STATISTICAL ANALYSIS PLAN

Study Title:	A Randomized, Double-blind, Placebo-controlled Phase 2 Study to Evaluate the Testicular Safety of Filgotinib in Adult Males with Moderately to Severely Active Inflammatory Bowel Disease
Name of Test Drug:	Filgotinib
Study Number:	GS-US-418-4279
Protocol Version (Date):	Original: 31 January 2017 Amendment 1: 06 July 2017 Amendment 2: 22 August 2018 Amendment 3: 18 January 2019 Amendment 4: 18 March 2020
Analysis Type:	Interims (Week 26 and Week 26 plus Reversibility) and Final Analysis
Analysis Plan Version:	Version 1.0
Analysis Plan Date:	13 January 2021
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
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BLQ	below the limit of quantitation
BMI	body mass index
CD	Crohn's disease
CDAI	Crohn's Disease Activity Index
CI	confidence interval
COVID-19	Coronavirus Disease 2019
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CVEAC	Cardiovascular Event Adjudication Committee
DB	double-blind
DMC	Data Monitoring Committee
DVT	deep vein thrombosis
EAIR	exposure-adjusted incidence rates
ECG	electrocardiogram
eCRF	electronic case report form
ET	early termination
FDA	Food and Drug Administration
FSH	follicle-stimulating hormone
FU	follow-up
GLPG	Galapagos, NV
HLGT	high-level group term
HLT	high-level term
IBD	inflammatory bowel disease
IC	inclusion criteria
ID	Identification
IPD	important protocol deviation
ITT	Intent-to-Treat
IUT	Internal Unblinded Team
IWRS	interactive web response system
LH	luteinizing hormone
LLT	low-level term
LOQ	limit of quantitation
LTE	long term extension

M	Million
MACE	major adverse cardiac events
MAR	missing at random
MCS	Mayo Clinic Score
MedDRA	Medical Dictionary for Regulatory Affairs
MH	Mantel-Haenszel
MI	multiple imputation
MST	MedDRA search term
MTX	Methotrexate
OI	opportunistic infection
OL	open-label
PE	pulmonary embolism
PGA	Physician's Global Assessment
PK	Pharmacokinetics
pMCS	partial Mayo Clinic Score
PT	preferred term
PTM	placebo-to-match
PYE	patient-years of exposure
Q1	first quartile
Q3	third quartile
RB	rectal bleeding
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SE	standard error
SF	stool frequency
SMQ	standardized MedDRA query
SOC	system organ class
TEAEs	treatment emergent adverse events
TFLs	tables, figures and listings
UC	ulcerative colitis
ULN	upper limit of normal
US	United States
VTE	venous thromboembolism event
WHO	World Health Organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the statistical analysis methods and data presentations to be used in tables, figures, and listings (TFLs) for Study GS-US-418-4279 [MANTA]. This SAP is based on study protocol Amendment 4 dated 18 March 2020 and the electronic case report form (eCRF). The SAP will be finalized prior to unblinding for the Week 26 interim analysis after at least 200 evaluable subjects pooled across studies GS-US-418-4279 [MANTA] and GLPG0634-CL-227 [MANTA-RAy] complete Week 13; and evaluable subjects from Study GS-US-418-4279 complete Week 26 or permanently discontinue from study drug. Any changes made after the finalization of the SAP will be documented in the clinical study report (CSR).

1.1. Study Objectives

The primary objective of the study is as follows:

- To evaluate the effect of filgotinib on testicular function as defined by the proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 13

The secondary objectives of the study are as follows:

- To evaluate the effect of filgotinib on testicular function as defined by the proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 26
- To evaluate the effect of filgotinib on sperm total motility at Weeks 13 and 26
- To evaluate the effect of filgotinib on total sperm count at Weeks 13 and 26
- To evaluate the effect of filgotinib on the change from baseline in sperm concentration at Weeks 13 and 26
- To evaluate the effect of filgotinib on ejaculate volume at Weeks 13 and 26
- To evaluate the effect of filgotinib on percent normal sperm morphology at Weeks 13 and 26

The exploratory objectives of the study are as follows:

- To evaluate the reversibility of observed effects of filgotinib on testicular function in subjects who experience a ≥ 50% decrease in sperm concentration and/or motility and/or morphology
- To evaluate the effect of filgotinib on hormones, including luteinizing hormone (LH), follicle-stimulating hormone (FSH), inhibin B, and total testosterone at Weeks 13 and 26
- To evaluate the safety and tolerability of filgotinib
- To characterize the plasma pharmacokinetics (PK) of filgotinib and its metabolite (GS-829845, formerly CCI)

1.2. Study Design

This is a randomized, double-blind, placebo-controlled, Phase 2 study in adult males with moderately to severely active inflammatory bowel disease (IBD).

Up to 250 males were planned to be randomized in a 1:1 ratio to receive up to 26 weeks of filgotinib 200 mg or placebo-to-match (PTM) once daily, to ensure approximately 200 evaluable subjects. Protocol Amendment 3 randomization was stratified by type of disease (UC or CD), concurrent use of methotrexate (MTX; yes or no), and sperm concentration (15 to 25 million [M]/mL, > 25 to 50 M/mL, or > 50 M/mL). Subjects may have been on protocol-specified concomitant therapy. Original and Protocol Amendment 1 randomization was stratified by concurrent use of MTX (yes or no) only. No subjects were randomized under Protocol Amendments 2 or 4.

The main inclusion criteria relevant to this study are:

- Male subjects between the age of 21 and 65 (inclusive) on the day of signing informed consent
- The mean of 2 separate evaluable semen samples collected at the screening visit must meet the following minimum criteria (Protocol Amendment 4, Section 6.14): semen volume ≥ 1.5 mL, total sperm/ejaculate ≥ 39 M/ejaculate, sperm concentration ≥ 15 M/mL, sperm total motility ≥ 40%, and percent normal sperm morphology ≥ 30% using World Health Organization (WHO) 2010 criteria {World Health Organization (WHO) 2010}).

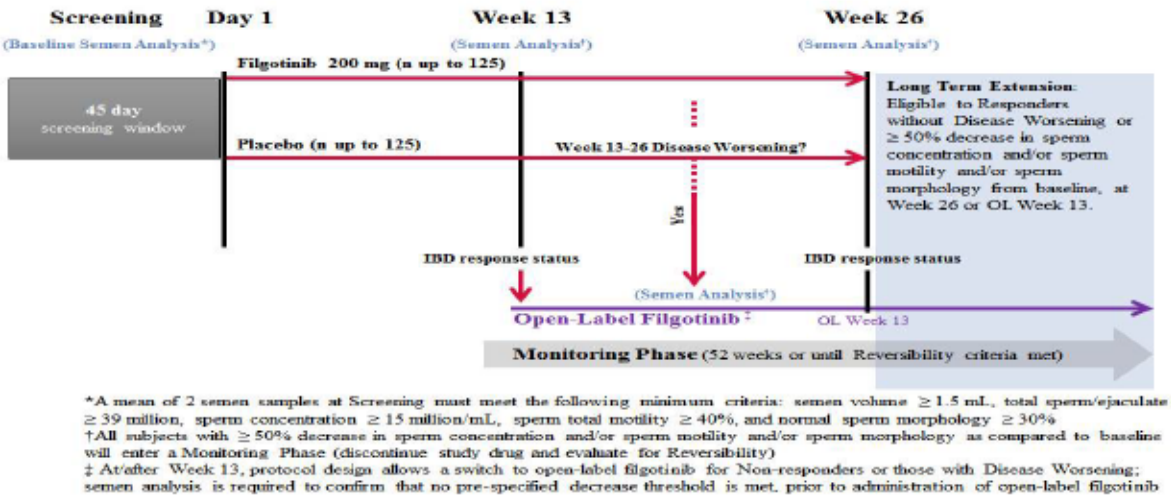
The main exclusion criterion relevant to this study is:

- Previously or currently documented problems with male reproductive health, including but not limited to primary hypogonadism, secondary hypogonadism, or reduced fertility

A complete list of eligibility criteria is provided in Sections 4.2 and 4.3 of Study Protocol Amendment 4. The Screening period was up to 45 days prior to randomization. The assessments planned to be performed at each visit are detailed in the study procedures table (Appendix 1).

The study includes 5 parts: Part A (Double-blind Phase: Day 1 through Week 13 study visit); Part B (Double-Blind Phase: after Week 13 through Week 26 study visit); Open-Label (OL) Filgotinib Phase (OL Day 1 to OL Week 13 visit); Long Term Extension (LTE, double-blind or OL for up to 195 weeks); and Monitoring Phase (up to 52 weeks). A schematic of the study design is provided in Figure 1-1.

Figure 1-1. GS-US-418-4279 (MANTA): Study Schema (Protocol Amendment 4)



*A mean of 2 semen samples at Screening must meet the following minimum criteria: semen volume ≥ 1.5 mL, total sperm/ejaculate ≥ 39 million, sperm concentration ≥ 15 million/mL, sperm total motility ≥ 40%, and normal sperm morphology ≥ 30%
†All subjects with ≥ 50% decrease in sperm concentration and/or sperm motility and/or sperm morphology as compared to baseline will enter a Monitoring Phase (discontinue study drug and evaluate for Reversibility)
‡ At/after Week 13, protocol design allows a switch to open-label filgotinib for Non-responders or those with Disease Worsening; semen analysis is required to confirm that no pre-specified decrease threshold is met, prior to administration of open-label filgotinib

Subjects will visit the clinical study center at Screening, Day 1, Week 2, Week 4, Week 8, and Week 13 (Part A of study). Subjects who are IBD responders at Week 13 without meeting pre-specified sperm decrease thresholds will continue on blinded treatment and return for Week 20 and Week 26 assessments. A pre-specified sperm decrease threshold was defined as a ≥ 50% decrease from baseline in sperm concentration prior to Protocol Amendment 2; and a ≥ 50% decrease from baseline in sperm concentration and/or motility and/or morphology at/after Protocol Amendment 2. Since no subject demonstrated a ≥ 50% decrease in sperm motility or morphology prior to Protocol Amendment 2, this SAP will hereafter refer to a "pre-specified sperm decrease threshold" as a subject demonstrating a ≥ 50% decrease from baseline in sperm concentration and/or motility and/or morphology. Subjects who are non-responders at Week 13 and subjects who have disease worsening after Week 13 up to Week 26 who do not meet pre-specified sperm decrease thresholds will enter the Open-Label Phase and return for visits at Week 2, Week 4, Week 8, and Week 13 during the open-label phase. Subjects who are responders at Week 26/Open-label Week 13 without disease worsening and without meeting pre-specified sperm decrease thresholds may continue taking study drug in the LTE phase for up to 195 weeks (or until they have disease worsening or meet pre-specified sperm decrease threshold[s]).

Subjects who meet pre-specified sperm decrease threshold(s) at any postbaseline visit will stop study drug, switch to a standard of care regimen selected by the investigator (in accordance with the protocol) and enter the Monitoring Phase. Subjects will undergo semen evaluations every 13 weeks starting from their entry into the Monitoring Phase for up to 52 weeks or until reversibility is met, whichever is achieved first. Reversibility is defined as a return to > 0.5 x (mean of 2 evaluable samples collected at screening visit) of all sperm parameter(s) qualifying the subject for the monitoring phase—based on 2 evaluable samples collected at the monitoring phase visit.

Prior to Protocol Amendment 2, subjects who switched to OL filgotinib and subjects who discontinued study drug after meeting pre-specified sperm decrease threshold(s) continued to attend visits scheduled relative to the date of their first dose of double-blind study drug (eg, the Week 26 visit was scheduled approximately 26 weeks after the first dose of double-blind study drug). However, because of the time required to collect and evaluate semen samples, subjects may not have switched or discontinued, respectively, until after the Week 13 visit. Therefore, a full 13 weeks (ie, a full spermatogenic cycle) may not have passed before their next scheduled semen assessment under the new condition (ie, since switching to OL filgotinib or discontinuing study drug, respectively). At/after Protocol Amendment 2, study visits in the OL Filgotinib Phase and Monitoring Phase are scheduled relative to the first day of starting OL filgotinib and the last day of study drug, respectively, allowing a full 13 weeks before the first semen assessment under the new condition.

Please refer to the study protocol for detailed descriptions of the visit schedule and respective evaluations.

1.3. Sample Size and Power

Results of this study will be pooled with the results of a separate study being conducted in subjects with rheumatic diseases (GLPG0634-CL-227 [MANTA-RAy]) with the same objectives. The total planned number of subjects in both studies combined will be up to approximately 250 subjects.

A sample size of approximately 100 evaluable subjects per group (ie, 100 subjects in the filgotinib 200 mg group and 100 subjects in the placebo group) is adequate for the purposes of estimating cumulative distribution curves and producing a 95% confidence interval (CI) width that is reasonably narrow for the proportion of subjects in each group who experience a $\geq$ 50% decrease in sperm concentration compared to baseline. An evaluable subject is defined as a subject with 2 semen samples at both baseline and at Week 13 that are eligible for mean calculation for sperm concentration, with visit window assignment for the mean calculation based on the date of the first of 2 evaluable sample collections. This 200 subject sample size is suggested in FDA's guidance on testicular toxicity studies {U. S. Department of Health and Human Services (DHHS) 2018}. Assuming a 20% rate of non-evaluable subjects at Week 13, up to 250 subjects may be enrolled to obtain 200 evaluable subjects.

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee Analyses

An external DMC reviews the progress of the study, performs unblinded interim reviews of safety data, and provides a recommendation to Gilead whether the nature, frequency, and severity of adverse effects associated with study treatment warrant the early termination of the study in the best interests of the participants, whether the study should continue as planned, or the study should continue with modifications. The initial unblinded data review meeting occurred after approximately the first 50 subjects completed the Week 13 visit. Following this, subsequent unblinded data review meetings were to occur approximately every 4 months (if enrollment supports the need). Pooled sperm/semen data from Studies GS-US-418-4279 [MANTA] and GLPG0634-CL-227 [MANTA-RAy] were reviewed after approximately 100 subjects complete Week 13.

The DMC may additionally request to review blinded sperm/semen data prior to the first unblinded DMC meeting and/or in between unblinded review meetings. Ad-hoc DMC meetings may be scheduled as needed. Ad-hoc DMC meetings may be triggered for reasons outlined in the DMC Charter and/or protocol.

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

While the DMC will be asked to advise Gilead regarding the future conduct of the study, including possible early study termination, Gilead retains final decision-making authority on all aspects of the study.

