demographic variables will be summarized by the induction dose and respective maintenance dose, and overall. Baseline age from the Induction Study will also be used for age in the Maintenance Study.

A by-subject demographic listing, including the informed consent date, will be provided by subject ID number in ascending order.

5.2. Baseline Disease Characteristics

For the Induction Studies (Cohort A Induction Study and Cohort B Induction Study), Cohorts A and B Induction Studies combined, and Maintenance Study, the following baseline disease characteristics will be summarized using Safety Analysis Sets by the same grouping as the demographic tables, using descriptive statistics for continuous variables, and using number and percentage of subjects for categorical variable:

- Duration of UC (in years) from date of diagnosis to first dosing date
- Duration of UC (< 1 year, ≥ 1 and < 3 years, ≥ 3 and < 7 years, ≥ 7 years)

- Have a history of pancolitis (yes, no, unknown)
- MCS score (Induction studies only)
- MCS score (≤ 8, ≥ 9) (Induction studies only)
- Partial MCS score (Induction studies only)
- Endoscopic subscore of 3 (yes, no) (Induction studies only)
- Participation in Cohort A or Cohort B (Maintenance Study Only)
- Fecal calprotectin (µg/g) (re-baseline value for Maintenance Study)
- Fecal calprotectin (≤ 250 µg/g, > 250 µg/g) (Induction studies only)
- hs-CRP (mg/L) (re-baseline value for Maintenance Study)
- hs-CRP (≤ 3 mg/L, > 3 mg/L) (Induction studies only)
- UC treatment history

  Prior use of systemic corticosteroids (yes, no)

  Prior use of immunomodulators (yes, no)

  Number of prior exposure to biologic agent listed in Appendix 2 (0, 1, 2, ≥ 3)

  Prior use of TNFα antagonist listed in Appendix 2 (yes, no; for Cohort B Induction Study and Cohorts A and B Induction Studies combined) and for subjects with prior use:

  - Number of TNFα antagonist used
    - 1
    - 2
    - ≥ 3
  - Worst outcome of prior use of TNFα antagonist
    - Treatment failure
    - Intolerance, including both allergic and non-allergic intolerance
    - Other

Prior use of vedolizumab (yes, no; for Cohort B Induction Study and Cohorts A and B Induction Studies combined) and for subjects with prior use:

- Worst outcome of prior use of vedolizumab
  - Treatment failure
  - Intolerance, including both allergic and non-allergic intolerance
  - Other

Prior use of both TNFα antagonist and vedolizumab (yes, no; for Cohort B Induction Study and Cohorts A and B Induction Studies combined)

Note: The worst outcome of a prior treatment is treatment failure, followed by intolerance, and then other outcomes.

Prior failure of both TNF-alpha antagonist and vedolizumab (for Cohort B Induction Study and Cohorts A and B Induction Studies combined)

- Yes (dual refractory, defined as those who have failed at least 2 classes of biologic therapies [any TNFα antagonist and vedolizumab])
  - US/Korea Males
  - Subjects other than US/Korea Males
- No
  - US/Korea Males
  - Subjects other than US/Korea Males

- Corticosteroid and Immunomodulator Treatment at baseline (at re-baseline for Maintenance Study)

  Concomitant use of systemically absorbed corticosteroids and/or immunomodulator at baseline/re-baseline, including: a) systemically absorbed corticosteroids only, b) immunomodulators only, c) both, and d) neither;

  Prednisone equivalent dose for subjects who are on systemically absorbed corticosteroids at baseline/re-baseline (mg/day)

  Prednisone equivalent dose for subjects who are on systemically absorbed corticosteroids at baseline/re-baseline (> 0 to 10 mg/day, > 10 to 20 mg/day, > 20 mg/day)

  5-aminosalicylates (yes, no)

In addition, another baseline disease characteristics table for Maintenance Study using the FAS will be generated, including the above baseline characteristics except those marked as "Induction studies only" and the following additional variables:

- MCS score at re-baseline
- MCS score at re-baseline ($\leq 8, \geq 9$)
- Partial MCS score at re-baseline
- Endoscopic subscore of 3 at week 10 (yes, no)
- Fecal calprotectin at re-baseline (≤ 250 μg/g, > 250 μg/g)
- hs-CRP at re-baseline (≤ 3 mg/L, > 3 mg/L)
- EBS remission at Week 10 (yes, no)
- MCS remission at Week 10 (yes, no)
- Endoscopic subscore of 0 at Week 10 (yes, no)
- Geboes histologic remission at Week 10 (yes, no)

5.3. Medical History

General medical history and IBD family history data will be collected at screening and listed only. General medical history data will be coded using the current version of Medical Dictionary for Regulatory Activities (MedDRA).

6. EFFICACY ANALYSES

6.1. General Considerations

The efficacy analysis will be conducted on the FAS, defined in Section 3.1.2, unless otherwise specified.

The definitions of the dichotomous efficacy endpoints for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study) are listed in Table 6-1. While the endoscopic appearance will be read by both study site investigator and central reader, only the centrally-read endoscopic subscore will be used in the calculations of Mayo Clinic Score (MCS) and other related efficacy endpoints unless otherwise stated.

Appendix 5 includes a detailed definition of study treatment failure and the corresponding data handling rules for efficacy analysis.

Table 6-1. Definitions of Dichotomous Efficacy Endpoints

Study	Type	Event Endpoint	Definition
Induction	Primary	EBS remission at Week 10	An endoscopic subscore of 0 or 1*, rectal bleeding subscore of 0, and at least a 1 point decrease in stool frequency from baseline to achieve a subscore of 0 or 1 at Week 10
Induction	Key Secondary	MCS remission at Week 10	A MCS of 2 or less and no single subscore higher than 1 at Week 10
Induction	Key Secondary	An endoscopic subscore of 0 at Week 10	An endoscopic subscore of 0 at Week 10
Induction	Key Secondary	Geboes histologic remission at Week 10	Based on the Geboes Scale (Appendix 6), all of the following must be met to be considered in Geboes histologic remission at Week 10: Grade 0 of ≤ 0.3, Grade 1 of ≤ 1.1, Grade 2a of ≤ 2A.3, Grade 2b of 2B.0, Grade 3 of 3.0, Grade 4 of 4.0, and Grade 5 of 5.0 {Geboes 2000}.
Induction	Key Secondary	MCS remission (alternative definition) at Week 10	Rectal bleeding, stool frequency, and PGA subscore of 0 and an endoscopic subscore of 0 or 1; overall MCS of ≤ 1 at Week 10
Induction	Exploratory	Novel histologic outcomes at Week 10	This novel histologic outcome incorporates both Geboes and non-Geboes criteria. Based on the Geboes Scale (Appendix 6), all of the following must be met to achieve the following novel, exploratory histologic outcome at Week 10: Grade 0 of ≤ 0.3, Grade 1 of ≤ 1.1, Grade 2a of 2A.0, Grade 2b of 2B.0, Grade 3 of 3.0, Grade 4 of 4.0 and Grade 5 of 5.0, along with the absence of basal plasmacytosis.
Induction	Exploratory	EBS remission (alternative definition) at Week 10	An endoscopic subscore of 0 or 1*, rectal bleeding subscore of 0, and stool frequency subscore of 0 or 1 at Week 10

Study	Type	Event Endpoint	Definition
Induction	Exploratory	MCS response at Week 10	An MCS reduction of ≥ 3 points and at least 30% from baseline score with an accompanying decrease in rectal bleeding subscore of ≥ 1 point or an absolute rectal bleeding subscore of 0 or 1 at Week 10
Induction	Exploratory	Endoscopic response at Week 10	An endoscopic subscore of 0 or 1 at Week 10
Maintenance	Primary	EBS remission at Week 58	An endoscopic subscore of 0 or 1* , rectal bleeding subscore of 0, and at least a 1 point decrease in stool frequency from induction baseline to achieve a subscore of 0 or 1 at Week 58
Maintenance	Key Secondary	6-month corticosteroid-free EBS remission at Week 58	EBS remission with no corticosteroid use for the indication of UC for at least 6 months prior to Week 58 among subjects who are on corticosteroid at re-baseline (baseline of maintenance study). Subjects who weaned off steroids but required re-initiation within 6 months prior to Week 58 assessment will be considered to have not met this endpoint.
Maintenance	Key Secondary	Sustained EBS remission at Week 58	EBS remission at both Week 10 and Week 58
Maintenance	Key Secondary	MCS remission at Week 58	An MCS of 2 or less and no single subscore larger than 1 at Week 58
Maintenance	Key Secondary	An endoscopic subscore of 0 at Week 58	An endoscopic subscore of 0 at Week 58
Maintenance	Key Secondary	Geboes histologic remission at Week 58	Based on the Geboes Scale (Appendix 6), all of the following must be met to be considered in Geboes histologic remission at Week 58: Grade 0 of ≤ 0.3, Grade 1 of ≤ 1.1, Grade 2a of ≤ 2A.3, Grade 2b of 2B.0, Grade 3 of 3.0, Grade 4 of 4.0, and Grade 5 of 5.0.
Maintenance	Key Secondary	MCS remission (alternative definition) at Week 58	Rectal bleeding, stool frequency, and PGA subscore of 0 and an endoscopic subscore of 0 or 1; overall MCS of ≤ 1 at Week 58
Maintenance	Exploratory	6-month corticosteroid-free MCS remission at Week 58	MCS remission with no corticosteroid use for the indication of UC for at least 6 months prior to Week 58 among subjects who are on corticosteroid at re-baseline
Maintenance	Exploratory	Sustained MCS remission at Week 58	MCS remission at both Week 10 and Week 58
Maintenance	Exploratory	Novel histologic outcomes at Week 58	This novel histologic outcome incorporates both Geboes and non-Geboes criteria. Based on the Geboes Scale (Appendix 6), all of the following must be met to achieve the following novel, exploratory histologic outcome at Week 58: Grade 0 of ≤ 0.3, Grade 1 of ≤ 1.1, Grade 2a of 2A.0, Grade 2b of 2B.0, Grade 3 of 3.0, Grade 4 of 4.0 and Grade 5 of 5.0, alone with the absence of basal plasmacytosis.

Study	Type	Event Endpoint	Definition
Maintenance	Exploratory	EBS remission (alternative definition) at Week 58	An endoscopic subscore of 0 or 1*, rectal bleeding subscore of 0, and stool frequency subscore of 0 or 1 at Week 58
Maintenance	Exploratory	MCS response at Week 58	An MCS reduction of ≥ 3 points and at least 30% from induction baseline score with an accompanying decrease in rectal bleeding subscore of ≥ 1 point or an absolute rectal bleeding subscore of 0 or 1 at Week 58
Maintenance	Exploratory	Endoscopic response at Week 58	An endoscopic subscore of 0 or 1 at Week 58

* Based on centrally read endoscopy.

MCS response will be used as one of the response criteria at Week 10 to decide whether subjects are eligible to be re-randomized into Maintenance Study.

Endoscopic response will be evaluated at the futility analysis at the second DMC meeting.

Summary of Dichotomous Efficacy Endpoints

For the Cohort A Induction Study, the Cohort B Induction Study, and the Maintenance Study, numbers and percentages of subjects achieving each of the dichotomous efficacy endpoints defined above, and numbers and percentages of subjects not achieving those endpoints for the following reasons will be summarized by treatment.

1) Subjects who are observed non-responders;

2) Subjects who are non-responders due to study treatment failure

3) Subjects who are non-responders due to study drug discontinuation led by protocol-specified disease worsening (maintenance study only)

4) Subjects who do not have sufficient measurements to determine the endpoint due to study drug discontinuation for other reasons

5) Subjects who do not have sufficient measurements to determine the endpoint while on study drug

Non-Responder Imputation (NRI)

For analysis of all the above defined efficacy endpoints, subjects who do not have sufficient measurements to determine the endpoint will be considered non-responders (ie, non-responder imputation [NRI]).

