STATISTICAL ANALYSIS PLAN

Study Title:	A Randomized, Double-blind, Placebo-controlled Phase 2 Study to Evaluate the Effect of Filgotinib on Semen Parameters in Adult Males with Active Rheumatoid Arthritis, Psoriatic Arthritis, Ankylosing Spondylitis or Non-radiographic Axial Spondyloarthritis
Name of Test Drug:	Filgotinib
Study Number:	GLPG0634-CL-227
Protocol Version (Date):	Original: 05 February 2019 Amendment 1: May 26, 2020
Analysis Type:	Interims (Week 26 and Week 26 plus Reversibility) and Final Analysis
Analysis Plan Version:	Version 1.0
Analysis Plan Date:	12 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
AS	ankylosing spondylitis
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
BASDAI	Bath Ankylosing Spondylitis Disease Activity Index
BLQ	below the limit of quantitation
BMI	body mass index
CDAI	Clinical Disease Activity Index
CI	confidence interval
COVID-19	Coronavirus Disease 2019
csDMARD	conventional synthetic disease modifying anti-rheumatic drug
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
CVEAC	Cardiovascular Event Adjudication Committee
DAPSA	Disease Activity in Psoriatic Arthritis
DMC	Data Monitoring Committee
DVT	deep vein thrombosis
EAIR	exposure-adjusted incidence rate
ECG	electrocardiogram
eCRF	electronic case report form
ET	early termination
EXT	extension phase
FDA	Food and Drug Administration
FSH	follicle-stimulating hormone
FU	follow-up
GLPG	Galapagos, NV
HLGT	high-level group term
HLT	high-level term
hsCRP	high-sensitivity C-Reactive Protein
IC	inclusion criteria
ID	Identification
IPD	important protocol deviation
ITT	Intent-to-Treat
IUT	Internal Unblinded Team
IWRS	interactive web response system
JAK	Janus Kinase

LH	luteinizing hormone
LLT	low level term
LOQ	limit of quantitation
M	Million
MACE	major adverse cardiac events
MAR	missing at random
MedDRA	Medical Dictionary for Regulatory Affairs
MH	Mantel-Haenszel
MI	multiple imputation, myocardial infarction
MP	Monitoring Phase
MST	MedDRA search term
MTX	Methotrexate
nrAxSpA	non-radiographic axial spondyloarthritis
OI	opportunistic infection
PE	pulmonary embolism
PhGADA	Physician's Global Assessment of Disease Activity
PK	pharmacokinetics
PsA	psoriatic arthritis
PT	preferred term
PTM	placebo-to-match
PYE	patient-years of exposure
Q1	first quartile
Q3	third quartile
RA	rheumatoid arthritis
SAE	serious adverse event
SAP	Statistical Analysis Plan
SD	standard deviation
SE	standard error
SMQ	standardized MedDRA query
SOC	system organ class
TEAEs	treatment emergent adverse events
TFLs	tables, figures and listings
ULN	upper limit of normal
US	United States
VAS	visual analogue scale
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 GLPG0634-CL-227 [MANTA-RAy]. This SAP is based on Amendment 1 of the protocol dated 26 May 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 Week13; and evaluable subjects from Study GLPG0634-CL-227 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 rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS) or non-radiographic axial spondyloarthritis (nrAxSpA) who have had an inadequate response to prior arthritis therapy as outlined in the inclusion criteria, have an adverse prognosis of their condition per the investigator, and who may benefit from treatment with a Janus Kinase (JAK)-inhibitor. Subjects may be on protocol-specified concomitant therapy.

Up to 250 males were planned to be randomized in a 1:1 ratio to receive filgotinib 200 mg or placebo-to-match (PTM) once daily for 13 weeks, after which an extension phase was to start. Subjects who met a pre-specified sperm decrease threshold (eg, ≥ 50% decrease in sperm concentration or motility or morphology) at any time during the study were to be followed off-study drug for reversibility in the Monitoring Phase.