2.2. Internal Independent Safety Review

A Sponsor Internal Unblinded Team (IUT), independent of the blinded study team, will be assembled. The Sponsor IUT will monitor blinded cumulative interval sperm/semen data and may be granted access to unblinded clinical data at pre-specified timepoints to review sperm/semen data. This internal review team will be supported by an external Urology expert. To mitigate the risk of inadvertently releasing treatment assignment to sites and subjects, the internal team will keep unblinded information confidential and will not communicate any information to the blinded study team, site staff or subjects. To date, the Sponsor IUT has not been utilized to review unblinded semen data at pre-specified timepoints to monitor safety (DMC has performed this duty). If the Sponsor IUT is utilized to review unblinded data for the purposes of monitoring safety prior to unblinding of the study, the committee's specific activities will be defined by a mutually agreed upon charter, which will define the committee membership, conduct, and meeting schedule. Data unblinding due to medical emergency will follow standard Gilead procedures.

2.3. Week 26 Analysis

The unblinded Week 26 analysis of Study GS-US-418-4279 will be performed after at least 200 evaluable subjects pooled across studies GS-US-418-4279 [MANTA] and GLPG0634-CL-227 [MANTA-RAy] complete Week 13 and evaluable subjects from Study GS-US-418-4279 complete Week 26 or permanently discontinue from study drug. This analysis will be conducted in order to fulfill ex-US regulatory commitments.

A Data Integrity and Communication Plan will be developed prior to unblinding for any interim analysis. The GS-US-418-4279 study team will remain blinded to double-blind treatment assignments (where possible by study design) until the final analysis when the database has been locked, and the study has been completely unblinded.

2.4. Week 26 and Reversibility Analysis for Pre-specified Sperm Decreases Up to Week 26

A separate unblinded Week 26 analysis of Study GS-US-418-4279 will be performed after all subjects reach Week 26 and all subjects with a ≥ 50% decline in sperm concentration and/or motility and/or morphology up to Week 26 (or Open-Label Week 13) enter the monitoring phase and either meet reversibility criteria or are followed off-treatment for 52 weeks, whichever occurs sooner. This analysis will be conducted at the request of the US Food and Drug Administration.

The Data Integrity and Communication Plan will continue to be in place, and the GS-US-418-4279 study team will remain blinded to double-blind treatment assignments (where possible by study design) until the final analysis when the database has been locked and the study has been completely unblinded.

2.5. Final Analysis

After all subjects have completed the study (eg, completed Week 195 and did not meet pre-specified sperm decrease thresholds [a ≥ 50% decrease in sperm concentration and/or motility and/or morphology]; met pre-specified sperm decrease threshold criteria and either met reversibility criteria or completed 52 weeks of off-study drug follow-up for reversibility; or prematurely discontinued from 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.

All statistical tests will be 2-sided and performed at the 5% significance level unless otherwise specified.

By-subject listings will be presented for all subjects in the All Randomized Analysis Set and be sorted by subject ID number, visit date, and time (if applicable). Data collected on log forms, such as AEs, will be presented in chronological order within the subject. The treatment group to which subjects were randomized or the treatment sequence received will be presented in the listings. Age, sex, race, ethnicity, start and end date of the double-blind study drug and open-label study drug (if applicable) will be included in the listings, as space permits. A by-subject listing will be provided for subjects with any visits or assessments impacted by Coronavirus Disease 2019 (COVID-19).

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 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 Set

The All Randomized Analysis Set includes all subjects who were randomized in the study. This is the primary analysis set for randomization summaries and by-subject listings.

3.1.2. Semen Analysis Sets

The Semen Analysis Set includes all randomized and treated (received $\geq$ 1 dose of double-blind study drug) subjects who have 2 semen samples that are eligible for mean calculation at baseline and at the Week 13 analysis visit with the date of the first chronologic semen sample used for purposes of assigning analysis visit windows.

An individual semen sample is eligible for mean calculation if it met criteria outlined in the protocol with respect to ejaculation-free time and days between evaluable semen collections. Subject ID 10603-35043 had a third sample collected 17 days after the second sample at Week 13 when it was noted that his first sample had been collected prior to the lower bound of

the Week 13 analysis visit window (first 2 semen samples were collected within 14 days of each other; third sample within Week 13 analysis visit window). This sample will be considered an exception and will be "eligible for mean calculation". Four additional subjects (Subject IDs 14620-35106, 14930-35115, 14620-35116, 14620-35119) had Week 13 semen sample collections late for the Week 13 analysis visit window due to COVID-19 shelter in place orders and travel restrictions. These samples will be used for the Week 13 timepoint as subjects had more than 13 weeks of exposure to double-blind study drug when semen samples were collected.

The Week 26/LTE Semen Analysis Set includes all subjects who took ≥ 1 dose of study drug and have 2 evaluable samples at baseline and at ≥ 1 post baseline measurement at/after Week 26 or Open-Label Week 13. This analysis set will be used for analyses of sperm and hormone data by treatment sequence.

The primary analysis of semen data for the primary and secondary endpoints at double-blind Week 13 will be based on the Semen Analysis Set and subjects with 2 evaluable semen samples which are required at each semen sample collection time point. A supportive analysis (Section 6.4) will also include subjects in the following analysis set:

The Intent-to-Treat (ITT) Semen Analysis Set includes all randomized subjects who have 2 semen samples that are eligible for mean calculation at Baseline.

Since all subjects who were randomized took at least 1 dose of study drug and also had 2 evaluable semen samples at baseline, the Intent-to-Treat Semen Analysis Set will include the same subjects as the Safety Analysis Set.

3.1.3. Safety Analysis Sets

The Safety Analysis Set includes all randomized subjects who took ≥ 1 dose of double-blind study drug. This is the analysis set utilized for safety analyses during the double-blind phase of the study.

The Open-Label Safety Analysis Set includes all randomized subjects who took ≥ 1 dose of open-label filgotinib study drug in the Open-Label Phase of the study. This analysis set will be used for selected safety displays during open-label filgotinib treatment (eg, brief summary of AEs, interruption of open-label filgotinib, discontinuation of open-label filgotinib).

The As-Treated Safety Analysis Set includes all subjects in the Safety Analysis Set; subjects initially randomized to placebo who switched to open-label filgotinib will also be included for the filgotinib group. This analysis set will be utilized to summarize all available on-treatment safety data during the study.

3.1.4. Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analysis Set includes all subjects who took ≥ 1 dose of filgotinib (double-blind or open-label study drug) who have at least 1 nonmissing postdose concentration value for filgotinib and/or its metabolite GS-829845. This is the primary analysis set for all PK analyses.

3.2. Subject Grouping

For analyses based on the All Randomized and ITT Semen Analysis Sets, subjects will be grouped according to their randomized treatment. For all other analysis sets, subjects will be grouped according to the actual treatment received. The actual treatment received will differ from the randomized treatment only when the actual treatment differs from randomized treatment for the entire treatment duration of the study drug administration period.

The following subject groupings will be applied to the semen and hormone data according to the actual treatment received:

- For the double-blind treatment period, data from subjects who took double-blind treatment will be summarized by treatment group (ie, filgotinib 200 mg and placebo).

- Semen and hormone data collected after Week 13 will be summarized by treatment sequence (ie, double-blind filgotinib, double-blind filgotinib to open-label filgotinib, double-blind placebo, and double-blind placebo to open-label filgotinib) for subjects with baseline and ≥ 1 visit at/after Week 26. Analysis visit windows for the double-blind placebo to open-label filgotinib group will be windowed from the first dose date of open-label filgotinib and will be 13-weeks prior to those who remained on original treatment. All other treatment sequences will have analysis visit windows assigned based on days from first dose of double-blind study drug.

- Subjects with a decrease of ≥ 50% in sperm concentration and/or motility and/or morphology from baseline will have study drug stopped and an investigator selected standard of care treatment that is in accordance with the protocol started. Semen samples will be collected off study drug treatment every 13 weeks during the Monitoring Phase until reversibility is met or for up to 52 weeks, whichever occurs first. Reversibility is met when all sperm parameter(s) qualifying the subject to enter the Monitoring Phase return to greater than 50% of the baseline value (ie, greater than 0.5 x [mean of 2 evaluable samples collected at the screening visit]). Data will be summarized for each sperm parameter meeting a pre-specified sperm decrease threshold at the visit when study drug was stopped. The number (percent) who met reversibility criteria and the cumulative number (percent) who met reversibility criteria at each Monitoring Phase visit will be presented separately for subjects who met pre-specified sperm decreases at Week 13 (summarized by double-blind treatment group) and for subjects meeting pre-specified sperm parameter decrease thresholds at the Week 26/OL Week 13 analysis visit or at/after Week 39/OL Week 26 and up to last dose date of any study drug + 7 days (summarized by treatment sequence).

The following subject groupings will be applied to safety data according to the actual treatment received (filgotinib or placebo) unless otherwise stated:

- An "as-treated" analysis will be performed for safety data. Subjects who took double-blind placebo and switched to open-label filgotinib will have data reported during the placebo period summarized with the placebo group, and data reported during the open-label filgotinib period summarized with the filgotinib group (baseline will be open-label baseline). Subjects who switch from double-blind filgotinib to open-label filgotinib and subjects who stay on double-blind study drug for the duration of study will be summarized by their double-blind treatment group (baseline will be double-blind baseline).

  Adverse events will additionally be displayed for the double-blind treatment phase (to align with the period summarized for the primary analysis of semen data). Data will be summarized by double-blind treatment group (ie, filgotinib 200 mg or placebo) from first dose date of double-blind study drug up to the first dose date of open-label filgotinib (minus 1 day for AEs) for subjects who switch to open-label filgotinib; to last dose date of double-blind study drug + 30 days for subjects who stop double-blind study drug and do not switch to open-label filgotinib; or all available data at time of interim analysis (for subjects ongoing on double-blind treatment at the time of an interim analysis).

  Select AE summaries (eg, brief summary of AEs while on open-label filgotinib, AEs leading to interruption of open-label filgotinib, and AEs leading to discontinuation of open-label filgotinib) will be displayed for subjects in the Open-Label Filgotinib Safety Analysis Set. AEs occurring from first to last dose date of open-label filgotinib plus 30 days (for those permanently discontinued from drug) or all events on/after open-label first dose date for those ongoing on open-label filgotinib will be included.

3.3. Strata and Covariates

Subjects randomized under Protocol Amendment 3 were randomly assigned to treatment groups via the interactive web response system (IWRS) in a 1:1 ratio using a stratified randomization schedule. Stratification was based on the following variables:

- Type of IBD disorder (UC, CD)
- Concurrent use of MTX (Yes, No)
- Sperm concentration (mean of 2 evaluable semen samples) measured at the Screening visit: 1) 15 to 25 M/mL; 2) > 25 to 50 M/mL; 3) > 50 M/mL

Subjects randomized under the original protocol or under Protocol Amendment 1 were randomized based on a stratification factor of concurrent use of MTX (yes, no). These subjects will be assigned to disease type of UC and the respective sperm concentration stratum based on the mean of 2 evaluable sperm concentration measurements collected at the screening visit, respectively.

If there are discrepancies in stratification factor values between the IWRS and the clinical database, the values recorded in the clinical database will be used for analysis.

3.4. Examination of Subject Subgroups

Subgroup analyses for the primary endpoint are specified in Section 6.3.

3.5. Missing Data and Outliers

3.5.1. 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 dosing date of study drug, imputation rules are described in Section 4.2.1. The handling of missing or incomplete dates for AE onset is described in Section 8.1.5.2, and for prior and concomitant medications in Section 8.4.

3.5.2. Outliers

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

3.6. Data Handling Conventions and Transformations

In general, age (in years) on the date of the first dose of any study drug will be used for analyses and presentation in listings. If a randomized subject was not dosed with any study drug, the randomization date will be used instead of the first dosing date of study drug for age calculation. For screen failures, the date the informed consent was signed will be used for age calculation. If only the 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.

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 to calculate 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 to calculate 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 to calculate summary statistics.
- A value that is 1 unit above the LOQ will be used to calculate 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 to calculate descriptive statistics if the datum is reported in the form of "≤ x" or "≥ x" (where x is considered the LOQ).

Plasma concentration values that are below the limit of quantitation (BLQ) will be presented as "BLQ" in the concentration data listing.

3.7. Analysis Visit Windows

3.7.1. Definition of Study Day

The First Dosing Date of Filgotinib/Placebo is defined as the date when subjects take their first dose of filgotinib/placebo, either double-blind or open-label drug, whichever is the earliest as recorded in the Study Drug Administration eCRF.

The Last Dosing Date of Filgotinib/Placebo is defined as the date when subjects take the last dose of filgotinib/placebo, either double-blind or open-label drug, whichever is later as recorded in the Study Drug Administration eCRF.

Study Day will be calculated from the first dosing date of double-blind filgotinib/placebo and derived as follows:

- For postdose study days: Assessment Date – On-Filgotinib/Placebo Study Day 1 + 1
- For days prior to the first dose: Assessment Date – Open-Filgotinib/Placebo Study Day 1

For subjects who are on placebo and switch to open-label filgotinib, the On-Filgotinib Study Day 1 is defined as the first dose date of open-label filgotinib.

Follow-up (FU) Day for subjects who discontinue study drug and/or enter the Monitoring Phase will be calculated from the last dosing date of any study drug, which is defined as the date when a subject takes their last dose of any study drug considering both double-blind and open-label drugs. The FU day is derived as Assessment Date – Last Dosing Date. Therefore, Follow-up Day 1 is the first day after permanent discontinuation of all study drugs.

Baseline is defined as the last available observation taken on or prior to the first dose date of double-blind filgotinib/placebo study drug.

Open-label Baseline is defined as the last available observation taken on or prior to the first dosing date of open-label filgotinib (used for safety assessments for placebo subjects who switch to open-label filgotinib with the exception of semen and hormone parameters). For semen parameters, including sperm concentration, total sperm count, ejaculate volume, sperm total motility, and sperm morphology, the baseline value is the mean of 2 evaluable samples from the screening visit. The baseline value for sperm/semen parameters is used to evaluate whether subjects have met pre-specified sperm parameter decrease thresholds (ie, $\geq 50\%$ decrease in sperm concentration and/or motility and/or morphology). For hormone parameters, the baseline value will be the last value on/prior to starting double-blind study drug, to be consistent with the rules that are used for sperm/semen data.

3.7.2. Analysis Visit Windows

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

The analysis visit windows for lipid profile, serum immunoglobulin and urinalysis are provided in Table 3-1. The analysis visit windows for pMCS, CDAI, vital signs, weight, hematology and chemistry laboratories are provided in

Table 3-2. The analysis visit windows for LH, FSH, inhibin B, and total testosterone are provided in Table 3-3. The analysis windows for "on-treatment" and "reversibility" semen parameters are provided in Table 3-4 and Table 3-5, respectively.