6.1.1. Calculations of Mayo Clinic Scores (MCS) and Partial MCS (pMCS) with the Subscores

The MCS is an instrument designed to measure disease activity in ulcerative colitis (Table 6-2) . The MCS system is a composite index of 4 disease activity variables. Each variable is scored individually on an integer scale of 0 to 3, inclusive, with higher scores indicating greater disease activity. The individual components of the MCS include stool frequency (SF), rectal bleeding (RB), endoscopic subscore, and the physician's global assessment (PGA). SF and RB are determined using an electronic daily diary, which collects subject reported components directly. The efficacy endpoints described in Table 6-1, except for histologic based endpoints, are all based on the four individual components. MCS is calculated as the sum of the 4 subscores, ranging from 0 to 12. A pMCS is calculated as the sum of the 3 subscores excluding the endoscopic subscore, ranging from 0 to 9. For further information on the MCS and calculation rules, reference is made to Appendix 4.

Table 6-2. Mayo Clinic Score

Variables	Score			
	0	**1**	**2**	**3**
Endoscopic Findings	Normal or inactive	Mild disease	Moderate disease	Severe disease
Rectal Bleeding	None	Streaks of blood in stool less than half the time	Obvious blood in stool half or more than half of the time	Blood alone passes
Stool Frequency	Normal number of daily stools for patient	1-2 stools per day more than normal	3-4 stools per day more than normal	≥ 5 stools per day more than normal
Physician Global Assessment	None	Mild disease	Moderate disease	Severe disease

6.2. Primary Efficacy Endpoints

6.2.1. Definition of the Primary Efficacy Endpoints

The primary efficacy endpoints are the proportion of subjects achieving EBS remission at Week 10 for the Induction Studies (Cohort A Induction Study and Cohort B Induction Study) and at Week 58 for Maintenance Study as defined in Table 6-1. MCS subscores collected at screening will be used as baseline values for evaluation of EBS remission at both Week 10 and Week 58.

6.2.2. Statistical Hypotheses for the Primary Efficacy Endpoints

The primary efficacy objective of the Induction Studies (Cohort A Induction Study and Cohort B Induction Study) and the Maintenance Study are to evaluate efficacy of filgotinib in establishing EBS remission at Weeks 10 and 58. These objectives are translated into the following primary statistical hypotheses at Weeks 10 and 58:

- H1**:** The EBS remission rate in the filgotinib 200 mg group is equal to the EBS remission rate in the placebo group.
- H2**:** The EBS remission rate in the filgotinib 100 mg group is equal to the EBS remission rate in the placebo group.

Each null hypothesis is tested against the alternative hypothesis that the EBS remission rate is different between the respective filgotinib dose group and the placebo group.

6.2.3. Analysis of the Primary Efficacy Endpoint

For the individual Induction Studies (Cohort A Induction Study and Cohort B Induction Study), a stratified Cochran-Mantel-Haenszel (CMH) test will be used to compare the treatment effect between the filgotinib 200 mg group and placebo and between the filgotinib 100 mg group and placebo separately. The CMH tests will be stratified by the following factors:

- Concomitant use of oral, systemically absorbed corticosteroids at Day 1
- Concomitant use of immunomodulators at Day 1
- Exposure to biologic agent as listed in Appendix 2 (≤ 1, > 1; for Cohort B Induction Study only)

For the Maintenance Study, a CMH test will be used to compare the treatment effect between filgotinib 200 mg and placebo among the subjects from the Cohort A and B Induction Studies combined being treated with filgotinib 200 mg. The CMH test will be stratified by the following factors:

- Participation in Cohort A or Cohort B
- Concomitant use of oral, systemically absorbed corticosteroids at re-baseline
- Concomitant use of immunomodulators at re-baseline

Similarly, a CMH test with the same stratification factors will be used to compare the treatment effect between filgotinib 100 mg and placebo among the subjects from the Cohort A and B Induction Studies combined being treated with filgotinib 100 mg.

The stratified CMH chi-square p-value will be provided for each of the above comparison. Strata with low numbers of subjects might be aggregated for the CMH test. The 2-sided 95% confidence interval (CI) of EBS remission rate based on normal approximation method with a continuity correction will be provided for each treatment group. In addition, non-stratified risk difference estimated along with its 2-sided 95% CI using the normal approximation (ie, the Wald method) with a continuity correction for the difference in proportions will be provided. Reference Section 3.5 for the significance level that will be used to declare a statistically significant treatment effect for each filgotinib group. Stratification variables based on the eCRF data will be used for the analysis.

Bar-charts of proportions of subjects achieving EBS remission at Week 10 and Week 58 will be provided by treatment group with corresponding p-values. In addition to the primary analysis method described above, the proportion of subjects in EBS remission who had a stool frequency subscore of 0 versus 1 will be calculated and summarized.

6.2.4. Sensitivity Analysis of the Primary Efficacy Endpoints

Sensitivity analyses of the primary efficacy endpoint will be performed for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study). For all of the sensitivity analyses, the same statistical method that was used for the primary analysis, will be used for the treatment comparison.

6.2.4.1. Per-Protocol Analyses

To evaluate the impact of study conduct on the primary analysis, the primary endpoint of EBS remission rates of each individual study will be analyzed based on the corresponding PP analysis sets as defined in Section 3.1.3.

6.2.4.2. Locally Read Endoscopic Subscore Analyses

To evaluate the potential disparity between centrally read endoscopy scores versus locally read scores, the EBS remission rates using locally (ie, investigator) read endoscopic subscore will be analyzed based on the FAS.

6.2.4.3. Missing Data Imputation Analyses

Subjects who do not meet EBS remission due to study treatment failure or due to meeting protocol-specified disease worsening criterion will not be considered as missing in this section, but will be considered as not achieving EBS remission for the sensitivity analysis listed here.

To explore the distribution of the missing data across cohorts and treatment groups, missing EBS, MCS, and pMCS at Week 10 for Induction Studies and at Week 58 for Maintenance Study by baseline/re-baseline characteristic and disease status (severity/duration) including variables listed from below, in each treatment group will be summarized.

Cohort A Induction Study at Baseline

- Duration of UC (years)
- MCS score
- Endoscopic subscore of 3 (yes, no)
- Concomitant use of oral, systemically absorbed corticosteroids (yes, no)
- Concomitant use of immunomodulators (yes, no)

Cohort B Induction Study at Baseline

- Duration of UC (years)
- MCS score
- Endoscopic subscore of 3 (yes, no)
- Concomitant use of oral, systemically absorbed corticosteroids (yes, no)
- Concomitant use of immunomodulators (yes, no)
- Exposure to biologic agent listed in Appendix 2 (≤ 1, > 1)

Maintenance Study

- Duration of UC at baseline (years)
- MCS score at re-baseline
- Endoscopic subscore of 3 (yes, no) at re-baseline
- Concomitant use of oral, systemically absorbed corticosteroids at re-baseline (yes, no)
- Concomitant use of immunomodulators at re-baseline (yes, no)
- Participation in Cohort A or Cohort B

To evaluate the impact from missing data on the EBS remission rates at Weeks 10 and 58, the following missing value imputations will be used:

1) Observed-cases only imputation: Observed cases will be used for analysis without any imputation. Only subjects in the FAS with both baseline and Week 10 (or Week 58) data will be included for analysis.

2) Missing Success: impute subjects in the FAS who did not have sufficient data to decide on EBS remission as achieving the event.

3) Missing Success for placebo group and Missing Failure for filgotinib groups: impute subjects in the FAS who did not have sufficient data to decide on EBS remission as achieving this event for placebo group and as not achieving this event for filgotinib groups.

4) Using multiple imputation method to impute subjects in the FAS who did not have sufficient data to decide on EBS remission status. The EBS remission status for subjects who have missing data at Week 10 for Induction Studies or Week 58 for the Maintenance Study will be imputed in the multiple imputation procedure. A logistic regression model will be utilized to perform the imputation. The baseline values of EBS subscores (endoscopic, rectal bleeding and stool frequency subscores), treatment and stratification factors will be used as the independent variables in the model.

6.2.4.4. Analysis Excluding US/Korea Non-Dual Refractory Males

Given the restriction in the US/Korea of only allowing dual refractory males to receive filgotinib 200 mg, there exists an inherent imbalance in the Cohort B Induction Study between subjects randomized to filgotinib 200 mg and the placebo group with respect to US/Korea non-dual refractory males who will be present in the placebo group, but not in the filgotinib 200 mg group. Recognizing this imbalance and the theoretical potential for non-dual refractory subjects to have better disease prognosis and a higher chance of response, a sensitivity analysis excluding US/Korea non-dual refractory males from the placebo group in the FAS for Cohort B Induction Study will be conducted using the stratified CMH test for the treatment comparison between the filgotinib 200 mg and the placebo group on EBS remission at Week 10. This analysis will be the primary analysis for evaluating the efficacy of the filgotinib 200 mg dose group in the Cohort B Induction Study in support of a Japanese marketing application.

6.2.5. Subgroup Analysis of the Primary Efficacy Endpoint

The following subgroup analyses will be performed for the primary efficacy endpoints for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study):

Subgroups based on study stratification include:

Cohort A Induction Study Stratification Factors:

- Concomitant use of oral, systemically absorbed corticosteroids at Day 1 (yes, no)
- Concomitant use of immunomodulators at Day 1 (yes, no)

Cohort B Induction Study Stratification Factors:

- Exposure to biologic agent listed in Appendix 2 (≤ 1, > 1)
- Concomitant use of oral, systemically absorbed corticosteroids at Day 1(yes, no)
- Concomitant use of immunomodulators at Day 1(yes, no)

Maintenance Study Stratification Factors:

- Participation in Cohort A or Cohort B
- Concomitant use of oral, systemically absorbed corticosteroids at re-baseline (yes, no)
- Concomitant use of immunomodulators at re-baseline(yes, no)

Other subgroups for each individual study include:

- Age group on the date of first dose of the study drug (< 65 years, ≥ 65 years)
- Sex at birth (female, male)
- Race (Asian, Black or African American, White, and Other)
- Geographic region (the United States [US], non-US)
- Induction baseline hs-CRP (≤ 3 mg/L, > 3 mg/L)
- Induction baseline fecal calprotectin (≤ 250 μg/g, > 250 μg/g)
- Duration of UC (years from date diagnosed to first dose date of Induction study drug) (< 1 year, ≥ 1 to < 3 years, ≥ 3 to < 7 years, ≥ 7 years)
- Baseline disease activity (based on screening MCS) (6-8, 9-12)
- Previous exposure to TNFα antagonist (yes, no) (only for Cohort B Induction Study and Maintenance Study)
- Prior TNFα antagonist failure (yes, no) (only for Cohort B Induction Study and Maintenance Study)
- Previous exposure to vedolizumab (yes, no) (only for Cohort B Induction Study and Maintenance Study)
- Prior vedolizumab failure (yes, no) (only for Cohort B Induction Study and Maintenance Study)
- Dual refractory (yes, no) (only for Cohort B Induction Study and Maintenance Study)
- The following 4 subgroups (only for Cohort B Induction Study and Maintenance Study)

  US/Korea males who are dual refractory

  US/Korea males who are not dual refractory

  Non-US/Korea males (US/Korea females and subjects from countries other than US and Korea) who are dual refractory

  Non-US/Korea males who are not dual refractory

For a subject, if the value of the grouping variable cannot be determined, this subject will be excluded from the corresponding subgroup analysis. Non-stratified risk difference between treatment groups will be evaluated for each of the subgroups using Fisher's exact test. A forest plot will graphically present the non-stratified risk difference and 95% CI using normal approximation with a continuity correction on the treatment differences (filgotinib placebo) in EBS remission rates for each of the subgroups.

Additional subgroup analyses for geographic region may be added based on the distribution of study enrollment prior to study unblinding.