Randomization was stratified according to the type of rheumatologic condition (RA or Spondyloarthritis [PsA, AS, nrAxSpA]), by concurrent use of methotrexate ([MTX], yes or no), and by sperm concentration (mean of 2 evaluable samples) collected at the screening visit according to the following strata:

- 15 to 25 million [M]/mL
- > 25 to 50 M/mL
- > 50 M/mL

The main inclusion criteria relevant to this study were:

- Male subjects between the ages 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 have met the following minimum criteria (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% (morphology using World Health Organization [WHO] 1992 criteria {World Health Organization (WHO) 1992}; all other parameters using WHO 2010 criteria {Cooper 2010, World Health Organization (WHO) 2010})

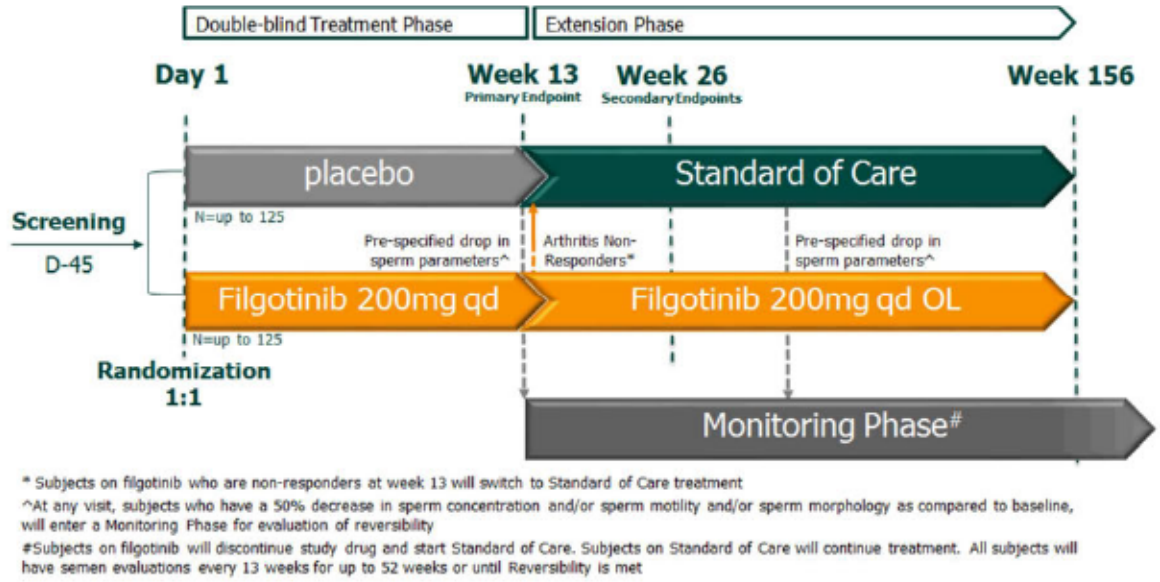
The main exclusion criterion relevant to this study was:

- 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 the Study Protocol. Male subjects with a rheumatologic condition who provided written informed consent were screened to determine eligibility. The Screening period was up to 45 days prior to randomization and could have been extended in consultation with the medical monitor.

The assessments planned to be performed at each visit are detailed in the study procedures table (Appendix 1). A schematic of the study design is provided below.

Figure 1-1. Study Schema



There are 3 distinct parts to the study:

1) Double-blind Treatment Phase (Day 1 through Week 13 study visit)

2) Extension Phase (After Week 13 Study Visit and up to Week 156)

3) Monitoring Phase (up to 52 weeks)

After fulfilling all selection criteria, subjects were to be randomized to filgotinib 200 mg or PTM in a 1:1 ratio at Day 1 and enter the Double-blind Treatment Phase. Based on the outcomes of the Week 13 assessments (ie, individual subject's response of the underlying rheumatic condition to the assigned treatment and/or observed changes in sperm parameters during the Double-blind Treatment Phase), subjects may enter the Extension or Monitoring Phase.