Visit Window Rules for "As Randomized" (eg, while on double-blind treatment) Displays

Sperm/semen and hormone data will have visit windows assigned based on study day calculated from first dose date of double-blind study drug.

Visit Window Rules for "As-Treated" Safety Displays

For subjects who took double-blind placebo and switched to open-label filgotinib, the date of first double-blind study drug will be used to calculate study days for visit windows for the placebo period (summarized with placebo treatment group) while the first open-label filgotinib dose date will be used to calculate study days for visit window assignments for the filgotinib period (summarized with filgotinib treatment group). Subjects who took double-blind filgotinib and switched to open-label filgotinib and subjects who remained on double-blind study drug throughout the study will have study days calculated relative to the first dose date of double-blind study drug for "as-treated" visit window assignments.

Visit Window Rules for "Open-Label Filgotinib" Safety Displays

Open-label filgotinib analysis visit windows assigned based on the study day calculated from first dose date of open-label filgotinib study drug.

Visit Window Rules for "Week 26/LTE" Displays

For subjects who took double-blind placebo and switched to open-label filgotinib, the date of the first open-label filgotinib dose date will be used to calculate study days for open-label analysis visit windows. Subjects who switch from double-blind to open-label filgotinib and those who remain on double-blind study drug will have analysis visit windows assigned based on the study day calculated from first dose date of double-blind study drug.

Visit Window Rules for Sperm/Semen Parameters During the Monitoring Phase

Sperm/semen data collected during the Monitoring Phase to assess reversibility of sperm parameters with a $\geq 50\%$ decrease from baseline will have the last dose date of any study drug used to calculate study days for Monitoring Phase visit windows.

Table 3-1. Analysis Visit Windows for Lipid Profile, Serum Immunoglobulin, and Urinalysis

Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Baseline	1	None	1
Week 13	92	2	137
Week 26	183	138	228
Week 39	274	229	319
Week 52	365	320	410
Week 65	456	411	501
Week 78	547	502	592
Week 91	638	593	683
Week 104	729	684	774
Week 117	820	775	865
Week 130	911	866	956
Week 143	1002	957	1047
Week 156	1093	1048	1138
Week 169	1184	1139	1229
Week 182	1275	1230	1320
Week 195	1366	1321	1411
Week 208	1457	1412	1502
Week 221	1548	1503	≥ 1548

Table 3-2. Analysis Visit Windows for Partial MCS, CDAI, Vital Signs, Weight, Hematology, and Chemistry Laboratories

Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Baseline	1	None	1
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	74
Week 13	92	75	116
Week 20	141	117	162
Week 26	183	163	228
Week 39	274	229	319
Week 52	365	320	410
Week 65	456	411	501
Week 78	547	502	592
Week 91	638	593	683
Week 104	729	684	774
Week 117	820	775	865
Week 130	911	866	956
Week 143	1002	957	1047
Week 156	1093	1048	1138
Week 169	1184	1139	1229
Week 182	1275	1230	1320
Week 195	1366	1321	1411
Week 208	1457	1412	1502
Week 221	1548	1503	≥ 1548
FU-4*	Last Dose Date any study drug + 30 days	Last Dose Date any study drug + 8 days	Last Dose Date any study drug + 35 days

Note: Data will be included for "on-treatment" visits up to the last dose date of any study drug + 7 days.
*FU-4 visit for subjects who have stopped all study drugs and had a 30-day safety FU visit ± 5 days. FU-4 visit window is applicable for hematology, chemistry, vitals, and weight. FU-4 visit will be displayed in listings only.

Table 3-3. Analysis Visit Windows for LH, FSH, Inhibin B, and Total Testosterone

Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Baseline	1	None	1
Week 4	29	2	84
Week 13	92	85	112
Week 26	183	170	260
Week 39	274	261	351
Week 52	365	352	442
Week 65	456	443	533
Week 78	547	534	624
Week 91	638	625	715
Week 104	729	716	806
Week 117	820	807	897
Week 130	911	898	988
Week 143	1002	989	1079
Week 156	1093	1080	1170
Week 169	1184	1171	1261
Week 182	1275	1262	1352
Week 195	1366	1353	1443
Week 208	1457	1444	1534
Week 221	1548	1535	≥ 1548
FU-4*	Last Dose Date any study drug + 30 days	Last Dose Date any study drug + 8 days	Last Dose Date any study drug + 35 days

Note: Primary analysis includes data up to first dose date of open-label filgotinib (if non-missing) or to last dose date of double-blind drug + 7 days (for those who do not switch to open-label filgotinib).
Secondary analysis includes all "on-treatment" data up to the last dose date of any study drug + 7 days.
*FU-4 visit for subjects who have stopped all study drugs and had a 30-day safety FU visit ± 5 days will be displayed in listings only.

Table 3-4. Analysis Visit Windows for On-Treatment Semen Parameters

Analysis Visit	Nominal Day (Study Day)	Lower Limit (Study Day)	Upper Limit (Study Day)
Baseline	1	None	1
Week 13	92	85	112
Week 26	183	170	260
Week 39	274	261	351
Week 52	365	352	442
Week 65	456	443	533
Week 78	547	534	624
Week 91	638	625	715
Week 104	729	716	806
Week 117	820	807	897
Week 130	911	898	988
Week 143	1002	989	1079
Week 156	1093	1080	1170
Week 169	1184	1171	1261
Week 182	1275	1262	1352
Week 195	1366	1353	1443
Week 208	1457	1444	1534
Week 221	1548	1535	≥ 1548

Note: Primary analysis includes data up to first dose date of open-label filgotinib (if non-missing) or to last dose date of double-blind drug + 7 days (for those who do not switch to open-label filgotinib).
Secondary analysis includes all "on-treatment" data up to the last dose date of any study drug + 7 days.

Table 3-5. Analysis Visit Windows for Semen Parameters in Monitoring Phase (Calculated from Last Dose Date of Any Study Drug)

Analysis Visit	Nominal Day (FU Day)	Lower Limit (FU Day)	Upper Limit (FU Day)
MP Week 13	91	8	104
MP Week 26	182	105	195
MP Week 39	273	196	286
MP Week 52	364	287	377

MP = Monitoring Phase; FU = Follow-Up.
NOTE: All visit windows for monitoring phase will be calculated from last dose date of any study drug.
FU Day = visit date – last dose date of any study drug.

Investigator Assessment of ECG Visit Assignment

Nominal visits from the clinical database will be used to assign timepoints for the investigator's assessment of electrocardiogram. Baseline will use the nominal visit "Screening"; Week 26 will use nominal visits "Week 26", "Open-Label Week 13", or "Early Termination" if the other visits are not available and the subject is not enrolled in LTE; and LTE Week 195 or "Early Termination" visit will use nominal visits of "LTE Week 195" or "Early Termination" if LTE Week 195 is not available and subject is enrolled in LTE Phase.

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

Depending on the statistical analysis method utilized, single values may be required from 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 value will be the last nonmissing value on or prior to the first dosing date of study drug, unless specified differently.
- If multiple measurements occur on the same day, the last nonmissing value prior to the time of first dosing 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 value.
- For postbaseline values:
  - 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, the last available record on or prior to the date of the first dose of study drug 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 findings).
- For postbaseline visits, if there are multiple records with the same time or no time recorded on the same day, the value with the worst severity within the window will be selected (eg, abnormal will be selected over normal for safety findings).

For semen data, since 2 evaluable semen samples are required at each semen sample collection time point, the average of the measurements will be taken and the study day associated with the earliest valid measurement will be used for assignment of the visit to an analysis visit window. For the sensitivity analysis of sperm concentration at Week 13 that includes subjects with only 1 evaluable measurement at a collection timepoint, the date of the single sample will be used to calculate the study day for assignment to an analysis visit window. If multiple collection time points fall within a visit window, the above rule will be applied.

4. SUBJECT DISPOSITION

4.1. Subject Randomization and Disposition

Summaries described under this section will be based on the All Randomized Analysis Set and randomized treatment groups.

A summary of subject randomization will be provided by randomized treatment group and overall for each country, and investigator within a country. The summary will present the number and percentage of subjects randomized. For each column, the denominator for the percentage calculation will be the total number of subjects randomized for that column.

A similar randomization table will be provided by randomization stratum. The denominator for the percentage of subjects in the stratum will be the total number of randomized 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. A listing of 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 provided.

The randomization schedule used for the study will be provided as an appendix to the CSR.

The flow through the study for an individual subject will be determined by the response of the underlying IBD to assigned treatment and by the subject meeting a pre-specified sperm decrease threshold (eg, a $\geq$ 50% decrease in sperm concentration and/or motility and/or morphology from baseline). Subjects meeting a pre-specified sperm decrease threshold will stop study drug and switch to a standard of care selected by the investigator (in accordance with the protocol) and enter the Monitoring Phase of the study. The mean value at a study visit for each of the sperm parameters (measured from 2 separate evaluable semen collection samples) will be used to determine whether a pre-specified sperm decrease threshold has been met. In the special case where the 2 semen collections at a visit are evaluable, but data for an individual sperm parameter (ie, motility is not evaluable due to technical problems when performing semen analysis) is not evaluable and collection of a third semen sample at the visit is not possible, the 1 evaluable value will be used to determine whether a pre-specified sperm decrease threshold was met for that sperm parameter, while the mean of the 2 samples will be used for sperm parameters with 2 evaluable collections.

A summary of subject disposition will be provided by randomized treatment group and overall. This summary will present the number of subjects screened (rescreens counted once), the number of subjects randomized, and the number of subjects in the safety analysis set. In addition, the number and percentage of subjects in each category below will be summarized:

Double-Blind Study Drug Completion Status Through Week 13

- Completed blinded study drug through Week 13
  - Switched to Monitoring Phase after Week 13 (subject met pre-specified sperm decrease threshold[s])
  - Continued double-blind treatment after Week 13 (responder, subject did not meet pre-specified sperm decrease threshold[s])
  - Switched to Open-Label Filgotinib Phase after Week 13 (non-responder, subject did not meet pre-specified sperm decrease threshold[s])
- Did not complete blinded study drug through Week 13 with reasons for premature discontinuation of study drug

Double-Blind Study Drug Completion Status (> Week 13 Through Week 26)

- Completed blinded study drug through Week 26
  - Switched to Monitoring Phase after Week 26 (subject met pre-specified sperm decrease threshold[s])
  - Continued blinded study drug (responder and no disease worsening, subject did not meet pre-specified sperm decrease threshold[s])
  - Discontinued study drug at Week 26 (non-responder, subject did not meet pre-specified sperm decrease threshold[s])
- Did not complete blinded study drug through Week 26 with reasons for premature discontinuation of study drug

Open-Label Filgotinib Phase Study Drug Completion Status

- Completed open-label study drug through Open-Label Week 13
  - Switched to Monitoring Phase after Open-Label Week 13 (subject met pre-specified sperm decrease threshold[s])
  - Continued open-label filgotinib after Open-Label Week 13 (responder and no disease worsening, subject did not meet pre-specified sperm decrease threshold[s])
  - Discontinued Open-Label Filgotinib Phase at Open-Label Week 13 (non-responder, subject did not meet pre-specified sperm decrease threshold[s])
- Did not complete open-label study drug through Open-Label Week 13 with reasons for premature discontinuation of study drug

LTE Double-Blind Study Drug Completion Status

- Completed LTE Double-Blind Phase of study
- Ongoing in Long Term Extension (LTE) Double-Blind Phase of study (for interim analysis only, includes subjects who completed 26 weeks of double-blind study drug and are ongoing on double-blind drug in the extension phase at the time of interim analysis)
- Did not complete LTE Double-Blind Phase of study and reason for premature discontinuation

LTE Open-Label Filgotinib Study Drug Completion Status

- Completed LTE on open-label filgotinib
- Ongoing in LTE open-label filgotinib phase of study (for interim analysis only, includes subjects who completed Open Label Week 13 and are ongoing on open-label filgotinib in the extension phase at the time of interim analysis)
- Did not complete LTE open-label phase of study and reason for premature discontinuation

Monitoring Phase Completion Status

- Completed Monitoring Phase of study
  - Met reversibility criteria for all sperm parameters meeting pre-specified sperm decrease threshold[s]
  - Completed 52-weeks of off-treatment follow-up
- Continuing Monitoring Phase of study (for interim analysis only), represents those who remain under observation in the Monitoring Phase at the time of data cutoff date)
- Did not complete Monitoring Phase and reason for premature discontinuation

Note for programming: Reasons for Early Termination from Monitoring Phase and Completion of Monitoring Phase will be pulled from the Study Completion eCRF.

Study Completion Status

- Completed study
- Ongoing in study (for interim analysis only, includes subjects currently participating in the study at the time of an interim analysis)
- Did not complete study with reasons for premature discontinuation of study

For the status of study drug, study completion and reasons for premature discontinuation of study drug or study, 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 who entered into the study phase for the column being summarized.

The number of subjects screened for the study that did not meet eligibility criteria will be summarized overall and by inclusion/exclusion criteria not met (multiple reasons may be marked for an individual screen). Screen failure reasons for tuberculosis and for failure to meet IBD criteria (across different versions of the protocol) will be counted once for the category. Subjects who were rescreened will be summarized based on their last chronologic screening attempt. Screens that did not explicitly fail ≥ 1 eligibility criteria but were not randomized will be summarized by the reason the subject was not randomized.

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)

Screen Failure for Sperm/Semen Parameters

The number of subjects who had ≥ 1 sperm/semen abnormality for inclusion criteria (IC) #8 based on the mean of 2 evaluable semen samples at the screening visit will be displayed. Subjects who were rescreened will be included in analysis based on their last chronologic screening attempt. Inclusion criteria #8 was defined as: semen volume ≥ 1.5 mL; total sperm/ejaculate ≥ 39 M/ejaculate; sperm concentration ≥ 15 M/mL; sperm total motility ≥ 40%; and normal sperm morphology ≥ 30% [WHO 1992 criteria] for subjects randomized under protocol Amendment 3 or normal sperm morphology ≥ 4% [WHO 2010 criteria] for subjects randomized under the original protocol or Protocol Amendment 1. Sperm morphology was assessed for inclusion/exclusion criteria based on the protocol version the subject enrolled under. The number and percentage of subjects not meeting each individual sperm/semen parameter (denominator for percentage calculation is number of subjects who had ≥ 1 sperm/semen abnormality) will be displayed. Multiple sperm/semen parameters may not qualify for a given screening attempt.

A listing of screened subjects with 2 evaluable semen samples at the screening visit who screen failed for IC #8 will be provided with sperm/semen parameters not meeting criteria flagged in the listing.

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.