6.3. Secondary Efficacy Endpoints

6.3.1. Definition of the Secondary Efficacy Endpoints

The key secondary efficacy endpoints for Induction Studies are:

- The proportion of subjects achieving MCS remission at Week 10
- The proportion of subjects achieving an endoscopic subscore of 0 at Week 10
- The proportion of subjects achieving Geboes histologic remission at Week 10
- The proportion of subjects achieving MCS remission (alternative definition) at Week 10

The key secondary efficacy endpoints for Maintenance Study are:

- The proportion of subjects achieving MCS remission at Week 58
- The proportion of subjects achieving sustained EBS remission, defined as EBS remission at both Weeks 10 and 58
- The proportion of subjects achieving 6-month corticosteroid-free EBS remission at Week 58
- The proportion of subjects achieving an endoscopic subscore of 0 at Week 58
- The proportion of subjects achieving Geboes histologic remission at Week 58
- The proportion of subjects achieving MCS remission (alternative definition) at Week 58

Please refer to Table 6-1 for detailed definition for all secondary efficacy endpoints for each individual study (Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study).

6.3.2. Analysis Methods for Secondary Efficacy Endpoints

The same statistical method that was described under Section 6.2.3 for testing the primary efficacy endpoint will be utilized for testing the key secondary efficacy endpoints. The key secondary efficacy endpoints will be tested in the order listed in Section 3.5. The key secondary efficacy endpoints will be tested only after the primary efficacy endpoint is shown to be statistically significant at the significance level specified in Section 3.5. Each key secondary efficacy endpoint will be tested at the same significance level that is used for the primary efficacy endpoint; testing will be stopped after the first failure to reject the null hypothesis of no difference for statistical inference. If the null hypothesis testing for the primary efficacy endpoint is not rejected, the key secondary efficacy endpoint comparisons will be considered exploratory.

Bar-charts of proportions of subjects achieving each of the key secondary efficacy endpoints by treatment group with corresponding p-values will be provided.

6.4. Exploratory Efficacy Endpoints

As the exploratory efficacy endpoint comparisons are not part of the gatekeeping strategy described under Section 3.5 to control the family-wise type I error, all testing described in this section will be done in an exploratory fashion.

Analysis results of biomarkers except for hs-CRP, fecal calprotectin, and fecal lactoferrin, as exploratory endpoints, will be included into a separate biomarker analysis report outside of the study CSR. Analysis of hs-CRP, fecal calprotectin and fecal lactoferrin is described in Section 9.

6.4.1. Definition of the Exploratory Efficacy Endpoints

A subset of exploratory efficacy endpoints will be described in the CSR; for the Induction Studies these include:

- Change from baseline in HRQoL scores at Week 10
- The proportion of subjects achieving novel histologic outcomes at Week 10
- The proportion of subjects achieving endoscopic response at Week 10
- The proportion of subjects achieving MCS response at Week 10
- Change from baseline in MCS at Week 10
- Change from baseline in pMCS at Week 10
- The proportion of subjects achieving EBS remission (alternative definition) at Week 10

A subset of exploratory efficacy endpoints will be described in the CSR; for the Maintenance Study these include:

- The proportion of subjects achieving sustained MCS remission, defined as MCS remission at both Weeks 10 and 58
- The proportion of subjects achieving 6-month corticosteroid-free MCS remission at Week 58
- Change from baseline in HRQoL scores at Week 58
- The proportion of subjects achieving novel histologic outcomes at Week 58
- The proportion of subjects achieving endoscopic response at Week 58
- The proportion of subjects achieving MCS response at Week 58
- Change from baseline in MCS at Week 58
- Change from baseline in pMCS at Week 58
- The proportion of subjects achieving EBS remission (alternative definition) at Week 58

Please refer to Table 6-1 and Section 6.1 for detailed definitions of novel histologic outcomes, EBS remission (alternative definition), endoscopic response, and MCS response at Weeks 10 and Week 58, respectively, sustained MCS remission at Week 58, and 6-month corticosteroid-free MCS remission at Week 58 for Maintenance Study.

6.4.2. Analysis of Exploratory Efficacy Endpoints

6.4.2.1. Analysis of dichotomous efficacy endpoints

A stratified CMH test will be used to compare the percentage difference of subjects between each filgotinib group and placebo group. The CMH chi-square nominal p-value will be provided. The 2-sided 95% CI based on normal approximation method will be provided for each treatment group. In addition, non-stratified risk difference estimated along with its 95% 2-sided CI with a continuity correction will be provided.

Bar-charts of proportions of subjects achieving the exploratory event type efficacy endpoints will be provided by treatment group with corresponding nominal p-values.

6.4.2.2. Analysis of MCS and pMCS exploratory endpoints

For calculation of MCS and pMCS, please refer to rules described under Section 6.1.1 for MCS and pMCS with subscores.

MCS at screening will serve as induction baseline MCS. Partial MCS at Day 1 will serve as induction baseline pMCS. If pMCS at Day 1 is missing, the screening partial MCS will be used as induction baseline.

MCS at Week 10 will serve as maintenance re-baseline MCS. Partial MCS at Maintenance Day 1 will serve as Maintenance re-baseline pMCS. Re-baseline values of MCS and partial MCS will serve as Maintenance Study baseline for change from baseline calculation.

As 2 of the 3 partial MCS components (stool frequency and rectal bleeding) will be collected through the subject daily diary, partial MCS will be calculated at all visits when PGA is scheduled to be assessed.

Descriptive statistics will be used to summarize the absolute values and change from baseline values in MCS, endoscopic subscore, stool frequency subscore, rectal bleeding subscore, PGA and partial MCS by analysis visit and treatment group for Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study. Shift tables of MCS subscores from baseline to Week 10 for Induction Studies and from Week 10 to Week 58 for Maintenance Study will be provided.

Difference of mean change from baseline at Week 10 for Induction Studies and difference of mean change from re-baseline at Week 58 for Maintenance study between each filgotinib dose group and placebo for partial MCS will be tested in an exploratory fashion using an analysis of covariance (ANCOVA) model adjusting for stratification factors and baseline or rebaseline value, respectively. Last Observation Carried Forward (LOCF) approach will be used to impute the missing values. Estimated means of treatment effects and estimated differences in treatment effects between each filgotinib treatment group and the placebo group will be presented with 95% CIs and nominal p-values. Difference of mean change from baseline (or re-baseline) for partial MCS at Weeks 2, 4, and 6 for Induction Studies and at Weeks 14, 20, 26, 34, 42 and 50 for Maintenance Study will be analyzed in the same fashion. The model based estimated mean change from baseline (or re-baseline) with 95% CIs will be plotted over all analysis visits by treatment group.

6.4.2.3. Analysis of exploratory Health-Related Quality of Life (HRQoL) endpoints

HRQoL questionnaires include Short-Form 36 Health Survey (SF-36), Work Productivity and Activity Impairment questionnaire (WPAI), EuroQoL (health-related quality of life questionnaire) (EQ-5D), and inflammatory bowel disease questionnaire (IBDQ).

HRQoLs will be collected at baseline, Week 10, Week 26, and Week 58.

SF-36

The SF-36 is a health related quality of life instrument consisting of 36 questions belonging to 8 domains in 2 components and covers a 4-week recall period:

- Physical well-being, 4 domains: physical functioning (10 items), role physical (4 items), bodily pain (2 items), and general health perceptions (5 items)
- Mental well-being, 4 domains: vitality (4 items), social functioning (2 items), role emotional (3 items), and mental health (5 items).

The remaining item (health transition) is not part of the above domains but is kept separately. These scales will be rescaled from 0 to 100 (converting the lowest possible score to 0 and the highest possible score to 100), with higher scores indicating a better quality of life. The SF-36 is not disease specific and has been validated in numerous health states.

The change from baseline at Week 10 for Cohort A and B Induction Studies and the change from re-baseline at Weeks 26 and 58 for Maintenance Study in each of the 8 domains and each of the 2 components in SF-36 will be analyzed using an ANCOVA model. The model will include treatment, stratification factors and baseline/re-baseline score as covariates. LOCF approach will be used to impute missing values. Estimated means of treatment effects and estimated differences in treatment effects between each filgotinib treatment group and the placebo group will be presented with 95% CIs and nominal p-values.

Descriptive statistics will be used to summarize the absolute values and change-from baseline (or re-baseline) values for each domain as well as each component by treatment group and analysis visit.

WPAI

The WPAI is a designed to measure the effect of general health and symptom severity on work productivity and regular activities during the past seven days.

The questionnaire consists of 6 questions:

Q1: currently employed,

Q2: work time missed due to UC,

Q3: work time missed due to other reasons,

Q4: hours actually worked,

Q5: degree UC affected productivity while working (0-10 VAS; with 0 indicating no effect and 10 indicating UC completely prevented the subject from working),

Q6: degree UC affected productivity in regular unpaid activities (0-10 VAS; with 0 indicating no effect and 10 indicating UC completely prevented the subject's daily activities).

WPAI outcomes are expressed as impairment percentages, with higher numbers indicating greater impairment and less productivity, that is, worse outcomes, as the following domains:

- The percentage of work time missed (absenteeism) due to UC: 100xQ2/(Q2+Q4)
- The percentage of impairment while working (presenteeism) due to UC: 100xQ5/10
- The percentage of overall work impairment (work productivity loss) due to UC: 100x{Q2/(Q2+Q4)+[(1-Q2/(Q2+Q4))x(Q5/10)]}
- The percentage of activity impairment due to UC: 100xQ6/10.

Summary statistics of the absolute and change from baseline (or re-baseline) scores for each of the four domains will be displayed by treatment and visit.

The change from baseline at Week 10 for Cohort A and B Induction Studies and the change from re-baseline at Weeks 26 and 58 for Maintenance Study will be analyzed using similar ANCOVA models as for SF-36.

EQ-5D

The EQ-5D is a standardized instrument developed by the EuroQol Group as a measure of health-related quality of life that can be used in a wide range of health conditions and treatments.

The EQ-5D consists of 2 components: a descriptive system of the subject's health and a rating of his or her current health state using a 0 to 100 VAS. The descriptive system comprises the following 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension with 5 levels of severity is used for our study: no problems, slight problems, moderate problems, severe problems, and extreme problems. The VAS records the subject's self-rated health on a vertical VAS in which the endpoints are labeled "best imaginable health state" on the top and "worst imaginable health state" on the bottom. Higher EQ VAS indicates better health.

The EQ-5D data will be summarized by treatment group and visit as the following:

- A health profile: summary statistics will be provided, including number of patients and proportions of categorical responses for the 5 dimensions (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression)

- A self-perceived current health score: calculated from patient level EQ VAS responses (continuous variable). The mean, SD, min, median, and max scores will be provided.

The change from baseline at Week 10 for Cohort A and B Induction Studies and the change from re-baseline at Weeks 26 and 58 for Maintenance Study will be analyzed using similar ANCOVA models as for SF-36.

IBDQ

This disease-specific questionnaire comprises 32 questions divided into four health subscales: bowel symptoms (10 questions); systemic symptoms, including sleep disorders and fatigue (5 questions); emotional function such as depression, aggression, and irritation (12 questions); and social function, meaning the ability to participate in social activities and to work (5 questions).

Summary statistics of the absolute and change from baseline (or re-baseline) scores for IBDQ total score and each subscale score will be displayed by treatment and visit.

The change from baseline at Week 10 for Cohort A and B Induction Studies and the change from re-baseline at Weeks 26 and 58 for Maintenance Study will be analyzed using similar ANCOVA models as for SF-36.

6.5. Interim Efficacy Analyses

6.5.1. DMC Futility Analysis

After the initial DMC meeting, the second DMC meeting will include an interim futility analysis for each Induction Study after approximately 175 subjects reach Week 10 or discontinue from study. The futility analysis will be conducted to evaluate endoscopic efficacy. Overall safety will be evaluated at this timepoint as well. The proportion of subjects who achieve an endoscopic subscore of 0 or 1 for each treatment group will be evaluated. The DMC may recommend termination of a filgotinib dose group in the Induction Study that is being evaluated if the observed proportion of subjects who achieve endoscopic response in the filgotinib group is lower than that in the placebo group. If both filgotinib dose groups meet the futility criteria, the DMC may recommend stopping the study (ie, Cohort A Induction Study [biologic naïve], Cohort B Induction Study [treatment experienced], or both).