Double-blind Treatment Phase (Day 1 through Week 13 Study Visit)

During the Double-blind Treatment Phase, all subjects received blinded study drug for 13 weeks starting from the Day 1/Randomization Study Visit. At the Week 13 study visit, sperm parameters were evaluated to determine whether the subject had met any of the pre-specified sperm decrease thresholds (eg, ≥ 50% decrease in sperm concentration and/or motility and/or morphology). In addition, arthritis response status (Arthritis Responder [defined as improvement in physician's global assessment of disease activity (PhGADA) of at least 20% compared to Baseline (Day 1) at the specified assessment time] vs Arthritis Non-responder [defined as does not meet Arthritis Responder definition]) was determined. Based on these assessments, subjects were assigned to one of the following:

- Subjects whose semen parameters met any of the pre-specified sperm decrease thresholds, regardless of arthritis response status, discontinued study drug and entered the Monitoring Phase.

- Subjects who were arthritis non-responders and whose semen parameters did not meet any of the pre-specified sperm decrease thresholds discontinued blinded study drug and started Standard of Care treatment during the Extension Phase.

- Subjects who were Arthritis Responders and whose semen parameters did not meet any of the pre-specified sperm decrease thresholds were unblinded by the interactive web response system (IWRS) and entered the Extension Phase. Subjects on filgotinib were convert to open-label filgotinib and subjects who received placebo switched to Standard of Care treatment.

Extension Phase (After Week 13 through Week 156 Study Visit)

In the Extension Phase, subjects will receive open-label treatment (filgotinib or Standard of Care) and have a study visit every 13 weeks that includes semen analysis. The first scheduled visit of the Extension Phase occurs at Week 26.

In case semen parameters at any visit (including Week 156) meet any of the pre-specified sperm decrease thresholds, subjects will enter the Monitoring Phase.

During the Extension Phase, Standard of Care and background treatment concomitant to filgotinib may be optimized at the discretion of the investigator, taking into consideration the protocol restrictions (see Sections 5.3 and 5.4 of the study protocol).

In case a subject needs to be treated with a prohibited medication (see Table 5.1 from the study protocol), for instance to control their rheumatic condition, the subject will discontinue the study and complete an Early Termination (ET) visit (including semen collection). For subjects on filgotinib, this will be followed by a safety follow-up visit 30 days after last study drug dose.

All subjects who did not enter the Monitoring Phase will continue the study until the Week 156 visit.

Monitoring Phase (up to 52 weeks)

Subjects whose semen parameters meet any of the pre-specified sperm decrease thresholds will switch to the Monitoring Phase immediately. Subjects on double-blind treatment or open-label filgotinib who enter the Monitoring Phase will discontinue study drug and start Standard of Care. These subjects will also have a safety follow-up visit 30 days after the last study drug dose. Subjects on Standard of Care in the extension phase will continue taking a Standard of Care regimen in the Monitoring Phase.

All subjects who enter the Monitoring Phase will undergo semen evaluations every 13 weeks starting from their entry into the Monitoring Phase, for up to 52 weeks or until reversibility (defined as a return to > 0.5 x [mean of 2 evaluable samples collected at screening visit] for all sperm parameter[s] qualifying the subject for the monitoring phase—based on 2 evaluable samples collected at the monitoring phase visit) is met, whichever is achieved first.

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 inflammatory bowel disease (GS-US-418-4279 [MANTA]) 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 ≥ 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 the Sponsor 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] plus GLPG0634-CL-227 [MANTA-RAy] were reviewed after approximately 100 subjects completed 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 the Sponsor regarding the future conduct of the study, including possible early study termination, the Sponsor 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 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 Sponsor procedures.

2.3. Week 26 Analysis

The unblinded Week 26 analysis of Study GLPG0634-CL-227 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 GLPG0634-CL-227 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 GLPG0634-CL-227 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 plus Reversibility Analysis for Pre-specified Sperm Decreases Up to Week 26

A separate unblinded Week 26 analysis of Study GLPG0634-CL-227 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 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 GLPG0634-CL-227 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 156 and did not meet pre-specified sperm decrease threshold [≥ 50% decrease in sperm concentration and/or motility and/or morphology]; met pre-specified sperm decrease threshold(s) 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 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, double-blind treatment and responder/non-responder status at Week 13, or treatment sequence will be presented in the listings, as appropriate. Age, sex, race, ethnicity, start and end date of the double-blind study drug and open-label filgotinib or start and end date of extension phase (for subjects who were switched to standard of care in the extension phase) 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 (≥ 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 collected ≥ 48 hours and ≤ 7 days after the previous ejaculation date/time; and collected within 14 days of the previous chronologic semen sample collection at the visit when the sample is not the first collection at the visit. Three subjects (10545-60695, 12858-60706, 15922-60684) had a third semen sample collected at Week 13 that was 35, 22, and 24 days, respectively, after the second semen sample (the first 2 semen samples were collected within 14 days) due to delays in analysis time. These samples will be considered to be an exception and will be "eligible for mean calculation".