4.2.1. Duration of Exposure to Study Drug

Total duration of exposure to double-blind study drug (in weeks) will be defined as: ([last dosing date of double-blind study drug minus first dosing date of double-blind study drug plus 1] divided by 7), regardless of any temporary interruptions in study drug administration. Results for all calculations of duration will be displayed to 1 decimal place (eg, 4.5 weeks).

Total duration of exposure to open-label filgotinib study drug (in weeks) will be defined as: ([last dosing date of open-label filgotinib study drug minus first dosing date of open-label filgotinib study drug plus 1] divided by 7), regardless of any temporary interruptions in study drug administration.

Total duration of exposure to any filgotinib study drug (in weeks) will be defined as: ([last dosing date of any filgotinib study drug minus first dosing date of any filgotinib study drug plus 1] divided by 7), regardless of any temporary interruptions in study drug administration.

For subjects who took double-blind placebo and then switched to open-label filgotinib, the duration of exposure will be calculated from the first dose of double-blind placebo to the last dose of double-blind placebo for double-blind study drug; and from the first dose of open-label filgotinib to last dose of open-label filgotinib for open-label filgotinib and for any filgotinib.

The any filgotinib exposure (in weeks) for a subject who switched from double-blind filgotinib to open-label filgotinib will be calculated as: ([last dose date of open-label filgotinib minus first dose date of double-blind filgotinib + 1] divided by 7).

For subjects with a partial last dosing date (ie, month and year of last dose are known), the latest of the dispensing dates of study drug bottles, study drug start dates and end dates, and the imputed last dosing date (day imputed as 15) will be used as the final imputed last dosing date. If the subject died and the death date is complete (ie, not partial date) and before the imputed last dosing date, the complete death date will be used as the imputed last dosing date.

If only year is recorded (ie, month and day of last dose are missing), the latest of the dispensing month of study drug bottles, study drug start month, and study drug end month will be used to impute the unknown last dose month. If the subject died and the death date has month and year available and before the imputed last dose month, then the month of death will be used instead. With the month imputed, the aforementioned method will be used to impute the last dose date.

If subjects are continuing on study drug at the time of an interim analysis, the earliest of the date of death or data cutoff date for the interim analysis will be used to impute the last dosing date for the calculation of duration of exposure to study drug.

The total duration of exposure to double-blind study drug, open-label filgotinib, and any filgotinib study drug will be summarized using descriptive statistics and using the number (ie, cumulative counts) and percentage of subjects exposed through the following time periods: Baseline (Day 1), Week 2 (Day 14), Week 4 (Day 28), Week 8 (Day 56), Week 13 (Day 91), Week 20 (Day 140), Week 26 (182), Week 39 (273), and up to Week 221 (Day 1548) at 13-week intervals (as appropriate).

Summaries will be provided by double-blind treatment group for double-blind drug; and by treatment group and overall for "open-label filgotinib" and "any filgotinib" for the Safety Analysis Set. No formal statistical testing is planned.

4.2.2. Adherence to Study Drug

For the final analysis, the on-treatment adherence will be calculated for double-blind filgotinib 200 mg/PTM (tablets), throughout the double-blind phase for the Semen Analysis Set.

The total number of double-blind 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:

Total Number of Tablets Administered

$$= \left(\sum \text{No. of Tablets Dispensed}\right) - \left(\sum \text{No. of Tablets Returned}\right)$$

Bottles for double-blind drug dispensed throughout the study will be included in the calculations.

If a bottle is dispensed and the bottle is returned empty, then the number of tablets returned will be entered as zero. If a bottle is dispensed but not returned (missing), the number of tablets administered will be counted as zero by assuming that a subject did not take study drug as prescribed during the period for which the bottle was dispensed.

4.2.2.1. On-Treatment Adherence

The level of on-treatment adherence to the study drug (during double-blind treatment) 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 (while on double-blind treatment) will be expressed as a percentage using the following formula:

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

Note: If calculated adherence is greater than 100%, the result will be set to 100%.

Study drug expected to be administered for filgotinib 200 mg/PTM (tablets) = 1× total duration of exposure to double-blind study drug (days) up to last visit (eg, last double-blind dose date – double-blind first dose date + 1).

Descriptive statistics for the level of on-treatment adherence with the number and percentage of subjects belonging to adherence categories (eg, < 80%, ≥ 80% to < 90%, ≥ 90%) will be provided by treatment group during the double-blind phase of the study for the Semen Analysis Set for the final analysis. For the interim analyses, a listing of subjects who interrupted study drug as recorded on the study drug administration eCRF will be listed, since all double-blind bottles for the study will not have been returned.

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).

A listing of subjects who had interruptions in study drug dosing will be produced for both the interim and final analyses.

4.3. Protocol Deviations

A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion but were enrolled into the study for the All Randomized Analysis Set. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that the subject 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 ≥ 1 important protocol deviation (IPD) will be summarized by treatment group. In addition, important protocol deviations will be summarized by treatment group and IPD category for the All Randomized Analysis Set.

A by-subject listing will be provided for those subjects with an IPD. In addition, 2 separate by-subject listings will be provided for subjects with: 1) any important protocol deviation related to COVID-19, and 2) any non-important protocol deviation related to COVID-19.

5. BASELINE CHARACTERISTICS

5.1. Demographics

Subject demographic variables (ie, age, sex, race, and ethnicity) will be summarized for the double-blind phase of the study by treatment group and overall for the ITT/Safety Analysis Sets and the Semen Analysis Set using descriptive statistics for age, and using number and percentage of subjects for sex, race and ethnicity. Separate displays will present data for the "as-treated analysis" by treatment for the Safety Analysis Set and by treatment sequence for the Week 26/ LTE Semen Analysis Set. Subjects who were randomized to placebo and switched to open-label filgotinib will be counted in both the placebo and the filgotinib group for "as-treated" displays.

Age is calculated in years on the date of first study drug administration. If a subject does not receive study drug after randomization, the subject's age will be calculated from the Day 1/baseline assessment. A by-subject demographic listing, including the informed consent date, will be provided by subject ID number in ascending order.

5.2. Baseline Characteristics

Other baseline characteristics will include:

- Body weight (in kg)
- Height (in cm)
- Body mass index (BMI; in kg/m^2)
- Type of IBD disorder (UC versus CD)
- Duration of UC/CD (in years)
- MCS score (categorical for each score) for UC subjects only
- Endoscopic subscore of MCS (2 or 3) for UC subjects only
- CDAI score for CD subjects only
- Smoking status (former, current, never)
- Alcohol Use ─ frequency of drinking in past 12 months (no alcohol in my life, no alcohol in past 12 months, 1 or 2 times in past year, 3-11 times in the past year, 1-3 times a month, < 4 times a week, ≥ 4 times a week)
- Alcoholic drinks per day consumed on a drinking day ─ past 12 months (1 drink, 2 drinks, 3-4 drinks, ≥ 5 drinks)
- Number of children fathered

- Concurrent MTX use status (yes, no)
- Sperm concentration (15 to 25 M/mL, > 25 to 50 M/mL, > 50 M/mL)
- Sperm concentration (M/mL)
- Sperm total motility (%)
- Total sperm count (M/ejaculate)
- Ejaculate volume (mL)
- Sperm morphology (% normal) using WHO 1992 criteria
- Luteinizing hormone (mIU/mL)
- Follicle-stimulating hormone (mIU/mL)
- Inhibin B (pg/mL)
- Total testosterone (ng/dL)

Duration of UC/CD (in years) calculated as:

([first dose date] – [date of initial diagnosis of UC/CD] + 1 day) / 365.25

If the date of initial diagnosis is incomplete, then the following rules will be applied:

- missing day: use the first of the month
- missing month: use January

Baseline characteristics will be summarized by double-blind treatment group and overall using descriptive statistics for continuous variables and using number and percentage of subjects for categorical variables for the Safety/ITT Semen Analysis Sets and the Semen Analysis Set. Separate displays will be presented by treatment for the As-Treated Safety Analysis Set and by treatment sequence for the Week 26/LTE Semen Analysis Set.

A by-subject listing of baseline characteristics will be provided by subject ID number in ascending order.

5.3. Medical History

Disease-specific and general medical history data will be collected at screening. Medical history data will not be coded.

A by-subject listing of medical history will be provided by subject ID number in ascending order and abnormalities in chronological order.

6. SEMEN AND HORMONE ANALYSES

6.1. General Considerations

The definitions for the primary and secondary endpoints are listed in Table 6-1.

Table 6-1. Definitions for the Primary and Secondary Endpoints

Type	Event Endpoint
Primary	Proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 13
Secondary	Proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 26
Secondary	Change from baseline in sperm total motility at Weeks 13 and 26
Secondary	Change from baseline in total sperm count at Weeks 13 and 26
Secondary	Change from baseline in sperm concentration at Weeks 13 and 26
Secondary	Change from baseline in ejaculate volume at Weeks 13 and 26
Secondary	Change from baseline in percent normal sperm morphology at Weeks 13 and 26

Note: Value at a visit is the mean of 2 evaluable semen sample collections at a given collection timepoint for primary analysis.

When calculating the percentage change from baseline in sperm concentration, no rounding will be performed in the analysis. For the purposes of tables and listings, the percentage will be rounded to 1 decimal place.

Eligible subjects were required at screening to have the mean of 2 evaluable semen collections for each of the 5 sperm/semen parameters greater than or equal to the value associated with the 5th percentile of the WHO reference set per Inclusion Criteria #8. Subjects randomized under the original protocol/Protocol Amendment 1 were enrolled based on WHO 2010 criteria for sperm morphology. Upon consultation with the US FDA, morphology grading was changed to that specified by WHO 1992 in Protocol Amendment 2. Morphology slides collected using WHO 2010 criteria were re-read using WHO 1992 criteria for purposes of summarizing morphology data over time. Semen parameters other than sperm morphology used WHO 2010 criteria {World Health Organization (WHO) 2010}. Semen parameter eligibility criteria are provided in Table 6-2.

Table 6-2. Semen Parameter Eligibility Criteria

Parameter	Reference Range
Sperm concentration	≥ 15 million per milliliter (mL)
Sperm total motility	≥ 40%
Total sperm per ejaculate	≥ 39 million
Ejaculate volume	≥ 1.5 mL
Sperm morphology	≥ 30% normal*

Listed semen parameters use 2010 WHO Reference values {Cooper 2010, World Health Organization (WHO) 2010} for human semen characteristics except *(morphology) which utilizes the 1992 WHO sperm morphology criterion {World Health Organization (WHO) 1992}.
NOTE: Prior to amendment 2, subjects were required to have sperm morphology ≥ 4% normal at screening using the strict "Tygerberg" method per inclusion criteria #8. Morphology for semen samples collected prior to consent/re-consent for Protocol Amendment 3 were re-read using WHO 1992 methodology. WHO 1992 will be used for all postbaseline morphology analysis.

6.2. Primary Analysis of Primary and Secondary Endpoints

The primary analysis will be an observed cases analysis at Week 13 while subjects are on double-blind study drug as described in FDA's guidance for industry on testicular toxicity studies {U. S. Department of Health and Human Services (DHHS) 2018}.

The main estimators of a While on Treatment Strategy estimand will be utilized for the primary analysis of primary and secondary endpoints at Week 13:

BINARY ENDPOINT	WHILE ON-TREATMENT STRATEGY
Population	Semen Analysis Set (subjects with 2 evaluable semen samples collected at baseline and at the Week 13 analysis visit) as defined in Section 3.1.2 of this SAP.
Patient Level Outcome to be Measured	Binary outcome for a subject meeting a ≥ 50% decrease from baseline in sperm concentration at Week 13. Outcome is determined based on the mean of 2 evaluable semen samples at a visit.
Measure of intervention effect and handling of intercurrent events	Measure of treatment effect assuming that the treatment effect prior to the intercurrent event of treatment early termination is of interest. Data included to last dose date double-blind drug + 7 days will be considered "while on-treatment" for those who permanently discontinue double-blind study drug.
Population level summary measure	Difference in a binary outcome, comparing those assigned to filgotinib versus those assigned to placebo.
Estimators	Main estimator: A stratified Mantel-Haenszel test (stratification factor: sperm concentration [15 - 50 and > 50 M/mL]) will be used to estimate the difference in proportions (filgotinib – placebo) and corresponding 95% CI {Koch 1989}. Calculations are provided in Appendix 2. Note: When performing the stratified Mantel-Haenszel test, the 15 - 25 and > 25 - 50 M/mL strata for sperm concentration will be combined (eg, 15 - 50 M/mL) due to the small number of subjects in the lowest stratum; disease type will not be included as a stratification factor due to the small number of subjects with CD; and no subjects in this study received MTX.

The number and proportion of subjects with a ≥ 50% decrease in sperm concentration at Week 26 (observed cases) will be presented by double-blind treatment group. The number of subjects who have a ≥ 50% decrease in sperm concentration at any time up to Week 26 (to upper bound of analysis visit window) will be displayed to ensure that ≥ 50% decreases from baseline in sperm concentration that fell outside of the analysis visit windows are accounted for.

CONTINUOUS ENDPOINTS	WHILE ON TREATMENT STRATEGY
Population	Semen Analysis Set (subjects with 2 evaluable semen samples collected at baseline and at the Week 13 analysis visit) as defined in Section 3.1.2 of this SAP.
Patient Level Outcome to be Measured	Change and Percentage Change from Baseline (continuous outcome) at Week 13. Outcome is determined based on the mean of 2 evaluable semen samples at a visit, and analysis visit windows are assigned based on first of 2 evaluable semen samples.
Measure of intervention effect and handling of intercurrent events	Measure of treatment effect assuming that the treatment effect prior to the intercurrent event of treatment early termination is of interest. Data included to last dose date double-blind drug + 7 days will be considered "while on-treatment" for those who permanently discontinue double-blind study drug.
Population level summary measure	Median Difference between treatment groups, comparing those assigned to filgotinib versus those assigned to placebo in the change and percentage change from baseline at Week 13 for a continuous outcome.
Estimators	Main estimator: Quantile regression models will be used to estimate the median difference (filgotinib – placebo) in change and percentage change from baseline at Week 13, 95% CI for the sperm/semen parameter of interest. The CI=RANK option will be specified in PROC QUANTREG in SAS. For sperm concentration (M/mL), the model will include: baseline sperm concentration (continuous). For sperm total motility (%), percent normal morphology, total sperm count (M/ejaculate) and ejaculation volume (mL), the model will include: baseline value for parameter (continuous), and sperm concentration (15 – 50 M/mL, > 50 M/mL).

Eight-point summary statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) by double-blind treatment group will be presented for baseline value, absolute value at the visit, and change and percentage change from baseline for each sperm parameter at Week 13. A 95% distribution-free CI on the median change and percentage change from baseline for each semen parameter will be constructed by treatment group at Week 13 using the CIPCTLDF option in PROC UNIVARIATE in SAS.