Should the DMC recommend that the study be modified, a Gilead employee (the sponsor representative, as defined in the DMC charter) may be provided with unblinded study data in order to make a decision about study conduct. Unblinding of specific Gilead personnel will be documented per the appropriate standard operation procedures (SOP).

6.5.2. DMC Cohorts A and B End-of-Induction Analysis

After all subjects from Cohorts A and B have completed the Week 10 visit or have terminated prior to Week 10 and corresponding data entry is complete, an End of Induction safety and efficacy analysis will be performed, including cumulative safety analysis and analysis on primary endpoint. Gilead Sciences, Inc.'s (Gilead) blinded statistical programmers will provide unblinded external statistician, who is independent of Gilead and not a member of Gilead's Study Management Team, with the datasets and programs necessary to complete the analysis. The unblinded external statistician will apply the unblinded treatment codes to the datasets and generate the unblinded analysis results.

Both cohorts will be examined independently by DMC:

- Taking into account data in Cohort A and Cohort B, if both dosing groups (200 mg and 100 mg) in both cohorts (independently examined) fail to reach statistical significance compared to placebo on EBS remission, the DMC may recommend overall study discontinuation.

- If the condition above is not met, the DMC may recommend that the study continue without modification.

6.6. Change from Protocol-Specified Efficacy Analyses

Protocol amendment 5 (02 April 2019) Section 8.11, described that a review of the results of all primary and key secondary endpoints of Cohort A and B Induction Studies would be performed by a Gilead executive team in parallel with the DMC's review of the end of induction analysis.

However, this review by a Gilead executive team was not performed due to concerns raised by the Voluntary Harmonization Procedure (VHP) during protocol review. No Gilead executive team member or GS-US-418-3898 study team member had access to any unblinded study results prior to study completion and database finalization/lock.

EBS remission (alternative definition), endoscopic response, MCS response, changes from baseline in MCS and partial MCS at Weeks 10 and 58 are included as exploratory endpoints for Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study. The analyses on these endpoints are also included as exploratory analyses.

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

Clinical and laboratory AEs will be coded using the current version of MedDRA. System organ class (SOC), high-level group term (HLGT), high-level term (HLT), preferred term (PT), and lower-level term (LLT) will be provided in the AE dataset.

7.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, 4, or 5 according to toxicity criteria specified in the protocol. The severity grade of events for which the investigator did not record severity will be categorized as "missing" for tabular summaries and data listings. The missing category will be listed last in summary presentation.

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE case report form (CRF) to the question of "Related to Study Treatment." Relatedness will always reflect the investigator assessment of causality rather than the sponsor's. Events for which the investigator did not record relationship to study drug will be considered related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if AEs met the definitions of SAE specified in the study protocol. Serious adverse events captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Pharmacovigilance & Epidemiology Department before data finalization.

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent Adverse Events

Treatment-emergent adverse events (TEAEs) are defined as 1 or both of the following:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- Any AEs with an onset date on or after the study drug start date for each Induction Study (Day 1) and:

  No later than 30 days after the last dosing date of Induction Study, for subjects who did not take any study treatment from Maintenance Study, or

  Before the first dosing date of Maintenance Study, for subjects who took any study treatment in Maintenance Study

- Any AEs leading to premature discontinuation of Induction study drug

Maintenance Study

- Any AEs with an onset date on or after the study drug start date for the Maintenance Study and up to 30 days after last dosing date of Maintenance Study
- Any AEs leading to premature discontinuation of Maintenance study drug

Overall (Induction and Maintenance)

- Any AEs with an onset date on or after the study drug start date for the entire study and up to 30 days after last dosing date of either Induction Study or Maintenance Study, whichever is later
- Any AEs leading to premature discontinuation of either Induction or Maintenance study drug

7.1.5.2. Incomplete Dates

If the onset date of the AE is incomplete and the AE stop date is not prior to the first dosing date of study drug, then the month and year (or year alone if month is not recorded) of onset date determine whether an AE is treatment emergent. For first dosing date and last dosing date defined for the Induction Studies (Cohort A Induction Study and Cohort B Induction Study), Maintenance Study and Overall, please refer to Section 3.8.1. The event is considered treatment emergent if both of the following 2 criteria are met:

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

- The AE onset date is the same as or after the month and year (or year) of the first dosing date of study drug, and
- The AE onset date is the same as or before the month and year (or year) of the date corresponding to

  30 days after the last dosing date of Induction Study, for subjects who did not take any treatment from Maintenance, or

  the day before the first dosing date of Maintenance Study, for subjects who took any study treatment in Maintenance Study

Maintenance Study

- The AE onset date is the same as or after the month and year (or year) of the first dosing date of Maintenance Study and
- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after last dosing date of Maintenance Study

Overall (Induction and Maintenance)

- The AE onset date is the same as or after the month and year (or year) of the first dosing date of study drug, and
- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after last dosing date of either Induction Study or Maintenance Study, whichever is later

An AE with completely missing onset and stop dates, or with the onset date missing and a stop date later than the first dosing date of study drug, will be considered to be treatment emergent. In addition, an AE with the onset date missing and an incomplete stop date with the same or later month and year (or year alone if month is not recorded) as the first dosing date of study drug will be considered treatment emergent.

7.1.6. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized based on the Safety Analysis Set for Cohort A Induction Study, Cohort B Induction Study, Cohort A and B Induction Studies Combined, Maintenance Study, and Overall separately.

No formal statistical testing is planned.

7.1.6.1. Summaries of AE incidence

The number and percentage of subjects who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and treatment group. For other AEs described below, summaries will be provided by SOC, PT, and treatment group:

- TEAEs of Grade 3 or higher (by maximum severity)
- TEAEs of Grade 2 or higher (by maximum severity)
- All TE treatment-related AEs
- TE treatment-related AEs of Grade 3 or higher (by maximum severity)
- TE treatment-related AEs of Grade 2 or higher (by maximum severity)
- All TE SAEs
- All TE treatment-related SAEs
- All TEAEs leading to premature discontinuation of study drug
- All TEAEs leading to premature discontinuation of study
- All TE SAEs leading to death (ie, outcome of death)
- All TEAEs leading to temporary dose interruption of study drug

A brief, high-level summary of AEs described above will be provided by treatment group presenting the number and percentage of subjects who experienced the above AEs. All deaths observed in the study will be also included in this summary.

Multiple events will be counted only once per subject in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC and HLT within each SOC (if applicable), and then by PT in descending order of total frequency within each SOC. For summaries by severity grade, the most severe grade will be used for those AEs that occurred more than once in an individual subject during the study.

In addition, the following tables will be generated and summarized by PT only, in descending order of total frequency:

- TEAEs
- TEAEs of Grade 3 or higher
- TEAEs of Grade 2 or higher
- TE SAEs
- TE treatment-related AEs
- TE treatment-related SAEs
- TEAEs leading to premature discontinuation of study drug

Data listings will be provided for the following:

- All AEs, indicating whether the event is treatment emergent
- All AEs of Grade 3 or higher
- All AEs of Grade 2 or higher
- SAEs
- Deaths
- All AEs leading to death (ie, outcome of death)
- AEs leading to premature discontinuation of study drug
- AEs leading to premature discontinuation of study
- AEs leading to dose temporary interruption of study drug

7.1.7. Adverse Events of Interest

Adverse events of interest (AEI) include infections, gastrointestinal perforations, herpes zoster, malignancies (excluding non-melanoma skin cancers), non-melanoma skin cancers, and thromboembolic events. Summaries of the following treatment-emergent AEIs will be produced to enhance the analysis of safety data.

- Events of infections, presented in the following subcategories:

  AEs of infections, utilizing all AEs in the MedDRA Infections and Infestations SOC

  AEs of serious infections, using all AEs in the MedDRA Infections and Infestations SOC that are classified as SAEs

  AEs of herpes zoster, utilizing a MedDRA search term (MST) list developed by Gilead

  AEs of opportunistic infections (OIs), using a MST list developed by Gilead

- AEs of malignancies, excluding non-melanoma skin cancers, utilizing a MST list developed by Gilead

- AEs of non-melanoma skin cancers, utilizing a MST list developed by Gilead

- AEs of gastrointestinal perforation, utilizing a MST list developed by Gilead

- AEs of thromboembolism, presented in the following subcategories:

  AEs of venous thrombosis, utilizing a MST list developed by Gilead

  AEs of pulmonary embolism, utilizing a MST list developed by Gilead

  AEs of arterial thrombosis, utilizing the embolic and thrombotic events, arterial Standardised MedDRA Queries (SMQ)

  AEs of cerebrovascular events, utilizing the ischaemic central nervous system vascular conditions SMQ

The number and percentage of subjects with a reported event will be summarized for each treatment group by PT for each AEI category. Data listings for AEIs will also be provided.

7.1.8. Additional Analysis of Adverse Events

TEAE and SAE for the following subgroups of subjects will be summarized by SOC and PT for Cohort A Induction Study, Cohort B Induction Study, Cohort A and B Induction Studies Combined, Maintenance Study, and Overall separately:

- Age group (< 65 years, ≥ 65 years; on the date of first dose of the study drug)
- Sex at birth (female, male)
- Race (Asian, Black or African American, White, Other)
- Geographic region (US, non-US)
- Prior TNFα antagonist failure (yes, no) (not applicable for Cohort A Induction Study)
- Prior vedolizumab failure (yes, no) (not applicable for Cohort A Induction Study)
- Prior dual refractory (TNFα antagonist and vedolizumab) (yes, no) (not applicable for Cohort A Induction Study)
- Concomitant use of systemic corticosteroid and immunomodulator at Day 1 (or re-baseline for Maintenance Study), including

  Systemic corticosteroid only

  Immunomodulator only

  Both systemic corticosteroid and immunomodulator

  Neither systemic corticosteroid nor immunomodulator

7.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set for Cohort A Induction Study, Cohort B Induction Study, Cohort A and B Induction Studies Combined, Maintenance Study, and Overall (treatment-emergent laboratory abnormalities tables only for Overall) separately and will include data collected as specified below:

<u>Induction Studies (Cohort A Induction Study and Cohort B Induction Study)</u>

- No later than 30 days after the last dosing date of Induction Study, for subjects who did not take any study treatment from Maintenance Study, or
- On or before the first dosing date of Maintenance Study, for subjects who took any study treatment in Maintenance Study

Maintenance Study

- No later than 30 days after the last dosing date of Maintenance Study

Overall (Induction and Maintenance)

- No later than 30 days after the last dosing date of Induction or Maintenance Study, whichever is later

The analysis will be based on values reported in conventional units. When values are below the LOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics as specified in Section 3.7. Test results for hemolyzed samples will not be included in the analysis, but they will be listed in by-subject laboratory listings.

A by-subject listing for laboratory test results will be provided by subject ID number and time point in chronological order for hematology, serum chemistry, urinalysis, lipid profile, and serum immunoglobulin separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher on the CTCAE severity grade will be flagged in the data listings, as appropriate.

No formal statistical testing is planned.

7.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics will be provided by treatment group for each laboratory test specified in the study protocol within hematology and chemistry panels including creatinine clearance, and also laboratory tests from lipids panel including total cholesterol, LDL, HDL, triglycerides, non-HDL cholesterol (total cholesterol minus HDL cholesterol), LDL/HDL ratio, and IgA, IgM, IgG, and total Ig, as follows for each Induction Study and Cohort A and B Induction Studies Combined:

- Baseline values
- Values at each postbaseline time point
- Change from baseline at each postbaseline time point

In addition, descriptive statistics for the laboratory tests described from above will be provided by treatment group for the Maintenance Study:

- Re-baseline values
- Values at each postbaseline time point after re-baseline
- Change from re-baseline at each postbaseline time point after re-baseline

A baseline laboratory value will be defined as the last measurement obtained on or prior to the date of first dose of study drug. Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value. A re-baseline laboratory value will be defined as the last measurement obtained on or prior to the date of first dose of study drug for the Maintenance study, and change from re-baseline will be defined as the visit value minus the re-baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

Median (Q1, Q3) of the observed change from baseline (or re-baseline) values for aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin, alkaline phosphatase, serum creatine, creatinine clearance, creatine phosphokinase, white blood cell count, absolute neutrophils, absolute lymphocytes, hemoglobin, platelets, total cholesterol, LDL, HDL, total Ig, IgG, IgA and IgM, will be plotted using a line plot by treatment group and visit.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3.