The Week 26 Semen Analysis Set includes all subjects treated (≥ 1 dose of open-label filgotinib or standard of care in the extension phase) who have 2 evaluable samples at baseline and at Week 26.

The At/After Week 39 Semen Analysis Set includes all subjects treated (≥ 1 dose of open-label filgotinib or standard of care in the extension phase) who have 2 evaluable samples at baseline and at ≥ 1 post baseline analysis visit window on/after Week 39. The first of 2 evaluable semen samples for the postbaseline timepoint will be utilized to assign the postbaseline analysis visit window.

These are the primary analysis sets for all semen and hormone-related parameters.

The primary analysis of semen data for 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 (see Section 6.4 of this SAP) will also be performed for 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 for safety during the double-blind phase of the study.

The Extension Phase Safety Analysis Set includes all subjects who took ≥ 1 dose of open-label filgotinib or standard of care in the extension phase of the study. This is the analysis set for safety during the extension phase of the study.

3.1.4. Pharmacokinetic Analysis Set

The Pharmacokinetic (PK) Analysis Set includes all randomized subjects who took ≥ 1 dose of filgotinib 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 phase, data will be summarized by double-blind treatment group (ie, filgotinib 200 mg and placebo).
- For the Week 26 visit, semen and hormone data will be summarized by treatment sequence and Rheumatic Disease Responder/Non-responder status at Week 13 (ie, filgotinib responders who switch to open-label filgotinib, filgotinib non-responders at Week 13 who switch to standard of care, placebo responders who switch to standard of care, and placebo non-responders who switch to standard of care).
- Semen and hormone data collected at/after Week 39 in the extension phase will be summarized by treatment sequence (ie, double-blind filgotinib to open-label filgotinib, double-blind filgotinib to standard of care, and placebo to standard of care) for subjects with baseline and $\geq$ 1 visit at/after Week 39.
- Subjects with a decrease of $\geq$ 50% in sperm concentration and/or motility and/or morphology from baseline while on double-blind treatment or open-label filgotinib will have study drug stopped and an investigator selected standard of care treatment that meets protocol-specified criteria started. Subjects meeting a sperm decrease while on standard of care in extension phase will continue on a standard of care regimen. Semen samples will be collected off-study drug/control (eg, standard of care in extension phase) 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 parameter decreases at Week 13 (summarized by double-blind treatment group); for subjects meeting pre-specified sperm parameter decrease thresholds at Week 26 (summarized by treatment sequence and rheumatic disease responder/ non-responder status at Week 13) and at or after Week 39 (summarized by treatment sequence).

For the safety analyses, subjects will be grouped by double-blind treatment group (ie, filgotinib 200 mg or placebo) for safety events reported during the double-blind phase of the study.

Safety events reported during the extension phase will be summarized by treatment sequence (double-blind filgotinib to open-label filgotinib, double-blind filgotinib to standard of care, and placebo to standard of care).

3.3. Strata and Covariates

Subjects will be randomly assigned to treatment groups via the IWRS in a 1:1 ratio using a stratified randomization schedule. Stratification will be based on the following variables:

- Type of rheumatologic condition (RA, Spondyloarthritis [PsA, AS, nrAxSpA])
- Concurrent use of MTX (Yes, No)
- Sperm concentration (mean of 2 evaluable samples) measured at the Screening visit: 1) 15 to 25 M/mL; 2) > 25 to 50 M/mL; 3) > 50 M/mL

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 eCRF, “01 July” will be used for the unknown birth day and month for purposes 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 Double-Blind Filgotinib/Placebo is defined as the date when subjects take their first dose of double-blind filgotinib/placebo as recorded on the Study Drug Administration eCRF.