Based on FDA's guidance for industry on testicular toxicity studies {U. S. Department of Health and Human Services (DHHS) 2018} and communication with the FDA, a cumulative distribution plot for percentage change from baseline in sperm concentration at Week 13 will be constructed by double-blind treatment group. Cumulative distribution plots for percentage change from baseline at double-blind Week 13 for total sperm motility (%) and sperm morphology (% normal) will also be presented by double-blind treatment group.

A waterfall plot will be constructed to present the distribution of percentage change from baseline in sperm concentration at Week 13. Each line on the x-axis of the waterfall plot will represent the percentage change from baseline in sperm concentration for an individual subject with their treatment assignment displayed based on the color/symbol displayed for the line.

6.3. Subgroup Analysis for Sperm Concentration Parameter at Week 13

The following subgroup analyses will be performed by treatment group for the endpoints proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 13, and change and percentage change from baseline in sperm concentration at Week 13:

- Type of IBD disorder (UC, CD)
- Sperm concentration (mean of 2 evaluable samples) measured at Screening visit: (15 - 50 M/mL, > 50 M/mL)

A subgroup analysis of concurrent MTX use (yes, no) was not performed since no subjects took concurrent MTX.

6.4. Supportive Analysis for Primary and Secondary Endpoints at Week 13

In addition to the primary analysis, the main estimators of a hypothetical strategy estimand will provide support to the primary analysis of primary and secondary endpoints at the Week 13 timepoint. The estimands for binary and continuous endpoints are described below.

Estimand for Binary Endpoint

BINARY ENDPOINT	HYPOTHETICAL STRATEGY
Population	ITT Semen Analysis Set (subjects with 2 evaluable semen samples collected at baseline) as defined in Section 3.1.2 of this SAP.
Patient Level Outcome to be Measured	Binary outcome for a subject meeting a ≥ 50% decrease from baseline in sperm concentration at Week 13. Outcome is determined based on the mean of 2 evaluable semen samples at a visit. A subject with only 1 of the 2 evaluable semen samples at Week 13 will be included with their 1 sample [Subject ID 10556-35035]. If a measurement for a specific semen parameter is deemed ineligible at Week 13, the 1 evaluable result will be used.
Measure of intervention effect and handling of intercurrent events	Measure of treatment effect assuming that the intercurrent event of early termination prior to Week 13 does not occur. Subjects who had missing data at Week 13 will be assumed to be missing at random (MAR) and data will be imputed using PROC MI in SAS and a logistic regression model with treatment group, and baseline sperm concentration category (15 - 50 M/ml, > 50 M/ml) as independent variables in the model.
Population level summary measure	Difference in a binary outcome, comparing those assigned to filgotinib versus those assigned to placebo.
Estimators	Main estimator: Imputed datasets will be analyzed using the stratified Mantel-Haenszel method (stratification factors: sperm concentration [15 – 50 M/mL, > 50 M/mL]). Results will be combined using Rubin's rule {Rubin 1987} to estimate the difference in proportions (filgotinib – placebo) and corresponding 95% CI.

Estimand for Continuous Endpoints

CONTINUOUS ENDPOINTS	HYPOTHETICAL STRATEGY
Population	ITT Semen Analysis Set (subjects with 2 evaluable semen samples collected at baseline) as defined in Section 3.1.2 of this SAP.
Patient Level Outcome to be Measured	Change and Percentage Change from Baseline (continuous outcome) at Week 13. Outcome is determined based on the mean of 2 evaluable semen samples at a visit. A subject with only 1 of the 2 evaluable semen samples at Week 13 will be included with their 1 sample [Subject ID 10556-35035]. If a measurement for a specific semen parameter is deemed ineligible at Week 13, the 1 evaluable result will be used.
Measure of intervention effect and handling of intercurrent events	Measure of treatment effect assuming that intercurrent event of early termination prior to Week 13 does not occur. Subjects who had missing data at Week 13 due to early termination will be assumed to be MAR and data will be imputed by conditional quantile estimation using the quantile regression model specified for the primary analysis using the methods of Bottai and Zhen.
Population level summary measure	Median Difference between treatment groups, comparing those assigned to filgotinib versus those assigned to placebo in the change and percentage change from baseline at Week 13 for a continuous outcome.
Estimators	Main estimator: Imputed datasets will be analyzed using quantile regression models used for primary analysis to obtain estimates and standard errors (standard errors obtained through bootstrap methods). Results of the estimates and standard errors will be combined using Rubin's rule to estimate the median difference (filgotinib – placebo) in change or percentage change from baseline, 95% CI for the sperm/semen parameter of interest.

6.5. Other Displays of Sperm/Semen Data

6.5.1. Treatment-Sequence Semen Displays (Week 26/OL Week 13 and Extension Phase)

The proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration will be presented by treatment sequence and analysis visit. An exact 95% CI for the proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration at Week 26/OL Week 13 within treatment sequence will be presented based on the Clopper-Pearson method.

Descriptive statistics will be presented at Baseline, Week 26/OL Week 13 and subsequent visits by actual treatment sequence to explore the confounding effects of IBD response to treatment. A 95% distribution-free CI on the median change and percentage change from baseline in sperm/semen parameters at Week 26/OL Week 13 will be constructed by treatment sequence using the CIPCTLDF option in PROC UNIVARIATE in SAS. Study days will be used for assignment of analysis visit windows for subjects who stayed on randomized drug throughout study and for subjects who started on double-blind filgotinib and switched to open-label filgotinib. For subjects who switched from placebo to open-label filgotinib, days on open-label filgotinib will be used to assign analysis visit windows. Week 26 for the first 3 treatment

sequences will be displayed along-side OL Week 13 for placebo to OL filgotinib group; Week 39 alongside OL Week 26 for placebo to OL filgotinib group, etc. to align with the originally planned study weeks, but using the at/after Protocol Amendment 3 visit schedule to ensure that subjects who switched from placebo to open-label filgotinib had at least 85 days on filgotinib prior to their first semen collection while on filgotinib treatment (eg, a full sperm cycle).

A cumulative distribution plot for percentage change from baseline in sperm concentration at the Week 26/OL Week 13 visit will be constructed by treatment sequence for subjects who are still on study drug at Week 26/OL Week 13.

The number of days on filgotinib at the time of the first of 2 semen assessments at the Week 26/Open-Label Week 13 analysis visit will be summarized by treatment sequence for subjects with evaluable samples in the analysis visit window.

A listing of subjects in the placebo to open-label filgotinib treatment sequence who had nominal visits for Week 26 or Open-Label Week 13 that were outside of the analysis visit window for OL Week 13 will be provided—primarily to assess the number of subjects who were on original/ Protocol Amendment 1 visit schedule and did not have sufficient exposure (minimum of 85 days) on open-label filgotinib at the time of their first collection to fall within the Open-Label Week 13 analysis visit window

6.5.2. Pre-Specified Sperm Decrease Threshold Criteria and Confirmed Semen Abnormalities

Subjects who met a pre-specified sperm decrease threshold criteria (a $\geq$ 50% decrease in sperm concentration and/or motility and/or morphology) were to stop study drug and begin a standard of care regimen selected by the investigator in accordance with the protocol and be followed off study drug to evaluate for reversibility criteria.

A “confirmed semen abnormality” as defined by Jarvi and colleagues {Jarvi 2008} reflects the presence of abnormal sperm parameters potentially associated with decreased male fertility; the criterion is met when one or more of the following are observed: a sperm concentration $<$ 5 M/mL, or sperm motility $<$ 20%, or a percent normal sperm morphology of $<$ 10%.

The number and proportion of subjects meeting any of the pre-specified sperm decrease thresholds (from the baseline value [mean of 2 evaluable semen collections at screening]) or any of the components of a confirmed semen abnormality (mean of 2 evaluable samples at a study visit) at the double-blind Week 13 (primary endpoint) visit will be displayed by double-blind treatment group:

Sperm Decrease Criteria

- ≥ 50% decrease in sperm concentration
- ≥ 50% decrease in sperm total motility
- ≥ 50% decrease in sperm morphology [% normal] (WHO 1992 criteria)
- ≥ 50% decrease in sperm concentration and (≥ 50% decrease in total motility and/or morphology [% normal] at the same visit)

Confirmed Semen Abnormality Criteria (as defined by Jarvi et al)

- Sperm concentration < 5 M/mL
- Sperm total motility < 20%
- Sperm morphology (% normal) < 10%
- Confirmed Semen Abnormality (Sperm concentration < 5 M/mL or sperm total motility < 20% or sperm morphology [% normal] < 10% at the same visit)

The number and proportion of subjects meeting any of the pre-specified sperm decrease thresholds or components of a confirmed semen abnormality (as described above) at the Week 26/OL Week 13 analysis visit and at/after the Week 39/OL Week 26 visit will be displayed by treatment sequence. Treatment sequences to be presented are:

- double-blind (DB) filgotinib
- DB filgotinib to open-label (OL) filgotinib
- DB placebo
- DB placebo to OL filgotinib

A listing of subjects who met any of the components of a confirmed semen abnormality will be provided.

6.6. Exploratory Endpoints

6.6.1. Definition of the Exploratory Endpoints

The exploratory endpoints are:

- Percent of reversibility among subjects who experience a ≥ 50% decrease in sperm concentration, and/or motility, and/or morphology
- Change from baseline in hormones, including LH, FSH, inhibin B, and total testosterone at Weeks 13 and 26

6.6.2. Reversibility Analysis

Reversibility will be summarized separately for each sperm parameter that met a pre-specified sperm parameter decrease. The number (percent) who met reversibility criteria and the cumulative number (percent) who met reversibility criteria at each Monitoring Phase visit will be presented. Reversibility for subjects meeting pre-specified sperm parameter decrease threshold[s] at Week 13 will be displayed by double-blind treatment group; at Week 26/OL Week 13 and at/after Week 39/OL Week 26 will be displayed by actual treatment sequence. For the Week 26 interim analysis, an additional reversibility table for "completers" (ie, subjects who have completed the full 52 weeks of off-study drug follow-up, met reversibility criteria, or prematurely discontinued from study while being followed in the monitoring phase) will be produced if at least 1 subject has an evaluable monitoring phase visit and is ongoing in the monitoring phase at the time of data cutoff.

No formal testing or CI is planned. A by-subject data listing of semen data (with measurement where reversibility criteria was met flagged) will be provided for subjects who entered the Monitoring Phase.

Table 6-3 shows an example for hypothetical proportions of subjects (at final analysis) with reversed sperm concentration for a ≥ 50% decrease from baseline at/prior to Week 13 by monitoring visit number and last treatment received.

Table 6-3. Hypothetical Proportion of Subjects Who Met Reversibility Criteria by Monitoring Visit Number and Treatment Group (Subjects with a ≥ 50% Decrease in Sperm Concentration at Week 13)

Treatment before Monitoring	Monitoring Visit 1	Monitoring Visit 2	Monitoring Visit 3	Monitoring Visit 4	Reversibility in Monitoring Phase
Filgotinib	3 / 5	1 / 2	1 / 1	0 / 0	5 / 5
Cumulative Rate	3 / 5 = 60.0%	4 / 5 = 80.0%	5 / 5 = 100.0%	5 / 5 = 100.0%	5 / 5 = 100.0%
Placebo	3 / 6	1 / 3	1 / 2	0 / 1	5 / 6
Cumulative Rate	3 / 6 = 50%	4 / 6 = 68%	5 / 6 = 83.3%	5 / 6 = 83.3%	5 / 6 = 83.3%

6.6.3. Hormones

Similar analysis methods used to summarize continuous secondary semen parameters will be used to analyze hormones LH, FSH, inhibin B, and total testosterone. Baseline will be the last value prior to administration of any study drug.

For all continuous hormone parameters, 8-point summary statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) by double-blind treatment group will be provided for the value at the visit, change from baseline, and percentage change from baseline in hormone parameter at Weeks 4 and 13 (change and percentage change values will be calculated from double-blind baseline). A 95% distribution-free CI on the median change and percentage change from baseline in all hormone parameters will be constructed by treatment at Week 4 and Week 13 using the

CIPCTLDF option in PROC UNIVARIATE in SAS. Quantile regression will be used to estimate the difference (filgotinib minus placebo) in median change and percentage change from baseline in hormones at double-blind Weeks 4 and 13 and to construct a 95% CI. The model will adjust for the baseline value of the hormone parameter as a continuous variable, and sperm concentration (15 - 50 M/mL, and > 50 M/mL). Since no subjects took concurrent MTX, and very few CD subjects were enrolled, these stratification factors will not be included in the model. The CI=RANK option will be specified in PROC QUANTREG in SAS.

Shift tables from baseline to Week 4 and Week 13 will be summarized by double-blind treatment group for the Semen Analysis Set for LH, FSH, inhibin B, and total testosterone. The shift tables include shift analysis for hormone parameters below, within, and above the central laboratory defined normal range for each specific hormone at baseline to below, within and above the normal range at Week 4 and Week 13.

Treatment-Sequence Hormone Displays for Week 26/OL Week 13 and Extension Phase

Descriptive statistics will be presented at Baseline, Week 26/OL Week 13, and subsequent visits by actual treatment sequence to explore the confounding effects of response to treatment. Study days will be used for assignment of analysis visit windows for subjects who stayed on randomized drug throughout study and for subjects who started on double-blind filgotinib and switched to open-label filgotinib. For subjects who switched from placebo to open-label filgotinib, days on open-label filgotinib will be used to assign analysis visit windows at/after Open-Label Week 13. Week 26 for the first 3 treatment sequences will be displayed along-side OL Week 13 for placebo to OL filgotinib group; Week 39 alongside OL Week 26 for placebo to OL filgotinib group, etc. to align with the originally planned study weeks; but using the at/after Protocol Amendment 3 visit schedule to ensure that subjects were on open-label filgotinib for an entire sperm cycle for the placebo to open-label filgotinib group.

A summary of the 4 hormone parameters and change and percentage change from baseline by visit will be presented for subjects with evaluable semen data at both baseline and ≥ 1 of the analysis visits at/after Week 26/OL Week 13. Shift tables from baseline to Week 26/OL Week 13 for LH, FSH, inhibin B, and total testosterone will be summarized by treatment sequence for the Week 26/Extension Phase Semen Analysis Set.

A by-subject listing will be provided for hormones. A separate listing of the hormone data for subjects who entered the Monitoring Phase of the study will be provided.