7.2.2. Graded Laboratory Values

The CTCAE Version 4.03 will be used for assigning toxicity grades (0 to 4) to laboratory results for analysis. Grade 0 includes all values that do not meet the criteria for an abnormality of at least Grade 1. For laboratory tests with criteria for both increased and decreased levels, analyses for each direction (ie, increased, decreased) will be presented separately.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, and

- No later than 30 days after the last dosing date of Induction Study, for subjects who did not take any study treatment from Maintenance Study, or
- On or before the first dosing date of Maintenance Study, for subjects who took any study treatment in Maintenance Study

Maintenance Study

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from re-baseline at any time point after first dose of maintenance, up to 30 days after last dosing date of Maintenance Study.

Overall (Induction and Maintenance)

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline (Day 1) at any postbaseline time point, up to 30 days after permanent discontinuation of study drug of Induction or Maintenance study, whichever is later.

If the relevant baseline laboratory value is missing, any abnormality of at least Grade 1 observed within the time frame specified above will be considered treatment emergent.

7.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

Induction Studies (Cohort A Induction Study and Cohort B Induction Study)

Treatment-emergent marked laboratory abnormalities are defined as values that increase at least 3 toxicity grades from baseline at any postbaseline time point, and

- No later than 30 days after the last dosing date of Induction Study, for subjects who did not take any study treatment in the Maintenance Study, or
- On or before the first dosing date of Maintenance Study, for subjects who took any study treatment in the Maintenance Study

Maintenance Study

Treatment-emergent marked laboratory abnormalities are defined as values that increase at least 3 toxicity grades from re-baseline at any postbaseline time point after the first dose date of Maintenance Study, up to 30 days after permanent discontinuation of study drug.

Overall (Induction and Maintenance)

Treatment-emergent laboratory abnormalities are defined as values that increase at least 3 toxicity grades from baseline at any postbaseline time point, up to 30 days after permanent discontinuation of study drug of Induction or Maintenance study, whichever is later.

If the relevant baseline laboratory value is missing, any Grade 3 or 4 values observed within the timeframe specified above will be considered treatment-emergent marked abnormalities.

7.2.2.3. Summaries of Laboratory Abnormalities

Laboratory data that are categorical will be summarized using the number and percentage of subjects in the study with the given response at baseline and each scheduled postbaseline time point for Cohort A Induction Study, Cohort B Induction Study, and Cohorts A and B Induction Studies combined, and at re-baseline and each schedule time point afterwards for Maintenance Study.

The following summaries (number and percentage of subjects) for treatment-emergent laboratory abnormalities will be provided by lab test and treatment group for Cohort A Induction Study, Cohort B Induction Study, Cohorts A and B Induction Studies combined, Maintenance Study, and Overall; subjects will be categorized according to the most severe postbaseline abnormality grade for a given lab test:

- Graded laboratory abnormalities
- Grade 3 or higher laboratory abnormalities
- Marked laboratory abnormalities

For all summaries of laboratory abnormalities, the denominator is the number of subjects with nonmissing postbaseline values.

A by-subject listing of treatment-emergent Grade 3 or 4 laboratory abnormalities will be provided by subject ID number and time point in chronological order. This listing will include all test results that were collected throughout the study for the lab test of interest, with all applicable severity grades displayed.

7.2.3. Laboratory Evaluations of Interest

7.2.3.1. Liver-related Laboratory Evaluations

Liver-related abnormalities after initial study drug dosing will be examined and summarized using the number and percentage of subjects who were reported to have the following laboratory test values for postbaseline measurements for Cohort A Induction Study, Cohort B Induction Study, Cohorts A and B Induction Studies combined, and Maintenance Study:

- AST: (a) > 3 times of the upper limit of normal range (ULN); (b) > 5 x ULN; (c) > 10 x ULN; (d) > 20 x ULN
- ALT: (a) > 3 x ULN; (b) > 5 x ULN; (c) > 10 x ULN; (d) > 20 x ULN
- AST or ALT > 3 x ULN and total bilirubin > 2 x ULN

For individual laboratory tests, subjects will be counted once based on the most severe postbaseline values. For both the composite endpoint of AST or ALT and total bilirubin, subjects will be counted once when the criteria are met at the same postbaseline visit date. The denominator is the number of subjects in the Safety Analysis Set who have nonmissing postbaseline values of all relevant tests at the same postbaseline visit date. A listing of subjects who met at least 1 of the above criteria will be provided.

7.2.3.2. Complete Blood Count-Related Laboratory Evaluations

Complete blood count (CBC)-related abnormalities such as anemia, leucopenia, neutropenia, lymphopenia, and thrombocytopenia after initial study drug dosing will be examined and summarized using the number and percentage of subjects who were reported to have the following laboratory test values for postbaseline measurements for Cohort A Induction Study, Cohort B Induction Study, Cohorts A and B Induction Studies combined, and Maintenance Study:

- Hemoglobin: (a) postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or 4 at worst postbaseline; (c) baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

- White blood cell (WBC) count: (a) postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or 4 at worst postbaseline; (c) baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

- Absolute neutrophil count: (a) postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or 4 at worst postbaseline; (c) baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

- Lymphocyte count: (a) postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or 4 at worst postbaseline; (c) baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

- Platelet count: (a) postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or 4 at worst postbaseline; (c) baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

For Maintenance Study, re-baseline will be used in the summaries described above.

7.2.4. Shifts Relative to the Baseline Value

Shift tables will be presented by showing change in CTCAE grade for hemoglobin, platelet count, white blood cell (WBC) count, absolute lymphocyte count, absolute neutrophils, fasting triglycerides, and fasting total cholesterol.

Shift tables will be presented by showing change in lab normal range (low, normal, and high) for hematocrit, absolute monocyte count, absolute eosinophil count, and absolute basophil count. In addition, shift tables for fasting LDL and HDL will be presented using the following National Cholesterol Education Program (NCEP) ATP III categories {National Cholesterol Education Program (NCEP) 2001}:

- For LDL (mg/dL): <100, 100-129, 130-159, 160-189, and ≥ 190

- For HDL (mg/dL): < 40, ≥ 40 to < 60, and ≥ 60

In all shift tables, the number and proportion of subjects for each category of each laboratory result will be summarized by its baseline/rebaseline category for Cohort A Induction Study, Cohort B Induction Study, Cohorts A and B Induction Studies combined, and Maintenance Study.

For the Induction Studies, shift tables will be presented for Week 10. For the Maintenance Study, shift tables will be presented for Weeks 26 and 58.

7.3. Body Weight and Vital Signs

Descriptive statistics will be provided by treatment group for body weight, BMI and vital signs (resting blood pressure [systolic blood pressure and diastolic blood pressure], respiratory rate, pulse and temperature) as follows for each Induction Study and combined:

- Baseline value
- Values at each postbaseline time point
- Change from baseline at each postbaseline time point

In addition, descriptive statistics will be provided by treatment group for Maintenance Study

- Re-baseline values
- Values at each postbaseline time point after re-baseline
- Change from re-baseline at each postbaseline time point after re-baseline

A baseline value will be defined as the last available value collected on or prior to the date of first dose of study drug. Change from baseline to a postbaseline visit will be defined as the postbaseline value minus the baseline value. A re-baseline value will be defined as the last measurement obtained on or prior to the date/time of first dose of study drug for the Maintenance Study, and change from re-baseline will be defined as the visit value minus the re-baseline value. Body weight and vital signs measured at unscheduled visits will be included for the baseline/re-baseline value selection.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.8.3. No formal statistical testing is planned.

A by-subject listing of vital signs will be provided by subject ID number and time point in chronological order. Body weight , height, and BMI will be included in the vital signs listing, if space permits. If not, they will be provided separately.

7.4. Prior and Concomitant Medications

Medications reported at screening and during the study will be coded using the current version of the World Health Organization (WHO) Drug dictionary.

7.4.1. Prior Medications

Prior medications are defined as any medication taken before a subject took the first study drug at baseline/re-baseline.

General prior and UC-specific prior medications will be summarized separately by preferred name using the number and percentage of subjects for each treatment group and overall. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be ordered by preferred term in descending overall frequency. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medication with a start date prior to the first dosing date of study drug will be included in the prior medication summary regardless of when the stop date is. If a partial start date is entered the medication will be considered prior unless the month and year (if day is missing) or year (if day and month are missing) of the start date are after the first dosing date. Medications with a completely missing start date will be included in the prior medication summary, unless otherwise specified.

Summaries will be based on the Safety Analysis Set for Cohort A Induction Study, Cohort B Induction Study, Cohort A and B Induction Studies Combined, and Maintenance Study. No formal statistical testing is planned.

7.4.2. Concomitant Medications

General concomitant and UC-specific concomitant medications are defined as medications taken while a subject took study drug. Use of concomitant medications will be summarized by preferred name using the number and percentage of subjects for each treatment group. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be ordered by preferred term in descending overall frequency. For drugs with the same frequency, sorting will be done alphabetically.

First dosing date and last dosing date for Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study have been defined in Section 3.8. For the purposes of analysis, any medication with a start date prior to or on the first dosing date of study drug and continued to take after the first dosing date, or started after the first dosing date but prior to or on the last dosing date of study drug will be considered concomitant medications. Medications started and stopped on the same day as the first dosing date or the last dosing date of study drug will also be considered concomitant. Medications with a stop date prior to the date of first dosing date of study drug or a start date after the last dosing date of study drug will be excluded from the concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the date of first study drug administration will be excluded from the concomitant medication summary. If a partial start date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing start and stop dates will be included in the concomitant medication summary, unless otherwise specified. Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

All prior and concomitant medications (other than per-protocol study drugs) will be provided in a by-subject listing sorted by subject ID number and administration date in chronological order.

7.5. Electrocardiogram Results

A shift table of the investigators' assessment of ECG results at each visit compared with baseline (or re-baseline) values will be presented by treatment group for each Induction Study, Cohort A and B Induction Studies Combined, and Maintenance Study using the following categories: normal; abnormal, not clinically significant; abnormal, clinically significant. The number and percentage of subjects in each cross-classification group of the shift table will be presented. Subjects with a missing value at baseline (or re-baseline) or postbaseline will not be included in the denominator for percentage calculation.

No formal statistical testing is planned.

A by-subject listing for ECG assessment results will be provided by subject ID number and time point in chronological order.

7.6. Other Safety Measures

A data listing will be provided for subjects who become pregnant during the study.

7.7. Changes From Protocol-Specified Safety Analyses

There are no deviations from the protocol-specified safety analyses.

8. PHARMACOKINETIC (PK) ANALYSES

8.1. PK Sample Collection

Sparse plasma PK samples will be collected prior to dose at Weeks 10 and 58, post dose at Week 4 (at least 30 minutes and up to 3 hours after dosing), and at any time at Week 26.

An optional PK Substudy will be performed in a subset of subjects. In the PK substudy, the daily dose of study drug should be administered in the clinic (at one visit between Week 2 and Week 10, inclusive), and additional PK samples should be collected predose and at 0.5, 1, 2, 3, 4, and 6 hours post dose to determine the steady-state PK of filgotinib and its metabolite GS-829845.

8.2. PK Analyses Related to Intensive PK Sampling

Steady-state PK will be determined in subjects in the PK Substudy analysis set. Concentrations of filgotinib and its metabolite GS-829845 in plasma will be determined using validated bioanalytical assays.

8.2.1. Estimation of PK Parameters

PK parameters will be estimated using Phoenix WinNonlin® software using standard noncompartmental methods. The linear/log trapezoidal rule will be used in conjunction with the appropriate noncompartmental model, with input values for dose level, dosing time, plasma concentration, and corresponding real-time values, based on drug dosing times whenever possible.