The Last Dosing Date of Double-Blind Filgotinib/Placebo is defined as the date when subjects take their last dose of double-blind filgotinib/placebo as recorded in the Study Drug Administration eCRF.

The Last Dosing Date of Any Filgotinib is defined as the date when a subject takes their last dose of double-blind or open-label filgotinib as recorded in the Study Drug Administration eCRF.

The Start Date of Extension Phase is defined as the first dose date of open-label filgotinib (for rheumatic disease responders randomized to filgotinib who continued on open-label filgotinib in extension phase) from the study drug administration eCRF or last dose date of double-blind study drug +1 day for subjects who have a non-missing standard of care start date in the extension phase of the study. In some countries, it takes time to obtain a new standard of care drug for rheumatic disease when subjects undergo a medication switch (especially for biologics), so the "standard of care" is to continue background medications alone or in rare cases, to take no rheumatic disease medication until the new regimen is available.

The End Date of Extension Phase is defined as last dose date of open-label filgotinib as recorded on the Study Drug Administration eCRF for subjects on open-label filgotinib in the Extension Phase. The end date for standard of care in extension phase will be 1 day prior to the date the subject went into the Monitoring Phase (ie, date the site was notified of subject meeting a prespecified sperm decrease threshold). Subjects who do not participate in Monitoring Phase but have a Study Completion eCRF that states that the subject has either completed study or prematurely discontinued study will have the last date of standard of care estimated using the last of the following dates: last Covance lab date, last eCRF visit date, date of last sperm sample collection, and death date (if subject died).

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 – On-Filgotinib/Placebo Study Day 1

Follow-up (FU) Day will be calculated from last study drug (double-blind study drug or open-label filgotinib) dosing date or end date of extension phase for subjects on standard of care.

The FU day is derived as Assessment Date – Last Dose Date of study drug/end date of extension phase for subjects on standard of care

Therefore, Follow-up Day 1 is the first day after permanent discontinuation of all study drugs or end of Extension Phase for subjects on standard of care in extension phase.

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

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 parameters is used to determine whether subjects have met a sperm parameter decrease threshold (ie, ≥ 50% decrease from baseline in sperm concentration and/or motility and/or morphology).

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 urinalysis, PhGADA, and lipid profiles are provided in Table 3-1. The analysis visit windows for high-sensitivity C-reactive protein (hsCRP), 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 semen parameters are provided in Table 3-4 and Table 3-5, respectively.

Visit Window Rules for Double-Blind Phase Displays

Data collected during the double-blind phase of the study will have visit windows assigned based on study day calculated from first dose date of double-blind study drug.

Visit Window Rules for Extension Phase Displays

Data collected during the extension phase of the study will have visit windows assigned based on study day calculated from first dose date of double-blind study drug.

Visit Window Rules for FU-4 Visit (labs [including hormones], vital signs, height, weight, BMI)

Prior to protocol amendment 1, all subjects were to have a 30-day safety follow-up visit after their last dose date of any filgotinib (last of double-blind/open-label) or double-blind placebo. Under protocol amendment 1, responders assigned to placebo will no longer be required to attend a safety follow-up visit after Week 13. For subjects in the filgotinib 200 mg to OL filgotinib group, the follow-up visit will occur after the extension phase (ie, FU days); for rheumatic disease nonresponders continuing on standard of care in extension phase, the follow-up visit will be windowed based on days in extension phase (ie EXT days) since subjects will be off of study drug but on standard of care in extension phase. Subjects who meet a sperm decrease threshold at Week 13 or prematurely discontinue from double-blind study drug will have the Follow-Up Week 4 visit calculated from the last dose date of double-blind study drug.

Visit Window Rules for Monitoring Phase Sperm Parameters

Semen and hormone data collected in the Monitoring Phase will have visit windows calculated from the maximum of last dose date of any study drug (double-blind drug or open-label filgotinib) or date subject went into Monitoring Phase (the date the site receives notification of the subject meeting a sperm decrease threshold) for subjects on standard of care in extension phase.