6.7. Changes to Protocol-Specified Analyses of Semen and Hormone Data

The protocol specified an analysis of covariance (ANCOVA) model adjusting for baseline value and stratification factors to estimate treatment differences (95% CI) for the continuous endpoints at timepoints where analysis was to be conducted for semen parameters and hormone data. Quantile regression was used in place of an ANCOVA model as medians were thought to be a more appropriate estimate than means for this data which is potentially skewed with outliers.

Since no subjects took concurrent MTX in this study, the subgroup analysis by MTX (yes, no) was not performed. MTX use (yes, no) and disease type were not included as stratification factors for the stratified Mantel-Haenszel test for the primary endpoint or as a dependent variable in quantile regression models for sperm/semen parameters (secondary endpoints) or hormones (exploratory endpoints) since no subjects took MTX in this study, and only a few CD subjects were enrolled. The lowest 2 levels (15 - 25 M/mL and > 25 - 50 M/mL) for sperm concentration strata were combined in these analyses due to the small number of subjects in the lowest sperm concentration stratum.

7. ASSESSMENT OF RESPONDER/NON-RESPONDER AND DISEASE WORSENING

Composite scores including pMCS for UC and CDAI for CD will be listed using all the observed data based on the All Randomized Analysis Set.

7.1.1. Calculations of MCS and pMCS with the Subscores

The MCS is an instrument designed to measure disease activity in ulcerative colitis (Table 7-1). 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). Individual components of the MCS are described in Table 7-1 and 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; pMCS ranges from 0 to 9.

Table 7-1. 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 (per subject recall)	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

A Responder for UC is defined as a subject who has a reduction of ≥ 2 in pMCS compared to baseline at the specified assessment timepoint. A Non-responder will be defined as a subject who does not fulfill the definition of responder at the specified assessment timepoint.

Disease Worsening for UC is defined by a pMCS after Week 13 consisting of:

- an increase of ≥ 3 points to at least 5 points from the Week 13 value on 2 consecutive visits, or
- an increase to 9 points from the Week 13 value on 2 consecutive visits, if the Week 13 value was > 6 points

7.1.2. Calculation of CDAI Scores

The CDAI system is a composite index of 8 disease activity variables with scores ranging from 0 to over 600 based upon a composite of symptoms (eg, abdominal pain), signs (the presence of abdominal mass and weight), laboratory values (eg, hematocrit), and physician assessment amongst others. The CDAI has 3 patient reported outcome components: liquid or very soft stool frequency, abdominal pain and general wellbeing.

The total CDAI score for a specific visit is based on a weighted sum (rounded to the nearest integer) of all 8 variables as specified in Table 7-2. The CDAI total score will be set to 0 in the case the calculation leads to a CDAI total score of < 0. Each component will be selected within the analysis window as defined in Section 3.7.3.

Table 7-2. Calculation of Total CDAI Score

Variable no.	Variable	Variable description	Multiplier
1	Liquid or very soft stool (per subject recall)	Daily stool count is summed for 7 days	2
2	Abdominal pain (per subject recall)	Sum of 7 days of daily ratings as 0 = none, 1 = mild, 2 = moderate, 3 = severe	5
3	General wellbeing (per subject recall)	Sum of 7 days of daily ratings as 0 = generally well, 1 = slightly below par, 2 = poor, 3 = very poor, 4 = terrible	7
4	Complications	Number of listed complications: - Arthritis or arthralgia - Erythema nodosum, pyoderma gangrenosum or apthous stomatitis - Iritis or uveitis - Anal fissures or fistulae or abscess - Other fistula - Fever over 37.8 C [100 F] during past week	20 each
5	Use of anti-diarrheal	Use of diphenoxylate or loperamide or other opiate for diarrhea 0 = No, 1 = Yes	30
6	Abdominal mass	0 = none, 2 = questionable, 5 = definite	10
7	Hematocrit*	Males: 47 – Hct [%] Females: 42 – Hct [%] *Result must be greater than or equal to 0. If negative result enter 0	6 x difference
8	Weight#	Percentage deviation from standard weight is (1-[weight/standard weight])*100. # Limit of -10 (result must be ≥ - 10)	1
CDAI score			TOTAL

CDAI = Crohn's Disease Activity Index; Hct = hematocrit
{Sandborn 2002}

There are 2 lookup tables (1 for males; 1 for females) that are used to determine the standard weight for each subject. Two centimeters should be added to the eCRF collected height to account for height being collected shoeless. The look up tables were generated by a chart that is commonly used for CD studies for the Filgotinib program. The lookup tables exist as SAS datasets in the following celbook area: \\gsasdata\biometrics\global\cel\celbooks\IBD_ST_WT\0.

A Responder for CD is defined as a subject who has a reduction of ≥ 100 points in total CDAI score compared to baseline at the specified assessment timepoint. A subject with a baseline total CDAI score of ≥ 220 to ≤ 250 will be considered to be a Responder if a CDAI score of < 150 is attained at the specified timepoint. A Non-responder will be defined as a subject who does not fulfill the definition of responder at the specified assessment timepoint.

Disease Worsening for subjects with CD is defined by a total CDAI score after Week 13 consisting of:

- a score ≥ 220 on 2 consecutive visits, and
- an increase of ≥ 100 points from the Week 13 value on 2 consecutive visits

8. SAFETY ANALYSES

The safety analysis will be conducted on the Safety Analysis Set, defined in Section 3.1.3 unless otherwise noted.

8.1. Adverse Events and Deaths

Up to 3 different displays of adverse event data may be presented as described in Section 3.2 of this SAP:

"As-treated" analysis: Subjects who switched from placebo to open-label filgotinib will have adverse events occurring while on double-blind study drug summarized with placebo, and adverse events occurring on/after first dose date of open-label filgotinib to last dose date of filgotinib plus 30 days (if permanently discontinued), or all available data if ongoing at time of an interim analysis summarized with filgotinib. All other subjects will have adverse events occurring on/after first dose date of any study drug to last dose date of any drug plus 30 days (if permanently discontinued), or all available data if ongoing at time of interim analysis summarized by treatment received. This analysis will summarize all on-treatment adverse events and will be used to scan for all safety events during the study.

During Double-Blind Phase of the Study: Data will be included up to first dose date of open-label filgotinib minus 1 day for subjects who switched to open-label filgotinib; last dose date of double-blind study drug plus 30 days for those who have permanently discontinued from study drug; or all available data while on double-blind drug if ongoing at the time of an interim analysis. This analysis will summarize adverse events during the same timeframe as the primary analysis for sperm/semen data.

During Open-Label Filgotinib Phase of the Study: Data will be included from OL Day 1 to last dose date of open-label filgotinib plus 30 days for those who have permanently discontinued from open-label filgotinib; or all available data if ongoing at the time of an interim analysis. These displays will summarize data for subjects in the Open-Label Filgotinib Safety Analysis Set. This analysis appropriately adjusts the denominator for interruptions and discontinuations for open-label filgotinib to only those subjects who took open-label filgotinib.

8.1.1. Adverse Event Dictionary

Clinical and laboratory adverse events (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.

8.1.2. Adverse Event Severity

Adverse events are graded by the investigator as Grade 1, 2, 3, 4, or 5 according to Common Terminology Criteria for Adverse Events (CTCAE) Grading Scale. If CTCAE criteria do not exist for a specific event, Table 7-1 of the study protocol provides guidance for the investigator to grade the event. 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.

8.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE eCRF to the question of "Related to Study Treatment." Relatedness will always default to the investigator's choice, not that of the medical monitor. 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.

8.1.4. Serious Adverse Events

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

8.1.5. Treatment-Emergent Adverse Events

8.1.5.1. Definition of Treatment-Emergent Adverse Events

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

- Any AEs with an onset date on or after the study drug start date and no later than 30 days after permanent discontinuation of study drug
- Any AEs leading to premature discontinuation of study drug

In "as-treated" displays, subjects who took double-blind placebo and switched to open-label filgotinib with TEAEs with an onset date on/after first dose date of double-blind placebo and prior to the first dose date of open-label filgotinib will be attributed to the placebo period (and summarized with the placebo treatment group). TEAEs with an onset date on or after the first dose date of open-label filgotinib up to last dose date of open-label filgotinib + 30 days or AEs that led to premature discontinuation of OL filgotinib will be attributed to the filgotinib period (and summarized with the filgotinib treatment group).

8.1.5.2. 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 determine whether an AE is treatment emergent. The event is considered treatment emergent if both of the following 2 criteria are met:

- The AE onset 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 date of the last dose of study drug

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 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.

For "as-treated" displays, subjects who took double-blind placebo and switched to open-label filgotinib, reporting an AE with incomplete onset date that is prior to the month and year (or year) of the first dose date of open-label filgotinib will have that AE attributed to the placebo period (and summarized with the placebo treatment group). An AE with incomplete onset date that is in the same or after the month and year (or year) of the first dose date of open-label filgotinib up to last dose date of open-label filgotinib + 30 days will be attributed to the filgotinib period (and summarized with the filgotinib treatment group).

In addition, an AE with completely missing onset date or incomplete onset date that is same as the month and year (or year) of the first dose date of open-label filgotinib, and stop date that is prior to the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose of open-label filgotinib will be attributed to the placebo period (and summarized with the placebo treatment group). An AE with completely missing onset date or incomplete onset date that is same as the month and year (or year) of the first dose date of open-label filgotinib, and stop date that is missing, or the same as, or after the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose of open-label filgotinib up to last dose date of open-label filgotinib + 30 days will be attributed to the filgotinib period (and summarized with the filgotinib treatment group).

8.1.5.3. Analysis of Exposure-Adjusted Incidence Rates

The Exposure-Adjusted Incidence Rate (EAIR) is defined as:

$$Incidence\ rate\ per\ 100\ patient\ years\ of\ exposure\ (PYE) = \frac{Total\ number\ of\ subjects\ with\ an\ event}{Total\ PYE} \times 100.$$

The total patient-years of exposure (PYE) to a treatment is the sum of the individual subject's PYE, which is defined as:

- For subjects with an event:

$$PYE = (Event\ start\ date - \ first\ dosing\ date + 1)/\ 365.25.$$

When calculating PYE, a partial AE onset date will be imputed as follows:

- If the non-missing portion of the AE onset date (month and/or year) is the same as the portion of the first dosing date (month and/or year) of start of study drug, the AE onset date will be imputed as the first dosing date of study drug; otherwise it will be imputed as the 1st day of month (if only day is missing) or January 1st of the year (if day and month are missing).

- If the AE onset date is completely missing, the AE onset date will be imputed as the first dosing date of study drug.

- For subjects with no event:

$$PYE = (Calculation\ end\ date - \ first\ dosing\ date + 1)/\ 365.25,$$

For double-blind phase:

First dosing date = double-blind first dose date

Calculation end date = earliest non-missing date of (last dosing date of double-blind drug + 30, first dosing date of OL filgotinib (if applicable) – 1, data cutoff date, death date)

For as-treated analysis:

Filgotinib exposure for Placebo to OL filgotinib: first dosing date = first dose date OL filgotinib; calculation end date is earliest non-missing date among (last dose date of OL filgotinib + 30 days, data cutoff date, death date)

Placebo exposure for Placebo to OL filgotinib: first dosing date = first dose date placebo; calculation end day is earliest non-missing date among (first dose date OL filgotinib – 1, data cutoff date, death date).

Exposure for subjects remaining on double-blind drug: first dosing date = first dose date double-blind drug; calculation end date is earliest date among (last dose date double-blind drug + 30, data cutoff date, death date)

Filgotinib exposure for Filgotinib to OL filgotinib: first dosing date = first dose date double-blind drug; calculation end day is earliest non-missing date among (last dose date of open-label filgotinib + 30, data cutoff date, death date).

The exact 95% CI based on Poisson distribution {Ulm 1990} for EAIR is defined as:

$$\left(\frac{\chi^2_{2(Total\ number\ of\ subjects\ with\ an\ event),0.025}}{2\times Total\ PYE}, \frac{\chi^2_{2(1+Total\ number\ of\ subjects\ with\ an\ event),0.975}}{2\times Total\ PYE}\right)\times 100.$$

Partial last dosing dates of double-blind phase will be imputed as described in section 4.2.1.

In addition to by treatment group summary, the EAIR difference with 95% CI for filgotinib minus placebo will be presented. The Method of Variance Estimates Recovery (MOVER)-type confidence interval described in {Li 2011} will be used to calculate EAIR difference and 95% CI:

$$L = \frac{n_a}{T_a} - \frac{n_b}{T_b} - \sqrt{(\frac{n_a}{T_a} - l_a)^2 + (u_b - \frac{n_b}{T_b})^2}\ ,\ U = \frac{n_a}{T_a} - \frac{n_b}{T_b} + \sqrt{(u_a - \frac{n_a}{T_a})^2 + (\frac{n_b}{T_b} - l_b)^2},$$

where n_a is the total number of subjects with the event, T_a is the total PYE, l_a is the lower bound and u_a is the upper bound of the exact 95% CI based on Poisson distribution for group *a*. Similarly for group *b*.

8.1.6. Summaries of Adverse Events and Deaths

A brief, high-level summary of the number and percentage of subjects who experienced at least 1 TEAE in the categories described below will be provided by treatment for the double-blind phase, open-label filgotinib phase, and for an "as-treated" display:

- TEAE
- TEAEs with Grade 3 or higher (further broken down to Grade 3, Grade 4, and Grade 5)
- TEAEs with Grade 2 or higher
- TE treatment-related AE
- TE treatment-related AEs with Grade 3 or higher (further broken down to Grade 3, Grade 4, and Grade 5)
- TE treatment-related AEs with Grade 2 or higher
- TE SAE
- TE treatment-related SAE

- TEAE that led to premature discontinuation of double-blind study drug (displayed for double-blind phase analysis only)
- TEAE that led to premature discontinuation of open-label filgotinib study drug (displayed for open-label phase analysis only)
- TEAE that led to temporary interruption of double-blind study drug (displayed for double-blind phase analysis only)
- TEAE that led to temporary interruption of open-label study drug (displayed for open-label phase analysis only)
- TEAE that led to premature discontinuation of study
- All deaths observed in the study (if subject received ≥ 1 dose of open label filgotinib will be summarized with open-label filgotinib; if last study drug received was double-blind, will be summarized with double blind study drug)

The number and percentage of subjects who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and treatment for the double-blind phase and for an "as-treated" analysis.