All predose sample times before time-zero will be converted to 0. Predose samples may also serve as the 24-hour post dose sample if appropriate.

For area under the curve (AUC), samples BLQ of the bioanalytical assays occurring prior to the achievement of the first quantifiable concentration will be assigned a concentration value of 0 to prevent overestimation of the initial AUC. Samples that are BLQ at all other time points will be treated as missing data in WinNonlin. The nominal time point for a key event or dosing interval (τ) may be used to permit direct calculation of AUC over specific time intervals. The appropriateness of this approach will be assessed by the PK scientist on a profile-by-profile basis.

Pharmacokinetic parameters such as AUC_{inf}, λ_z and $t_{1/2}$ are dependent on an accurate estimation of the terminal elimination phase of drug. The appropriateness of calculating these parameters will be evaluated upon inspection of PK data on a profile-by-profile basis by the PK scientist.

8.2.2. PK Substudy

PK parameters will be generated for all subjects in the PK substudy analysis set. The analyte presented in Table 8-1 will be evaluated if data are available

Table 8-1. Study Treatments and Associated Analyte

Treatment Group	Treatment	Analyte
Treatment 1	Filgotinib (200 mg) tablet	Filgotinib, GS-829845
Treatment 2	Filgotinib (100 mg) tablet	Filgotinib, GS-829845

The analytes and parameters presented in Table 8-2 will be used to evaluate the PK objectives of the study. The primary PK parameters are [AUC_{tau}, and C_{max}] of filgotinib and GS-829845. The PK parameters to be estimated in this study are listed and defined in the Appendix 7.

Table 8-2. PK Parameters for Each Analyte

Analyte	Parameters
Filgotinib	AUC_{last}, AUC_{tau}, C_{max}, T_{max}, C_{tau}, and CL_{ss}/F, if appropriate
GS-829845	AUC_{last}, AUC_{tau}, C_{max}, T_{max}, and C_{tau}, if appropriate

Individual subject concentration data and individual subject PK parameters for filgotinib and GS-829845 will be listed and summarized using descriptive statistics by treatment. Summary statistics (subject number [n], mean, SD, coefficient of variation [%CV], median, min, max, Q1, and Q3) will be presented for both individual subject concentration data by time point and individual subject PK parameters by treatment. Moreover, the geometric mean, 95% CI, and the mean and SD of the natural log-transformed values will be presented for individual subject PK parameter data.

Individual concentration data listings and summaries will include all subjects with concentration data. The sample size for each time point will be based on the number of subjects with nonmissing concentration data at that time point. The number of subjects with concentration BLQ will be presented for each time point. For summary statistics, BLQ values will be treated as 0 at predose and one-half of the lower limit of quantitation (LLOQ) for postdose time points.

Individual PK parameter data listings and summaries will include all subjects for whom PK parameter(s) can be derived. The sample size for each PK parameter will be based on the number of subjects with nonmissing data for that PK parameter.

The following tables will be provided for each analyte by treatment:

- Individual subject concentration data and summary statistics
- Individual subject plasma PK parameters and summary statistics

For the PK Substudy, the following figures may be provided for each analyte by treatment:

- Mean (± SD) concentration data versus time (on linear and semilogarithmic scales)
- Median (Q1, Q3) concentration data versus time (on linear and semilogarithmic scales)

Individual, mean, and median postdose concentration values that are ≤ LLOQ will not be displayed in the figures and remaining points connected.

The following listings will be provided:

- PK sampling details by subject including actual dosing time and actual draw time, calculated time postdose of sample collection, differences in scheduled and actual draw times, sample age, and sample concentration

8.3. Statistical Analysis Methods

For both Pharmacokinetic Analysis Set (sparse PK) and Pharmacokinetic Substudy Analysis Set, individual subject concentration data for filgotinib and its metabolite GS-829845 will be listed and summarized using descriptive statistics (eg, sample size, arithmetic mean, geometric mean, coefficient of variation (%) standard deviation, median, minimum, and maximum) by treatment. For PK samples collected prior to dose, summary statistics will be generated at nominal visits and only the records with PK drawn ≥ 22 hours and ≤ 26 hours after the prior dose of study drug will be included. If the date/time of prior dose of study drug is missing, then the corresponding concentration records will be excluded from the summary. The number of subjects with concentration BLQ will be presented for each time point. For summary statistics, BLQ values will be treated as 0 at predose and one-half of the lower limit of quantitation (LLOQ) for postdose time points.

Plasma concentrations over time may be plotted in semi-logarithmic and linear formats as mean ± standard deviation, and median (Q1, Q3) for PK substudy.

For PK substudy, individual PK parameters (C_{max}, T_{max}, C_{tau} AUC_{last}, $AUC_{tau,}$ and CL_{ss}/F, as applicable) for both filgotinib and GS-829845 will be listed and summarized using descriptive statistics. Molar ratio of GS-829845 to filgotinib exposures (AUC_{last}, AUC_{tau}, C_{max}, and C_{tau}, as applicable) will also be calculated and summarized. It can be expreseed as GS-829845 exposure/filgotinib exposure × (425.51/357.43). Moreover, the geometric mean, 95% CI, and the mean and SD of the natural log-transformed values will be presented for individual subject PK parameter data.

Additional PK analysis may be performed and exposure-response relationships may be explored.

9. BIOMARKER ANALYSIS

The biological marker analysis will be conducted on the Biomarker Analysis Set, defined in Section 3.1.7.

Serum hs-CRP, fecal calprotectin, and fecal lactoferrin will be listed and summarized for Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study separately by treatment using descriptive statistics including baseline (or re-baseline) values, values at each postbaseline time point, change from baseline (or re-baseline) and percent change from baseline (or re-baseline) at each postbaseline time point. The definitions of baseline (or re-baseline) for each study are in Section 3.8.1.

Summaries of fecal calprotectin and serum hs-CRP for subjects who have baseline (or re-baseline) values satisfying the following criteria will also be provided for Cohort A Induction Study, Cohort B Induction study, and Maintenance Study:

- Fecal calprotectin > 250 μg/g
- Fecal calprotectin ≤ 250 μg/g
- hs-CRP > 3 mg/L
- hs-CRP ≤ 3 mg/L

Mean (Q1, Q3) of the observed change from baseline (or re-baseline) values for fecal calprotectin, fecal lactoferrin, and serum hs-CRP will be plotted by treatment and visit for Cohort A Induction Study, Cohort B Induction Study, and Maintenance Study separately as well.

10. REFERENCES

Bretz F, Maurer W, Brannath W, Posch M. A Graphical Approach to Sequentially Rejective Multiple Test Procedures. Stat Med 2009;28:586-604.

Geboes K, Riddell R, Ost A, Jensfelt B, Persson T, Lofberg R. A reproducible grading scale for histological assessment of inflammation in ulcerative colitis. Gut 2000;47 (3):404-9.

National Cholesterol Education Program (NCEP). Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Executive Summary. National Institute of Health May, 2001.

11. SOFTWARE

SAS® Software Version 9.4 (SAS Institute Inc., Cary, NC, USA.) is to be used for all programming of tables, listings, and figures.

nQuery Advisor® Version 6.0 (Statistical Solutions, Cork, Ireland.) was used for sample size and power calculation.

12. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
Original Final Draft (07 September 2016)			
27 April 2018	1	Added key secondary objectives of establishing MCS remission (alternative definition) at Week 10 for Cohort A and B Induction Studies and at Week 58 for Maintenance Study	To align with Protocol Amendment 4
	3	Modified definition for Per-Protocol Analysis Sets	To better conduct Per-Protocol analysis
	3, 7	Added Cohorts A and B pooled safety summary	To support safety evaluation of study drug across Induction Studies
	3, 7	Clarified overall safety summaries only include adverse events and laboratory abnormalities including data collected across Cohort A Induction Study, Cohort B Induction Study and Maintenance Study	To support safety evaluation of study drug across Induction and Maintenance Studies
	5	Modified the list of variables to be summarized in baseline characteristic tables	To provide better characterization of the investigated patient populations
	7	Clarified AEs of interest to be summarized	To better address safety concerns
	7	Added summary of complete blood count-related laboratory evaluations	To better address safety concerns
	7	Removed shift relative to baseline tables for laboratory assessment including hematology and chemistry	To simplify the analysis
	Appendix 1	Update Schedule of Assessments	To align with Protocol Amendment 4
	Global	Administrative and editorial changes have been made throughout the SAP, where appropriate	Improve clarity and consistency

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
28 November 2018	3.5 and 6.1	Changed the testing strategy of efficacy endpoints of two filgotinib dose groups from separate to sequential testing with the proposed testing order	In response to comments from FDA
	4.2.2	Further specified details of adherence rate calculation	Improve clarity
	5.2	Added variables of baseline characteristics	To provide better characterization of subjects
	6.2.5	Added subgroup analysis regarding biologic treatment history for Maintenance Study	To better characterize the treatment efficacy profile of Maintenance Study
	7.1.8	Added subgroup summary of AEs and SAEs	In response to comments from FDA
	7.2.4	Added summary of shift change in selective laboratory tests	In response to comments from FDA
	9	Planned summary of additional PK parameters	due to data availability
08 April 2019	1.3	Power calculation revised according to the current alpha spending strategy	In response to comments from FDA
	3	Changed the testing strategy of efficacy endpoint of two filgotinib dose groups to sequentially rejective multiple test procedures	In response to comments from FDA
	Global	Administrative and editorial changes have been made throughout the SAP, where appropriate	Improve clarity and consistency

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
28 April 2020	3.2	For Maintenance Study summary, the subject grouping would not pool all placebo subjects into one group.	Placebo groups not pooled for the Maintenance study in response to comments from FDA on the integrated safety analysis SAP and also in recognition of different patient populations of filgotinib 200mg induced responders, filgotinib 100 mg induced responders, and placebo induced responders
	4.1, 4.2, 5.1 and 5.2	Added enrollment and disposition summary for Cohorts A and B combined.	To report overall patient population and disposition for filgotinib UC program
	4.1	Added "Per-Protocol Analysis Set", "PK Analysis Set" and "Biomarker Analysis Set" into Maintenance Study disposition table.	To improve the completeness of the Maintenance Study disposition table.
	6.1 and 6.4	Added exploratory endpoints of EBS remission (alternative definition) at Week 10 for Induction Studies A and B and at Week 58 for Maintenance Study	In response to comments from FDA
	8.2.2 and 8.3	Removed PK parameters $C_{last,}$ T_{last}, $t_{1/2}$, λ_z and Vss/F; removed listing of individual data on determination of the plasma half-life and corresponding correlation coefficient	To remove parameters that are less appropriate/informative to characterize the steady state PK
	Appendix	Removed Appendix 7 – Appendix 12 from previous SAP	Appendix 7 – Appendix 12 in the previous SAP were MSTs for AEI. As internal process has evolved and the AEI flags are included into the raw datasets, those appendixes have been removed from the SAP.
	Global	Administrative and editorial changes have been made throughout the SAP, where appropriate	Improve clarity and consistency

13. APPENDICES

Appendix 1. Schedule of Assessments

Period	Screen	Treatment																			Follow-up	
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**PTx[a]**	**ET**
Week		**0**	**2**	**4**	**6**	**10**	**11**	**14**	**17**	**20**	**23**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**		
Study Day	**-30 to -1**	**1**	**15**	**29**	**43**	**71**	**78**	**99**	**120**	**141**	**162**	**183**	**211**	**239**	**267**	**295**	**323**	**351**	**379**	**407**		
Visit Window			**±3**	**±3**	**±3**	**±2**	**±3**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	
Written Informed Consent	X																					
Medical History & Demographics	X																					
Ulcerative Colitis Treatment History and disease extent	X																					
12-lead ECG	X					X						X								X		X[b]
Review of Inclusion/ Exclusion Criteria	X	X																				
Complete Physical Exam[c]	X																					
Symptom directed physical exam[c] (as needed)		X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
Vital Signs	X	X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
Height	X																					
Weight	X	X	X	X	X	X		X		X		X		X		X		X		X	X	X
Adverse Events	X	X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
Concomitant Medications	X	X	X	X	X	X	X	X		X		X		X		X		X		X	X	X