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

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	≥ 1093
FU-4^	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug, otherwise + 30 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug, otherwise + 8 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 35 Days

*PhGADA is only assessed at the Baseline and Week 13 analysis visit window. PhGADA is only collected to determine whether the subject is a responder or non-responder at the end of the double-blind phase. # Lipid profile is collected at baseline, and at Weeks 13, 26, 39, 65, 91, 117, and 143.
^ FU-4 visit collected based on days from last dose of open-label filgotinib (filgotinib 200 mg to OL filgotinib); EXT days for all (prior to Amendment 1) subjects / only non-responders (after Amendment 1) who switched to standard of care in extension phase; and from last dose date of double-blind study drug for those meeting sperm decrease threshold at Week 13 or prematurely discontinued in/at end of double-blind phase. Protocol specified an upper bound of + 5 days. FU-4 visit window is applicable for urinalysis.

Table 3-2. Analysis Visit Windows for hs-CRP, 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	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	≥ 1093
FU-4*	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 30 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 8 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 35 Days

Note: Data will be included for "on-treatment" visits up to the last dose date of standard of care in the extension phase (if non-missing) or up to last dose date of any study drug + 7 days.
*FU-4 visit collected based on days from last dose of open-label filgotinib (filgotinib 200 mg to OL filgotinib); EXT days for all (prior to Amendment 1) subjects / only non-responders (after Amendment 1) who switched to standard of care in extension phase; and from last dose date of double-blind study drug for those meeting sperm decrease threshold at Week 13 or prematurely discontinued in/at end of double-blind phase. Protocol specified an upper bound of + 5 days. FU-4 visit window is applicable for hematology, chemistry, vitals, and weight.

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	≥ 1093
FU-4*	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 30 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 8 Days	Last Dose Date OL filgotinib (filgo to OL filgo); last dose date DB drug*, otherwise + 35 Days

Note: Data will be included for "on-treatment" visits up to the last dose date of standard of care in the extension phase (if non-missing) or up to last dose date of any study drug + 7 days.
*FU-4 visit collected based on days from last dose of open-label filgotinib (filgotinib 200 mg to OL filgotinib); off study drug but in EXT phase for all (prior to Amendment 1) subjects / only non-responders (after Amendment 1) who switched to standard of care in extension phase; and from last dose date of double-blind study drug for those meeting sperm decrease threshold at Week 13 or prematurely discontinued in/at end of double-blind phase. Protocol specified an upper bound of + 5 days.

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	≥ 1093

Note: Primary analysis includes data up to first dose date of open-label filgotinib/first date in extension phase (if non-missing) or to last dose date of double-blind drug + 7 days (for those not continuing in Extension Phase).
Secondary analysis includes all "on-treatment" data up to the last dose date in extension phase (if non-missing) or to 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/End of Extension Phase for those on Standard of Care in Extension Phase)

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/end date in Extension Phase (date site receives confirmation of subject meeting sperm parameter decrease threshold) for subjects taking standard of care in extension phase who met pre-specified sperm decrease thresholds.
FU Day = visit date – last dose date of any study drug/last date in extension phase.

Investigator Assessment of Electrocardiogram (ECG) Visit Assignment

Nominal visits from the clinical database will be used to assign timepoints for the investigator's assessment of ECG. Baseline will use the nominal visit "Screening"; Week 13 or Early Termination will use a nominal visit of "Week 13" or "Early termination " if subject prematurely discontinued double-blind study drug; and LTE Week 156 or Early termination visit for subjects in the filgotinib 200 mg to open-label filgotinib 200 mg group will use nominal visits of "Extension Week 156" or "Early termination" for subjects who early terminated after starting open-label filgotinib (per protocol Amendment 1 subjects taking standard of care in extension phase are no longer required to complete an ECG at Week 156 or early termination).