The following 2 tables will be displayed for the double-blind phase of the study for subjects in the Safety Analysis Set:

- The number and percentage of subjects who experienced TEAEs leading to temporary interruption of double-blind study drug
- AEs leading to premature discontinuation of double-blind study drug

Similar tables will be displayed for the open-label filgotinib phase of the study for subjects in the Open-Label Filgotinib Safety Analysis Set:

- The number and percentage of subjects who experienced TEAEs leading to temporary interruption of open-label filgotinib
- AEs leading to premature discontinuation of open-label filgotinib

For other AEs described below, summaries (for the double-blind phase and for "as-treated" analysis) will be provided by SOC, PT, and treatment:

- All TEAEs
- All TEAEs by Concomitant Medication Use (Immunomodulator and Corticosteroid Use)
- TEAEs of Grade 3 or higher (by maximum severity)

- TEAEs of Grade 2 or higher
- TE treatment-related AEs
- TE Treatment-related AEs of Grade 3 or higher (by maximum severity)
- TE SAEs
- TE SAEs by Concomitant Medication Use (Immunomodulator and Corticosteroid Use)
- TE treatment-related SAEs
- TEAEs leading to premature discontinuation of study
- TEAEs leading to death

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 to the above summary tables, all TEAEs, SAEs, and TE treatment-related AEs will be summarized (for the double-blind phase and for "as-treated" analysis) by PT only, in descending order of total frequency.

In addition, data listings will be provided for the following:

- All AEs, indicating whether the event is treatment emergent
- All AEs of Grade 3 or higher
- SAEs
- Deaths
- COVID-19 Adverse Events
- All AEs leading to temporary interruption of double-blind study drug
- All AEs leading to temporary interruption of open-label filgotinib study drug
- AEs leading to premature discontinuation of double-blind study drug
- AEs leading to premature discontinuation of open-label filgotinib study drug

8.1.7. Adverse Events of Interest

Adverse events of interest (AEIs) 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 infection 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 Standardized MedDRA Query (SMQ)--narrow scope
  - Active tuberculosis, using an MST list developed by Gilead
  - Hepatitis B or C infections, using an 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 arterial or venous thrombosis and/or 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

For the double-blind phase, the number of subjects with a TE AEI (and SAE), the corresponding PYE, the calculated EAIR, and the exact 95% CIs (defined in Section 8.1.5.3) will be summarized for each treatment group by PT. The EAIR difference (filgotinib – placebo) and corresponding 95% CIs for the difference between treatment groups will also be presented for the double-blind phase of the study.

For the overall study, the number of subjects with a TE AEI in each category (and separately for SAEs), the corresponding PYE, the calculated EAIR, and the exact 95% CIs (defined in Section 8.1.5.3) will be summarized for each treatment group (filgotinib and placebo) by PT using an as-treated analysis. Filgotinib group includes the filgotinib exposure for the placebo to filgotinib sequence, and exposure for DB filgotinib and filgotinib to OL filgotinib treatment sequences. Placebo will include DB Placebo and the placebo exposure for the Placebo to OL filgotinib treatment sequence.

Data listings for each AEI will also be provided.

8.1.7.1. Cardiovascular Event Adjudication Committee

An independent cardiovascular event adjudication committee (CVEAC) reviews and adjudicates all potential MACE and thromboembolic events in a blinded manner. To identify potential MACE and thromboembolic events, the following AEs will be sent for adjudication. Please refer to the Cardiovascular Event Adjudication Committee Charter for more details.

- All AEs leading to death
- CV events (meeting serious criteria), utilizing a MST list developed by Gilead
- MI, utilizing a narrow scope Standardized MedDRA Query (SMQ)
- Unstable angina (meeting hospitalization criteria), utilizing a MST list developed by Gilead
- Transient ischemic attack, utilizing a MST list developed by Gilead
- Stroke, utilizing a MST list developed by Gilead
- Cardiac failure (meeting hospitalization criteria), utilizing a MST list developed by Gilead
- Percutaneous coronary intervention, utilizing a MST list developed by Gilead
- Embolic and thrombotic events, utilizing a narrow scope SMQ

The CVEAC will review the above AEs, and related clinical data to adjudicate whether the criteria for MACE (CV death, MI, and/or stroke), ASTE, and VTE have been met for each AE.

The number and percentage of subjects with positively adjudicated TE MACE, TE ASTE, and TE VTE will be summarized by treatment group using the adjudicated category if applicable for the final analysis of the study only.

A by-subject listing for subjects with potential events for adjudication (MACE, ASTE, and VTE) and their respective adjudication results will be provided for interim and final analyses of the study.

A by-subject listing of thromboembolic history and risk factors will be provided for subjects with potential events for adjudication (MACE, ASTE, and VTE), if applicable.

8.1.8. Analysis of Adverse Events by Treatment Sequence (Final Analysis only)

For the final analysis (when all long-term data is available), an additional "as-treated" analysis of All TEAEs, SAEs, and AEIs will be provided. The number and percentage of subjects with events will be displayed by overall treatment group and for each of the treatment sequences within a treatment group as shown below:

For placebo group,

All Placebo	DB Placebo E/T Prior to Week 26*	Placebo exposure: DB Placebo to OL Filgotinib	DB Placebo at/after Week 26*

For filgotinib group,

All Filgotinib	DB Filgotinib E/T Prior to Week 26*	Filgotinib exposure: DB Placebo to OL Filgotinib	DB Filgotinib to OL Filgotinib	DB Filgotinib at/after Week 26*

DB = double-blind; E/T = early termination of study drug; OL = open-label; * = subject never received OL filgotinib (either met sperm stopping criteria and entered monitoring phase or E/T for other reasons).

8.2. Laboratory Evaluations

Laboratory data collected during the study will be summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set and will be by actual treatment using an "as-treated" analysis. For "as-treated" displays, subjects on double-blind placebo who switch to open-label filgotinib will have data collected up to first dose date of filgotinib attributed to placebo (summarized with placebo) and data collected after first dose of open-label filgotinib to last dose date of filgotinib + 30 days attributed to filgotinib (summarized with filgotinib). Other subjects will have post-baseline data included up to last dose date + 30 days if terminated from study drug; or all available data at time of an interim analysis.

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. Hemolyzed test results 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, and urinalysis separately. Values falling outside of the relevant reference range and/or having a severity grade of 1 or higher based on CTCAE criteria will be flagged in the data listings, as appropriate.

No formal statistical testing is planned.

8.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics will be provided by treatment group for selected laboratory tests as follows:

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

A baseline laboratory value will be defined as the last measurement obtained on or prior to the first dose date of double-blind study drug. For "as-treated" displays, subjects who took double-blind placebo and switched to open-label filgotinib will have the baseline value used for placebo period and the Open-label Baseline will be used for the analysis for data collected after the first dose of open-label filgotinib. Data collected while on double-blind placebo will be summarized with placebo treatment, and data collected while on open-label filgotinib will be summarized with filgotinib. Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value (or minus OL baseline for the filgotinib period for "as-treated" analysis for subjects who switched from double-blind placebo to OL filgotinib). 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) change from baseline for selected laboratory tests will be plotted using a line plot by treatment group and time point.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.7.3.

8.2.2. Graded Laboratory Values

The CTCAE version 4.03 will be used to assign 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.

8.2.2.1. Treatment-Emergent Laboratory Abnormalities

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days for subjects who permanently discontinued study drug, or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

For "as-treated" displays and subjects who took double-blind placebo and switched to open-label filgotinib, treatment-emergent laboratory abnormalities for the placebo period (summarized with placebo group) are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, up to and including the first dose date of open-label filgotinib. Treatment-emergent laboratory abnormalities of the filgotinib period (summarized with filgotinib group) are defined as values that increase at least 1 toxicity grade from the Open-label Baseline up to and including the date of last dose of any study drug plus 30 days, or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

For all other subjects, treatment-emergent laboratory abnormalities are defined as values that increase 1 toxicity grade from the double-blind baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days (or all available data if ongoing at an interim analysis).

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.

8.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

Treatment-emergent marked laboratory abnormalities are defined as values that increase from baseline by at least 3 toxicity grades at any postbaseline time point, up to and including the date of the last dose of any study drug plus 30 days for subjects who permanently discontinued study drug or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

For "as-treated" displays and placebo-treated subjects who switch to open-label filgotinib, treatment-emergent marked laboratory abnormalities of the placebo period (summarized with placebo group) are defined as values that increase at least 3 toxicity grades from baseline at any postbaseline time point, up to and including the first dose date of open-label filgotinib. Treatment-emergent laboratory abnormalities during the open-label filgotinib period (summarized with filgotinib group) are defined as values that increase at least 3 toxicity grades from Open-label Baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days, or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

For all other subjects, treatment-emergent laboratory abnormalities are defined as values that increase 3 toxicity grades from the double-blind baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days (or all available data if ongoing at an interim analysis).

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.

8.2.2.3. Summaries of Laboratory Abnormalities

The following summaries (number and percentage of subjects) for treatment-emergent laboratory abnormalities will be provided by lab test and treatment group; subjects will be categorized according to the most severe postbaseline abnormality grade for a given lab test:

- Graded laboratory abnormalities (for the final analysis of the study only)
- Grade 3 or 4 laboratory abnormalities
- Marked laboratory abnormalities

For all summaries of laboratory abnormalities, the denominator is the number of subjects with nonmissing postbaseline values up to the first dose date of open label filgotinib for placebo subjects who switch to open-label filgotinib; up to 30 days after permanent discontinuation of drug; or to the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

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.

8.2.3. Liver-related Laboratory Evaluations

Liver-related abnormalities will be examined and summarized using an "as-treated" analysis of the number and percentage of subjects with at least 1 value in the reporting period for the following:

- AST: (a) > 3 times 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 their most severe postbaseline value up to last dose date plus 30 days. For 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 at least 1 nonmissing postbaseline value 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.

For "as-treated" displays and subjects randomized to placebo who switched to open-label filgotinib, the most severe postbaseline value collected up to the first dose date of open-label filgotinib will be summarized with the placebo group. The most severe postbaseline value collected after first dose date and up to the last dose date of open-label filgotinib plus 30 days will be summarized with the filgotinib group.

8.3. Body Weight and Vital Signs

Descriptive statistics will be provided by treatment using an "as-treated" analysis for body weight, BMI and vital signs (pulse, systolic and diastolic blood pressure, respiration rate, and temperature) as follows:

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

A baseline value will be defined as the last available value collected on or prior to the date of first dose of study drug. For "as-treated" displays and subjects who took double-blind placebo and switched to open-label filgotinib, the baseline value will be used to calculate change from baseline during the double-blind placebo period (summarized with placebo group). The Open-Label Baseline will be used to calculate change from baseline for values collected after the first dose date and up to last dose date of open-label filgotinib plus 30 days (summarized with filgotinib group). Change from baseline to a postbaseline visit will be defined as the postbaseline value minus the baseline value.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.7.3. 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.

8.4. Prior and Concomitant Medications

Medications collected at screening and during the study will be coded using the current version of the WHO Drug dictionary.

8.4.1. Prior Medications

Prior medications are defined as any medications taken before a subject took their first dose of study drug.

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 assumed to be a prior medication, unless otherwise specified.

Prior medications will be listed as part of the Prior and Concomitant Medications listing.

8.4.2. Concomitant Medications

Concomitant medications are defined as medications taken while a subject is also taking study drug. For "as-treated" displays and subjects randomized to placebo and switching to open-label filgotinib, the medications taken between first and last dose dates of double-blind placebo will be summarized with the placebo group. For the open-label phase, medications taken between first and last dose dates of open-label filgotinib will be summarized with the filgotinib group. Use of concomitant medications will be summarized by ATC 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 provided by preferred term in descending overall frequency. For drugs with the same frequency, sorting will be done alphabetically.

For the purposes of analysis, any medications with a start date prior to or on the first dosing date of study drug and continued to be taken 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 (ie, the medication is specifically marked as a prior medication on the eCRF).

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.

8.5. Electrocardiogram (ECG) Results

A shift table of the investigator's assessment of ECG results at Study Week 26/OL Week 13/ET and at LTE Week 195/ET compared with baseline values will be presented by treatment sequence (DB Filgotinib; DB Filgotinib to OL Filgotinib, DB Placebo; DB Placebo to OL Filgotinib) using the following categories: normal; abnormal (not clinically significant); abnormal (clinically significant); or missing. Baseline values will be ECGs collected with a nominal visit of "Screening"; Week 26 will have nominal visits of "Week 26", "OL Week 13"; or "Early Termination"--if early termination (ET) prior to Week 26 (subject did not enroll in LTE), and LTE Week 195/ET will have nominal visits of "LTE Week 195" or "Early Termination" if early termination after Week 26 (enrolled in LTE).

The number and percentage of subjects in each cross-classification group of the shift table will be presented by treatment sequence. Subjects with a missing value at 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 visit in chronological order.

9. PHARMACOKINETIC ANALYSES

Concentrations of filgotinib and GS-829845 in plasma will be determined using validated bioanalytical methods. The PK analysis will be conducted on the PK Analysis Set, defined in Section 3.1.4.

Individual subject concentration data for filgotinib and GS-829845 will be listed (including date and time of last filgotinib study drug administration; PK collection date and time; and time post- dose that the PK sample was collected).

10. REFERENCES

Cooper TG, Noonan E, von Eckardstein S, Auger J, Baker HW, Behre HM, et al. World Health Organization reference values for human semen characteristics. Human reproduction update 2010;16 (3):231-45.

Jarvi K, Dula E, Drehobl M, Pryor J, Shapiro J, Seger M. Daily vardenafil for 6 months has no detrimental effects on semen characteristics or reproductive hormones in men with normal baseline levels. J Urol 2008;179 (3):1060-5.

Koch GG, Carr GJ, Amara IA, Stokes ME, Uryniak TJ. Categorical Data Analysis. Chapter 13 in Berry, D.A. (ed.). Statistical Methodology in the Pharmaceutical Sciences. New York: Marcel Dekker, Inc., 1989:pp. 414-21.

Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons, Inc; 1987.

Sandborn WJ, Feagan BG, Hanauer SB, Lochs H, Lofberg R, Modigliani R, et al. A review of activity indices and efficacy endpoints for clinical trials of medical therapy in adults with Crohn's disease. Gastroenterology 2002;122 (2):512-30.

U. S. Department of Health and Human Services (DHHS), Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER). Testicular Toxicity: Evaluation During Drug Development Guidance for Industry. 2018.

World Health Organization (WHO). WHO Laboratory manual for the examination of human semen and sperm-cervical mucus interaction. Third Edition. 1992.

World Health Organization (WHO). WHO Laboratory manual for the Examination and processing of human semen. Fifth Edition. 2010.

11. SOFTWARE

SAS® Software Version 9.4. SAS Institute Inc., Cary, NC, USA.

nQuery Advisor(R) Version 4.0. Statistical Solutions, Cork, Ireland.

12. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision

13. APPENDICES

Appendix 1. Study Procedures Table
Appendix 2. Calculation of Difference in Proportions and 95% CI (Stratum-Adjusted Mantel-Haenszel Proportions)

Appendix 1. Study Procedures Table

Study Procedure	Screening[g]	Part A (Section 6.4 Protocol)					Part B (Section 6. 5 Protocol)		Open-Label Filgotinib Phase (Section 6.6 Protocol)						Long Term Extension (Section 6.10 Protocol)	30 Day Safety Follow-up[e]	ET[f]
		Day 1	Wk 2	Wk 4	Wk 8	Wk 13[a]	Wk 20	Wk 26[a]	Sperm Parameter Eligibility Confirmation[b]	OL Day 1	OL Wk 2	OL Wk 4	OL Wk 8	OL Wk 13[c]	LTE Week 13 through LTE Week 195[d]		
	45 days		±5 days	±5 days	±5 days	+ 5 days	±5 days	+ 5 days			±5 days	±5 days	±5 days	+ 5 days	±10 days	±5 days	
Informed Consent	X																
Medical History, Number of times subject impregnated women, IBD History, and Demographics	X																
Inclusion/exclusion criteria review	X	X															
Complete Physical Exam[g]	X																
Symptom-directed Physical Exam, as needed		X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Vital Signs[g] and Weight	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Height	X																
12-lead ECG[g]	X							X						X	X[h]		X
Flexible Sigmoidoscopy[i]	X																
Ileocolonoscopy[i]	X																
Complete Mayo Clinic Score (complete MCS)[j]	X																
Partial Mayo Clinic Score (partial MCS)[j]		X	X	X	X	X	X	X		X	X	X	X	X	X		X

Study Procedure	Screening[g]	Part A (Section 6.4 Protocol)					Part B (Section 6. 5 Protocol)		Open-Label Filgotinib Phase (Section 6.6 Protocol)						Long Term Extension (Section 6.10 Protocol)	30 Day Safety Follow-up[e]	ET[f]
		Day 1	Wk 2	Wk 4	Wk 8	Wk 13[a]	Wk 20	Wk 26[a]	Sperm Parameter Eligibility Confirmation[b]	OL Day 1	OL Wk 2	OL Wk 4	OL Wk 8	OL Wk 13[c]	LTE Week 13 through LTE Week 195[d]		
	45 days		±5 days	±5 days	±5 days	+ 5 days	±5 days	+ 5 days			±5 days	±5 days	±5 days	+ 5 days	±10 days	±5 days	
Simple Endoscopic Score for Crohn's Disease (SES-CD)	X																
Crohn's Disease Activity Index (CDAI)[k]	X	X	X	X	X	X	X	X		X	X	X	X	X	X		X
TB (QuantiFERON®) Test	X[l]														X[m]		
Chest X-ray[n]	X																
Urinalysis[o]	X	X				X		X		X				X	X		X
Urine drug screen[o]	X																
Hematology and Chemistry[o]	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Lipid profile (fasting)[o]		X				X		X		X				X	X[p]		
Serum Immunoglobulin[o]	X					X		X						X	X		
Endocrine: TSH, HbA1c[o]	X																
LH, FSH, inhibin B, total Testosterone[o] collection time between 07:00-11:00 in the morning	X	X		X		X		X		X		X		X	X	X	X
Stool[o]	X																
Concomitant Medications	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Assessment of Adverse Events	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X

Study Procedure	Screening[g]	Part A (Section 6.4 Protocol)					Part B (Section 6. 5 Protocol)		Open-Label Filgotinib Phase (Section 6.6 Protocol)						Long Term Extension (Section 6.10 Protocol)	30 Day Safety Follow-up[e]	ET[f]
		Day 1	Wk 2	Wk 4	Wk 8	Wk 13[a]	Wk 20	Wk 26[a]	Sperm Parameter Eligibility Confirmation[b]	OL Day 1	OL Wk 2	OL Wk 4	OL Wk 8	OL Wk 13[c]	LTE Week 13 through LTE Week 195[d]		
	45 days		±5 days	±5 days	±5 days	+ 5 days	±5 days	+ 5 days			±5 days	±5 days	±5 days	+ 5 days	±10 days	±5 days	
Semen Collection (2 samples as per collection instructions in Protocol Section 6.13)[q]	X[r]					X		X	X[s]					X[t]	X		X
Date and time of most recent ejaculation[u]	X					X		X	X					X	X		X
HIV, Hepatitis B, and Hepatitis C[v]	X																
HBV DNA[w]						X		X						X	X		
Randomization		X															
Study Drug Accountability				X	X	X	X	X		X		X	X	X	X		X
Study Drug Dispensation		X		X	X	X	X	X		X		X	X	X	X		
PK collection (sparse)			X[x]	X[y]		X[z]		X[z]			X[x]	X[y]		X[z]			
In-Clinic Dosing[aa]		X	X			X		X			X			X			

a Weeks 13 and 26 must occur after 13 weeks and 26 weeks of drug exposure, respectively. Therefore, visit window for these visits are +5 days
b Subjects on blinded study drug meeting criteria as a non-responder (Week 13) or for Disease Worsening (after Week 13 through 26), and not meeting the pre-specified decrease threshold in sperm parameters, are eligible for open-label filgotinib (refer to Appendix 11 and Section 3 of the protocol). For definitions of IBD response status, Disease Worsening, and/or pre-specified decrease threshold in sperm parameters please refer to Definition of Terms and Section 3.4 of the protocol.
c OL Week 13 must occur after 13 weeks of starting open-label filgotinib. Therefore, visit window is +5 days.
d Subjects in the LTE will undergo scheduled visits for safety assessments and semen sample collection every 13 weeks, starting with LTE Week 13. Semen monitoring will be collected until the Week 13 primary study results are analyzed.
e Any time a subject discontinues study drug, a safety follow-up visit, 30 days (± 5 days) after the last dose of study drug is required, even if the subject is entering the Monitoring Phase
f Any time a subject discontinues participation in the study prior to Week 26, OL Week 13, reaching reversibility or Monitoring Phase Week 52 (whichever occurs first), or LTE Week 195, an ET Visit is required. Subjects who are Responders at Week 26 or at OL Week 13 and choose not to continue in the LTE are also required to complete ET visit assessments.
g As detailed in protocol Appendix 7.

h 12-Lead ECG will be performed at LTE Week 195 or ET, whichever occurs sooner
i Flexible Sigmoidoscopy (UC patients) or ileocolonoscopy (CD patients) is only required if it has not been done in the last 90 days.
j Partial MCS includes all parts of the complete MCS except endoscopy. For definitions of IBD response status, please refer to Definition of Terms; for definitions of Disease Worsening, please refer to protocol Section 3.4 and Definition of Terms. Subject recall for the last 24 hours will be used for the patient reported components of the complete MCS and partial MCS. Subjects on blinded study drug meeting criteria for Non-Responder at Week 13 or Disease Worsening after Week 13 through Week 26 (and not meeting the pre-specified decrease threshold in sperm parameters) will be eligible for open-label filgotinib (see protocol Section 3). MCS and partial MCS will be collected for UC patients. Physician Global Assessment is a component of MCS and partial MCS.
k CDAI will be collected for CD patients. Subject recall for the last 7 days will be used for the patient reported outcomes of the CDAI. The CDAI score calculations should not be inclusive of days that involve an endoscopy procedure or preparation for that procedure.
l 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 Gilead Medical Monitor or designee.
m In the LTE subjects must have yearly QuantiFERON® testing. Yearly TB testing begins 1 year from the screening TB test date. If yearly TB testing falls between study visits, the yearly TB testing should be performed at the visit prior to 1 year from the screening TB test date. Subjects with newly positive (converted) QuantiFERON® [or centrally reported equivalent assay] TB test should be discontinued from study drug. Subjects who were previously treated for TB with a complete and adequate course of therapy as per local standard of care and as verified by the investigator do not need to have yearly QuantiFERON® tests. Subjects previously treated for TB should be screened at least yearly for signs and symptoms consistent with reactivation of TB. Any subject with active TB should be discontinued from study.
n 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
o As detailed in protocol Appendix 6. For visits that require fasting, subjects should not have any food or drink (except water) for at least 8 hours before the visit.
p Lipid profile should be performed every 26 weeks
q Semen samples will be collected on the visit day and/or after the visit. Those meeting the pre-specified decrease threshold in sperm parameters (as defined in the Definition of Terms) must enter the Monitoring Phase (see schedule below).
r The screening semen sample collection should coincide with the Day 1 (baseline) visit as much as possible.
s At time of confirmation of Non-responder status/Disease Worsening and prior to initiation of open-label filgotinib, Semen Analysis must confirm subject did not meet the pre-specified decrease threshold in sperm parameters.
t At time of confirmation of Non-responder status/Disease Worsening and prior to entry into the LTE, Semen Analysis must confirm subject did not meet the pre-specified decrease threshold in sperm parameters.
u This question will be asked prior to each semen collection.
v An HIV-1/HIV-2 antigen/antibody test, a Hepatitis C virus antibody test, a Hepatitis B surface antigen, a Hepatitis B surface antibody and a Hepatitis core antibody test will be conducted on all subjects; Subjects with positive Hepatitis B surface antigen (HBsAg) at Screening are excluded from the study. Subjects with positive HBV core Ab and negative HBsAg, require reflex testing for HBV DNA. Subjects with positive HBV DNA at Screening will be excluded. Subjects with positive HBV core Ab and negative HBV DNA are eligible per investigator judgment but may require prophylactic treatment in accordance with HBV treatment guidelines/local standard of care and require ongoing monitoring with blood tests for HBV DNA every 3 months. Subjects with evidence of active Hepatitis B during the study, as evidenced by HBV DNA positivity, will be discontinued from study drug as outlined in the protocol. Subjects with positive HCV antibody (Ab) at Screening, require reflex testing for HCV RNA. Subjects with positive HCV RNA at Screening will be excluded. Subjects with positive HCV Ab, but negative HCV RNA are eligible per investigator judgment.
w Subjects with positive HBV core Ab and negative HBV DNA at Screening require ongoing monitoring with blood tests for HBV DNA every 3 months
x Sparse plasma PK samples are collected at least 30 minutes post-dose.
y Sparse plasma PK samples can be collected at any time without regard to dosing.
z Sparse plasma PK samples are collected pre-dose.
aa The subject will take the first dose (Day 1) in the clinical study center. For PK collection purposes, at Weeks 2, 13, and 26, and OL Weeks 2 and 13, subjects should be instructed **not** to take their study drug, but rather to bring it with them to the clinic. Subjects will be instructed to take their dose during their scheduled visit, as detailed in protocol Section 6.1.6.

Study Procedure	Monitoring Phase (Section 6.7) Visits occur every 13 weeks (±5 days) for up to 52 weeks or until Reversibility is met, whichever is sooner	ET[a]
Symptom-directed Physical Exam, as needed	X	X
Vital Signs[b] and Weight	X	X
12-lead ECG		X
Partial Mayo Clinic Score (partial MCS)[c]		X
Crohn's Disease Activity Index (CDAI)[d]		X
Urinalysis[e]	X	X
Hematology and Chemistry[e]	X	X
LH, FSH, inhibin B, total Testosterone[e] collection time between 07:00-11:00 in the morning	X	X
Concomitant Medications	X	X
Assessment of Adverse Events[f]	X	X
Semen Collection (2 samples as per collection instructions in protocol Section 6.12)[g]	X	
Date and time of most recent ejaculation[h]	X	

a Not required if Reversibility is met or if subject completes Week 52
b As detailed in protocol Appendix 7.
c Partial MCS includes all parts of the complete MCS except endoscopy. For definitions of IBD response status, please refer to Definition of Terms; for definitions of Disease Worsening, please refer to protocol Section 3.4 and Definition of Terms. Subject recall for the last 24 hours will be used for the patient reported components of the complete MCS and partial MCS. MCS and partial MCS will be collected for UC patients. Physician Global Assessment is a component of MCS and partial MCS.
d CDAI will be collected for CD patients. Subject recall for the last 7 days will be used for the patient reported outcomes of the CDAI. The CDAI score calculations should not be inclusive of days that involve an endoscopy procedure or preparation for that procedure.
e As detailed in protocol Appendix 6
f As detailed in protocol Section 7.3
g Semen samples will be collected on the visit day and/or after the visit
h This question will be asked prior to each semen collection.

Appendix 2. Calculation of Difference in Proportions and 95% CI (Stratum-Adjusted Mantel-Haenszel Proportions)

The 95% confidence interval on the difference in proportions (proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration) between Treatment A (filgotinib) and Treatment B (placebo) will be constructed based on stratum-adjusted MH proportions {Koch 1989}:

$$P_A - P_B \pm Z_{(1-\alpha/2)} * SE(P_A - P_B),$$

where

- $(P_A - P_B) = \frac{\sum w_h d_h}{\sum w_h}$, is the stratum-adjusted MH proportion difference, where $d_h = p_{Ah} - p_{Bh}$ is the difference in the proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration between filgotinib and placebo in stratum h (h=1 to 2). Strata are (1) sperm concentration 15-50 M/mL; (2) sperm concentration > 50 M/mL.

- $w_h = \frac{n_{Ah} n_{Bh}}{n_{Ah} + n_{Bh}}$, is the weight based on the harmonic mean of sample size per treatment group for each stratum where n_{Ah} and n_{Bh} are the sample sizes of the Treatment Groups A (filgotinib) and B (placebo) in stratum h.

- $SE(P_A - P_B) = \sqrt{\frac{\sum w_h^2 \left[\frac{p^*_{Ah}(1-p^*_{Ah})}{n_{Ah}-1} + \frac{p^*_{Bh}(1-p^*_{Bh})}{n_{Bh}-1} \right]}{(\sum w_h)^2}}$, where $p^*_{Ah} = \frac{m_{Ah} + 0.5}{n_{Ah} + 1}$ and $p^*_{Bh} = \frac{m_{Bh} + 0.5}{n_{Bh} + 1}$ and m_{Ah} and m_{Bh} are the number of subjects with a ≥ 50% decrease in sperm concentration in the Treatment Groups A (filgotinib) and B (placebo) in stratum h.

- $\alpha = 0.05$ for this study

- $Z_{(1-\alpha/2)} = Z_{0.975} = 1.96$ is the 97.5th percentile of the normal distribution

If the computed lower confidence bound is less than –1, the lower bound is defined as –1. If the computed upper confidence bound is greater than 1, then the upper bound is defined as 1. If the row or column marginal totals are zero in any strata, then some regrouping of strata may be performed.

GS-US-418-4279_SAP_v1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Regulatory Affairs eSigned	18-Jan-2021 04:48:48
PPD	Clinical Research eSigned	19-Jan-2021 16:05:07
PPD	Biostatistics eSigned	19-Jan-2021 17:37:40