Period	**Screen**	**Treatment**																			**Follow-up**	
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**PTx[a]**	**ET**
Week		**0**	**2**	**4**	**6**	**10**	**11**	**14**	**17**	**20**	**23**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**		
Study Day	**-30 to -1**	**1**	**15**	**29**	**43**	**71**	**78**	**99**	**120**	**141**	**162**	**183**	**211**	**239**	**267**	**295**	**323**	**351**	**379**	**407**		
Visit Window			**±3**	**±3**	**±3**	**±2**	**±3**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	
Randomization		X					X															
Study Drug Dispensing		X		X			X	X		X		X		X		X		X				
Colonoscopy/Flexible Sigmoidoscopy with biopsy	X[d]					X														X		
Investigator Endoscopic Value (0,1,2,or 3)[e]	X					X														X		
Mayo Score (Partial)[f]		X	X	X	X		X	X		X		X		X		X		X				X
Mayo Score (Complete)	X					X														X		
eDiary instruction & review[g]	X	X	X	X	X	X	X	X		X		X		X		X		X		X		
Stool for *C. diff* toxin, pathogenic *E. coli, Salmonella* spp, *Shigella* spp, *Campylobacter* spp or *Yersinia* spp testing	X																					
Stool O&P[h]	X																					
Stool microbiome	X					X						X								X		
Fecal MMP-9	X					X						X								X		

Period	**Screen**	**Treatment**																			**Follow-up**	
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**PTx[a]**	**ET**
Week		**0**	**2**	**4**	**6**	**10**	**11**	**14**	**17**	**20**	**23**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**		
Study Day	**-30 to -1**	**1**	**15**	**29**	**43**	**71**	**78**	**99**	**120**	**141**	**162**	**183**	**211**	**239**	**267**	**295**	**323**	**351**	**379**	**407**		
Visit Window			**±3**	**±3**	**±3**	**±2**	**±3**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	
Fecal Lactoferrin, Calprotectin	X					X						X								X		
Urine drug screen[i]	X																					
Urinalysis	X	X				X														X		
Pregnancy Test[j]	X	X		X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
TB[k]	X																					
Chest x-ray [l]	X																					
HBV, HCV, HIV screening[m]	X																					
HBV DNA monitoring (Japan)[n]				X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	
HBV monitoring (other regions)[n]						X				X				X				X		X		
Hematology	X	X	X	X	X	X		X		X		X		X		X		X		X	X	X
Chemistry	X	X	X	X	X	X		X		X		X		X		X		X		X	X	X
Lipid profile (fasting)[o]		X				X						X								X		X
CRP	X	X	X	X	X	X		X		X		X		X		X		X		X	X	X
Serum immunoglobulin		X		X		X						X								X	X	X

Period	Screen	Treatment																			Follow-up	
Visit	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**PTx[a]**	**ET**
Week		**0**	**2**	**4**	**6**	**10**	**11**	**14**	**17**	**20**	**23**	**26**	**30**	**34**	**38**	**42**	**46**	**50**	**54**	**58**		
Study Day	**-30 to -1**	**1**	**15**	**29**	**43**	**71**	**78**	**99**	**120**	**141**	**162**	**183**	**211**	**239**	**267**	**295**	**323**	**351**	**379**	**407**		
Visit Window			**±3**	**±3**	**±3**	**±2**	**±3**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±5**	**±3**	**±3**	
Blood TCR/BCR repertoire samples[p]		X		X		X						X								X		
Plasma biomarker sample		X		X		X						X								X		
Serum biomarker sample		X		X		X						X								X		
Blood transcriptome sample		X		X		X						X								X		
PK collection (sparse)[q]				X		X						X								X		
vfPBMC[r]		X		X		X						X								X		
PK substudy[s]			X																			
CCI																						
HRQoL Surveys[u]		X				X						X								X		
HCRU questionnaire		X	X	X	X	X	X	X		X		X		X		X		X		X		

a The Post Treatment (PTx) visit should occur 30 days after the last dose of study drug. Only subjects who roll over into the LTE study (GS US 418 3899) will not complete PTx assessments.
b For subjects who terminate prior to Week 10
c A complete physical examination (PE) including, vital signs, body weight, and height will be performed at screening. (GU exam at investigator discretion). A symptom directed PE may be done as needed at other time points.
d Diary data and other eligibility criteria should be reviewed prior to scheduling endoscopy. Perform the screening colonoscopy/flexible sigmoidoscopy within 2 weeks of Day 1 (Day 14 to 1). Subject should meet stool frequency and rectal bleeding requirements based on diary data prior to endoscopy performance.
e The investigator (ie, local endoscopist) should enter in an endoscopic subscore (0, 1, 2, or 3) based on the Mayo Clinic Scoring System endoscopic scale (Protocol Appendix 4); this will be entitled the Investigator Endoscopic Value. The locally read score entered may not be used for eligibility or Mayo Clinic Score calculation purposes, and is not a substitute for the centrally read score, but is collected solely for exploratory purposes.

f Includes all parts of the Mayo score except endoscopy
g Subjects should begin filling out the eDiary the day after their initial screening visit and continue to fill it out throughout the remainder of the study.
h Stool samples should be collected to rule out infectious causes when disease worsens.
i Positive cocaine test disqualifies subject; positive amphetamines, barbiturates, benzodiazepines, and opioids require medical monitor review.
j All females meeting the childbearing potential criteria must have a serum pregnancy testing at screening and a urine pregnancy test must be completed in clinic every 4 weeks at a minimum. If any pregnancy test is positive, study drug should be immediately interrupted and the subject should have a serum pregnancy test in clinic.
k Proof of no active or untreated latent TB at screening. Subjects who are diagnosed with latent TB at screening must initiate an adequate course of prophylaxis as per local standard of care for a minimum of 4 weeks prior to randomization. Subject may initiate study drug dosing only after consultation with the medical monitor.
l Chest x ray (views as per local guidelines) taken at screening or within the 3 months prior to screening (with the report or films available for investigator review) without evidence of active or latent TB infection
m Hepatitis B surface Ag, surface Ab and core Ab, reflex HBV DNA, Hepatitis C Ab, reflex HCV RNA, HIV Ag/Ab, reflext HIV 1/2 Ab at Screening (Protocol Section 6.2.1).
n In Japan, subjects with negative HBsAg, positive HBcAb and/or positive HBsAb at Screening require HBV DNA monitoring every 4 weeks in accordance with local guidelines (Protocol Section 6.2.1). In other regions, subjects with negative HBsAg and positive HBcAb require HBV DNA monitoring every 3 months in accordance with local guidelines (Protocol Section 6.2.1)
o Fasting means no food or drink, except water, for 8 hours
p TCR: T cell receptor; BCR: B cell receptor
q This PK sample at Weeks 10 and 58 are collected at pre dose (within 2 hours prior to dosing). The PK sample at Week 4 is collected post dose (at least 30 minutes and up to 3 hours after study drug dosing . For this visit, it is preferred that study drug dosing is in clinic. The PK sample at Weeks 26 can be collected at any time without regard to dosing. For these visits, the time of dose taken on the day of and the dose taken prior to the PK sample being drawn will be noted in the eCRF.
r vfPBMC collection at US and Canadian sites only.
s Subjects who consent to optional PK substudy will have an additional plasma PK sample at any single visit from Week 2 and 10, collected pre dose and at 0.5, 1, 2, 3, 4 and 6 hours after supervised dosing in the clinic. For all visits with PK sampling, the time of dose taken prior to and on the day of visit will be noted in the eCRF
t [redacted]
u HRQoLs include SF 36, WPAI, EQ 5D, IBDQ.

Appendix 2. Biologics for UC Treatment

The drug names of biologics considered as UC treatment are listed below.

No.	Drug Class	Drug Name
1	TNFα antagonist	Adalimumab
2	TNFα antagonist	Golimumab
3	TNFα antagonist	Infliximab
4	TNFα antagonist	TNFα antagonist biosimilar (to adalimumab, golimumab, or infliximab)
5	Integrin receptor antagonist	Vedolizumab

Appendix 3. Mayo Scoring System for Assessment of Ulcerative Colitis Activity

Stool Frequency – *Each subject serves as his or her own control to establish the degree of abnormality of the stool frequency*	0	Normal number of stools for subject
	1	1 to 2 stools per day more than normal
	2	3 to 4 stools more than normal
	3	≥ 5 stools more than normal
Rectal Bleeding - *The daily bleeding score represents the most severe bleeding of the day*	**0**	No blood seen
	1	Streaks of blood with stool less than half the time
	2	Obvious blood with stool half or more than half of the time
	3	Blood alone passes
Endoscopic findings - *Assessed by Central Reader (include only for MCS assessment)*	0	Normal or inactive disease
	1	Mild Disease *(erythema, decreased vascular pattern)*
	2	Moderate Disease *(marked erythema, lack of vascular pattern, friability, erosions)*
	3	Severe Disease *(spontaneous bleeding, ulceration)*
Physician's Global Assessment - *The physician's global assessment acknowledges the three other criteria, the subject's daily recollection of abdominal discomfort and general sense of wellbeing, and other observation, such as physical findings and the subject's performance status.*	0	Normal
	1	Mild disease
	2	Moderate disease
	3	Severe disease

Appendix 4. Mayo Clinic Score (MCS) Calculation

Definition of the Mayo Clinic Score and Component Subscores

The Mayo Clinic Score (MCS) is a composite index of 4 disease activity variables; stool frequency, rectal bleeding, endoscopic findings, and the physician global assessment (PGA) . Each of the 4 variables is assigned an integer subscore from 0-3 as described in Table 6-2.

- The MCS is calculated as the sum of the 4 subscores for stool frequency, rectal bleeding, endoscopic findings, and the PGA.
- Partial MCS (pMCS) is the sum of the 3 subscores for stool frequency, rectal bleeding, and the PGA.
- EBS (Endoscopy, Bleeding, Stool Frequency) is a description of the 3 subscores for endoscopic findings, rectal bleeding, and stool frequency.

Calculation of Mayo Clinic Score and Component Subscores

In the GS-US-418-3898 study, subjects report daily stool frequency and rectal bleeding symptoms on the study Logpad (an electronic medical diary). The physician reports the PGA at specified visits on the study Sitepad. Results of endoscopic evaluation are reported by a central reader and site personnel enters the score on the Sitepad.

For the purposes of calculation of the specific MCS and component subscores an anchor date is assigned and stool frequency and rectal bleeding records prior to the anchor date are used in the calculation of the subscores.

For screening calculations of MCS and EBS, the anchor date is considered the date of the endoscopy. As endoscopic subscore will be only evaluated over screening and not at Day 1 baseline. Screening MCS and EBS will be used as baseline evaluation.

For all calculations at other visits, including baseline partial MCS, postbaseline MCS, EBS, and partial MCS, the anchor date is the date of the study visit.

Because the preparation for endoscopy procedure may impact the validity of the diary data, the patients reported daily stool frequency and rectal bleeding records collected 1 day prior to, the day of, and the day after the procedure will not be used in the calculation of stool frequency and rectal bleeding subscores. These above mentioned days are considered as non-evaluable days for those 2 subscores.

Calculation of MCS and EBS at Screening/Baseline

The calculation of the screening stool frequency subscore and screening rectal bleeding subscores for the purposes of calculation of baseline MCS and baseline EBS is as follows:

1) The endoscopic procedure date is defined as the E-Day 1.

2) A 7-day window is defined in which the start of the window begins 7 days prior to the endoscopic procedure date and the window ends on the day prior to the endoscopic procedure date.