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 supportive 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 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 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 analyzed 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 rheumatologic condition 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 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 sperm 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, 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 are counted only 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])
  - — Switched to Open-Label Filgotinib after Week 13 (Responder on double-blind filgotinib, subject did not meet pre-specified sperm decrease threshold[s])

— Switched to Standard of Care after Week 13 (Non-responder on double-blind filgotinib; or randomized to placebo in double-blind phase, subject did not meet pre-specified sperm decrease threshold[s])

— Has Week 13 Semen Sample/Terminated from Study. Subject ID 15915-60609 completed double-blind study drug at Week 13 and had 2 evaluable semen samples collected. He subsequently early terminated from study and did not enter extension phase of the study.

- Did not complete blinded study drug through Week 13 with reasons for premature discontinuation of study drug

Extension Phase Open-Label Filgotinib Study Drug Completion Status

- Completed Extension Phase of study (on open-label filgotinib)
- Ongoing in Extension Phase and on open-label filgotinib (for interim analysis only, includes subjects who completed 13 weeks of double-blind study drug and are ongoing on open-label filgotinib in the extension phase at the time of interim analysis)
- Did not complete Extension Phase of study (on open-label filgotinib) and reason for premature discontinuation

Extension Phase Standard of Care Completion Status

- Completed Extension Phase of study on standard of care
- Ongoing in Extension Phase of study on standard of care (for interim analysis only, includes subjects who completed the Double-blind Phase and are ongoing on Standard of Care in the extension phase at the time of interim analysis)
- Did not complete Extension Phase of study on standard of care and reason for premature discontinuation (from study completion form)

Note for programming: Subjects who started standard of care in extension phase and met a sperm decrease threshold at a visit other than "LT Week 156" will be considered to be prematurely discontinued from this phase for "Pre-Specified Decrease in Sperm Parameters". For subjects who started standard of care in extension phase, did not go into the monitoring phase, and did not complete study, the reason for premature discontinuation of study will be used for reason for early termination of extension phase. Subjects who started standard of care in extension phase and completed the study (without going into monitoring phase) will be considered to have completed the extension phase of the study on standard of care.

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). Subjects who were rescreened will be summarized based on their last chronologic screening attempt. Screens that did not 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) #7 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 #7 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]. 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 #7 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 standard of care in extension phase; and subjects who took double-blind filgotinib, were a non-responder at Week 13, and switched to standard of care in extension phase, the duration in the extension phase will be calculated as last date in extension phase minus first date in extension phase + 1 divided by 7 (eg, reported in weeks to 1 decimal place).

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 will be summarized using descriptive statistics and using the number (ie, cumulative counts) and percentage of subjects exposed through the following time periods: Day 1, Week 2 (Day 14), Week 4 (Day 28), Week 8 (Day 56), and ≥ Week 13 (Day 91).

The total duration of exposure to open-label filgotinib or weeks in the extension phase for those switching to SOC 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 13 (Day 91), Week 26 (Day 182), Week 39 (Day 273), and up to Week 143 (Day 1001) at 13-week intervals (as appropriate). Total duration of exposure to any filgotinib will follow the same categories as extension phase displays with the addition of Week 156 (Day 1092).

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

4.2.2. Adherence to Study Drug

The on-treatment adherence will be calculated for double-blind filgotinib 200 mg/PTM (tablets), through the double-blind phase of the study 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 during the double-blind phase will be determined by the data collected on the drug accountability eCRF using the following formula:

Total Number of Tablets Administered

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

All open-label filgotinib bottles dispensed will be excluded from the calculations.

If a bottle of double-blind study drug is dispensed and the bottle is returned empty, then the number of tablets returned will be entered as zero. If a bottle of double-blind study drug 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 study drug during the double-blind phase of the study 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 phase based on the study drug regimen.

The level of on-treatment adherence during the double-blind phase of the study 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).

Descriptive statistics for the level of on-treatment adherence during the double-blind phase of the study with the number and percentage of subjects belonging to adherence categories (eg, < 80%, ≥ 80% to < 90%, ≥ 90%) will be provided by treatment group for the Double-Blind Semen Analysis Set.

No formal statistical testing is planned.

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

A listing of subjects who had interruptions in study drug dosing as recorded on the study drug administration eCRF will be produced.