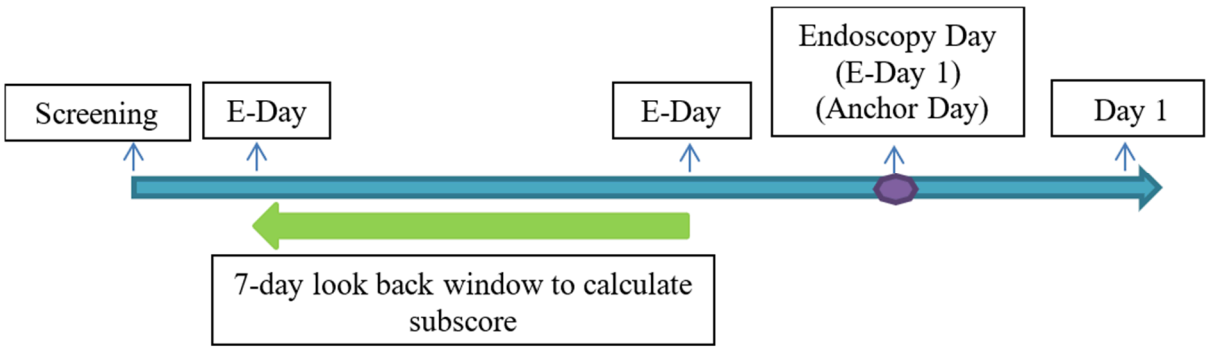
3) The subscore is calculated as the average of the 3 evaluable dairy data entries within the 7-day window closest to the endoscopic procedure date (discarding the day prior to endoscopy) and rounded to the nearest integer.

4) If a subject does not have 3 evaluable diary day entries within the 7-day window, this subject will have a missing subscore and missing baseline MCS and EBS.

5) The PGA subscore used in the baseline MCS is that recorded during the screening period (prior to Day1) .

A schema of this approach (Appendix Figure 1) and examples are included below in Appendix Table 1:

Appendix Figure 1. Stool Frequency and Rectal Bleeding Subscores Calculations at Screening



Appendix Table 1. Examples of Calculations for the Rectal Bleeding and the Stool Frequency Subscores of MCS and EBS at Baseline

	Diary Day Looking Back from Endoscopy Date							Anchor Day			
	-7	-6	-5	-4	-3	-2	-1	E-Day 1			
Example	3 Jun	4 Jun	5 Jun	6 Jun	7 Jun	8 Jun	9 Jun	10 Jun	Days for Calculation	Average Subscore	Final Subscore
Diary 1	M	M	2	M	1	2	X	X	2, 3, 5	1.67	2
Diary 2	M	M	M	M	0	1	X	X	2, 3	Missing	Missing
Diary 3	M	M	M	M	0	M	X	X	3	Missing	Missing
Diary 4	M	2	M	2	1	1	X	X	2, 3, 4	1.33	1

Days are named relative to Anchor Day Endoscopy Date (E Day 1), M Missing, X non evaluable
The diary data collected on E Day 1 and E Day 1 will be discarded as endoscopy and its preparation may impact the validity of the diary data.

Calculation of Partial MCS at Baseline, and Postbaseline MCS, EBS, and Partial MCS

The calculation of baseline partial MCS and postbaseline MCS, EBS and partial MCS use the study visit date as the anchor date. For baseline pMCS, the visit date is the Day 1 visit. The calculations are as follows:

1) Subject diary data collected within 7 days prior to the visit day will be used for calculation, and some of the days within this 7-day window may be considered non-evaluable due to the endoscopy procedure and its preparation (ie, the day of the endoscopy and the day prior to the endoscopy).

2) The average of the 3 evaluable diary day entries within the 7-day window closest to the visit day (V-Day 1), rounded to the nearest integer, will be considered the stool frequency and rectal bleeding subscore for that visit

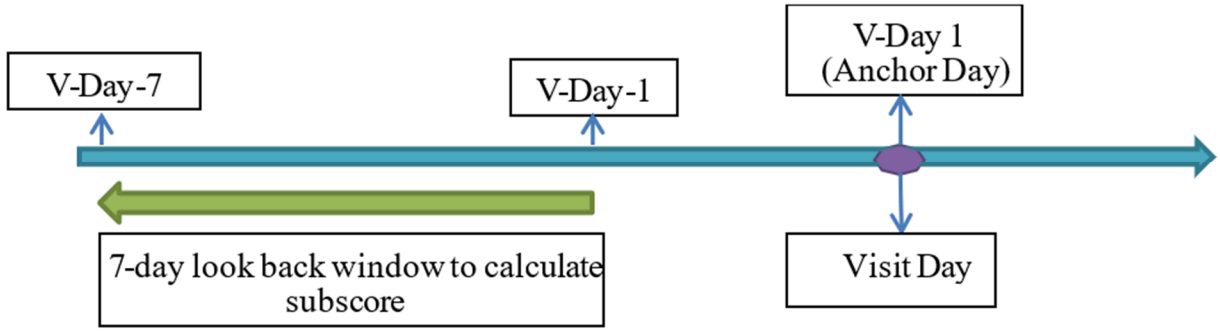
3) The PGA subscore used in the calculation of baseline partial MCS is the PGA subscore recorded on Day 1. At subsequent visits, the PGA subscore used in calculation of MCS and partial MCS is that recorded within the visit window

4) If there are 2 or less evaluable diary day entries for stool frequency or rectal bleeding within the 7-day window, no missing data imputation will be done. MCS, EBS and partial MCS will be considered as missing.

5) For MCS, if endoscopic subscore or PGA subscore is missing, MCS will be left as missing. For Partial MCS, if PGA is missing, partial MCS will be left as missing.

A schema of this approach and some examples are included below in Appendix Figure 2 and Appendix Table 2:

Appendix Figure 2. Stool Frequency and Rectal Bleeding Subscores Calculations at Baseline and Postbaseline



Appendix Table 2. Calculation of Partial MCS at Baseline, and Postbaseline MCS, EBS and Partial MCS

	Diary Day Looking Back from Visit Date							Anchor Day			
	-7	-6	-5	-4	-3	-2	-1	V-Day 1			
Example	3 Jun	4 Jun	5 Jun	6 Jun	7 Jun	8 Jun	9 Jun	10 Jun	Days for Calculation	Average Subscore	Final Subscore
Diary 1	E	X	3	M	M	M	0	X	-1, -5	Missing	Missing
Diary 2	4	4	X	E	X	3	3	X	-1, -2, -6	3.33	3
Diary 3	3	X	E	X	1	M	2	X	-1, -3, -7	2	2
Diary 4	2	3	4	4	X	E	X	X	-4, -5, -6	3.67	4
Diary 5	2	M	M	X	E	X	2	X	-1, -7	Missing	Missing
Diary 6	X	E	X	2	3	0	1	X	-1, -2, -3	1.33	1
Diary 7	M	3	X	E	X	M	M	X	-6	N/A	Missing

Days are named relative to Anchor Day (V Day 1 Visit Date), M Missing, X non evaluable, E Endoscopy Day

Sources of Information for Mayo Clinic Score Calculations

The sources of information used for the calculations of MCS and partial MCS are summarized in Appendix Table 3 (the calculated MCS and partial MCS from ERT data will not be used for data analysis):

Appendix Table 3. Sources of Information for Calculating MCS and Partial MCS

Variable in Calculation	Data Source
Endoscopy Procedure Date	ENDODAT in ENDO dataset from CRF data for protocol-directed endoscopy; ENDODAT in MAYO dataset from ERT data for non-protocol-directed endoscopy
Visit Date	VISITDAT in VISDT dataset from CRF data
Stool Frequency Subscore	QSSTRESN in DIARY dataset with QSTESTCD=”SF” from ERT data
Rectal Bleeding Subscore	QSSTRESN in DIARY dataset with QSTESTCD=”RB” from ERT data
Physician Global Assessment	QSSTRESN in MAYO dataset with QSTESTCD=”PGA” from ERT data
Endoscopic Score	MAYO in ENDO ROBARTS dataset from Robarts data

Note: Robarts is the vendor performing endoscopy central scoring;
ERT is the vendor that electronically collects patient and investigator reported outcomes

Appendix 5. Study Treatment Failure Rules

Study treatment failure rules apply to treatment failure that occurs during the course of the study. Study treatment failure rules will be applied to all efficacy endpoints unless otherwise specified. All efficacy data will be censored (set to missing) after treatment failure criteria are met, regardless of the observed data. Subjects who do not have sufficient measurements after censoring to determine the dichotomized endpoint(s) will be considered non-responder for corresponding endpoint(s). Study treatment failure rules override other data handling rules.

Subjects who have any of the following events will be considered a study treatment failure after the earliest event through the end of the Induction Study (if the event occurred during induction) or through the end of the Maintenance Study (if the event occurred during maintenance), regardless of the actual efficacy data collected. Induction baseline will be referenced when calculating the potentially effective use for induction studies and maintenance baseline (re-baseline) will be referenced when calculating the potentially effective use for Maintenance Study.

1) Potentially effective steroid use

Potentially effective corticosteroids, for the purpose of this SAP, are steroids that may impact disease under study. **Potentially effective steroids** include the following steroids when administered for an indication of UC:

a) Commencement of:

Any steroid administered intramuscularly (IM), intravenously (IV), orally (PO), or rectally (PR) at any dose for 7 or more continuous days

- This rule applies regardless if a change in drug, dose, or route (from IM to IV to PO or PR or vice versa) occurs within the seven continuous days; and it includes oral steroids with intended local actions (eg, budesonide)

b) Escalation of concomitant steroid dose above the baseline dose for 7 or more continuous days. The baseline steroid dose is defined as the dose at Day 1 of induction if the escalation occurs in the Induction Study, or Day 1 of maintenance (re-baseline) if the escalation occurs in the Maintenance Study. The prednisone equivalent dose will be used to determine escalation of concomitant systemic steroid dose even if there is a change in drug or route. For steroids with local actions (eg, budesonide or any rectally administered steroid), this rule will apply to scenarios where the post-baseline dose is above the baseline dose via the same drug and the same route for 7 or more continuous days.

Potentially effective steroid use only includes use of steroids administered via routes that are IM, PO, IV, or PR. The below steroids will **not** be considered potentially effective steroids regardless of indication for use:

i) ocular steroids (ie, eye drops)

ii) topical steroids (eg, cutaneously applied solely to the skin, or topically applied to the nasal mucosa)

iii) inhaled steroids (eg, inhalational fluticasone for asthma)

iv) intra-articular steroids (steroids administered directly into a joint)

v) neuraxial steroids (steroids administered into the epidural or spinal space)

2) Potentially effective immunomodulator use

Commencement of a different class of oral, IM, SC, or IV immunomodulator drugs (where the subject was not previously taking concomitant immunomodulators of the same class on Day 1 of induction or maintenance), including but not limited to 6-MP, azathioprine, MTX, 6-thioguanine, and prohibited immunomodulators including but not limited to cyclosporine, leflunomide, tacrolimus, thalidomide regardless of dose, for 7 continuous days. The use of an immunomodulator will be considered potentially effective when it is administered for an indication of UC.

3) Potentially effective biologic agent use

Commencement of a biologic agent including but not limited to TNFα antagonists, IL-12/23 antagonists, and vedolizumab (or similar agents), regardless of indication for use and duration.

Appendix 6. Geboes Histological Score Grades

	Subgrades	Structural (Architectural Change)
Grade 0	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**
	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

Appendix 7. PK Parameters

PK parameters evaluated in this study are listed below.

Parameter	Description
AUC_{last}	The area under the concentration versus time curve from time zero to the last quantifiable concentration
AUC_{tau}	area under the concentration versus time curve over the dosing interval
C_{max}	maximum observed concentration of drug in plasma
C_{tau}	observed drug concentration at the end of the dosing interval
T_{max}	time (observed time point) of C_{max}
CL/F	apparent oral clearance after administration of the drug: at single dose: CL/F = Dose/AUC_{inf} , where “Dose” is the dose of the drug at steady state: CL/F = Dose/AUC_{tau}, where “Dose” is the dose of the drug

GS-US-418-3898-SAP-v3.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Clinical Research eSigned	01-May-2020 23:30:45
PPD	Biostatistics eSigned	01-May-2020 23:56:35