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 deviations related to COVID-19, and 2) any non-important protocol deviations 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 double-blind treatment group and overall using descriptive statistics for age, and using number and percentage of subjects for sex, race and ethnicity for the Safety/ITT Semen Analysis Sets and Semen Analysis Set. Separate displays will present data by treatment sequence and overall for the Extension Phase Safety Analysis Set and the at/after Week 39 Semen Analysis Set. Displays for the Week 26 Semen Analysis Set will be by treatment sequence and Rheumatic Disease Responder/Non-Responder status at Week 13.

Age is calculated in years at the date of first double-blind 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. Other Baseline Characteristics

Other baseline characteristics include:

- Body weight (in kg)
- Height (in cm)
- Body mass index (BMI; in kg/m^2)
- Type of rheumatologic condition (RA, PsA, AS, nrAxSpA)
- Duration of inflammatory condition (in years)
- Physician's Global Assessment of Disease Activity (PhGADA) for all subjects
- Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) for subjects with AS or nrAxSpA, continuous and for categories (< 4, ≥ 4)
- Clinical Disease Activity Index (CDAI) for subject with RA, continuous and for categories (remission, low, moderate, and high) as described below
- Disease Activity in Psoriatic Arthritis (DAPSA) for subjects with PsA continuous and for categories (remission, low, moderate, and high) as described below
- High sensitivity C-Reactive Protein (hsCRP, mg/L)

- 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 (0, 1-2, 3-4, ≥ 5)
- Prior use of csDMARDs (any including MTX)
- Concurrent MTX use status (yes, no) with subcategories for MTX dose (< 20 mg, ≥ 20 mg)
- Sperm concentration measured (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) –WHO 1992 criteria
- Luteinizing Hormone (mIU/mL)
- Follicle-Stimulating Hormone (mIU/mL)
- Inhibin B (pg/mL)
- Total Testosterone (ng/dL)

Categories for CDAI (for RA) and DAPSA (for PsA) that will be summarized include:

Categories for Disease Activity	Clinical Disease Activity Index (CDAI for RA)	Disease Activity for Psoriatic Arthritis (DAPSA for PsA)
Remission	0 to 2.8	0 to 4
Low	2.9 to 10	5 to 14
Moderate	10.1 to 22.0	15 to 28
High	22.1 to 76.0	> 28

Duration of rheumatologic condition is calculated as:

([first dose date] – [date of initial diagnosis] + 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 for the double-blind phase of the study 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 present data by treatment sequence in extension phase and overall for the Extension Phase Safety Analysis Set and the at/after Week 39 Semen Analysis Set. Displays for the Week 26 Semen Analysis Set will be by treatment sequence and Rheumatic Disease Responder/Non-Responder status at Week 13.

No formal statistical testing is planned.

A by-subject listing of other 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 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 #7. Semen parameters other than sperm morphology use WHO 2010 criteria {World Health Organization (WHO) 2010}; while sperm morphology used WHO 1992 criteria {World Health Organization (WHO) 1992}. Semen parameter eligibility criteria are provided in Table 6-2.

Table 6-2. Semen Parameter Eligibility Criteria

Parameter	Normal 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}.

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 discontinuation 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 and do not continue to Extension Phase.
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 M/mL, > 50 M/mL], concurrent MTX use [yes, no]) will be used to estimate the difference in proportions (filgotinib – placebo) and corresponding 95% CI {Koch 1989}. Calculation for difference in proportions and 95% CI 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 RA.
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 discontinuation 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 and do not continue to Extension Phase.
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), MTX use (yes, no). 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), MTX use (yes, no) 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 semen 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 Rheumatic Disease disorder (RA, Spondyloarthritis [AS, PsA, nrAxSpA])
- Sperm concentration (mean of 2 evaluable samples) measured at Screening visit: (15 - 50 M/mL, > 50 M/mL)
- Concurrent MTX use (yes, no)

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 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	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. If a measurement for a specific semen parameter [Subject ID 15858-60640, motility] 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 due to early termination 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, baseline sperm concentration category (15 - 50 M/mL, > 50 M/mL), and concurrent MTX use (yes, no) 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], and concurrent MTX use [yes, no]). 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. If a measurement for a specific semen parameter [Subject ID 15858-60640, motility] 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 2013}.
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. Week 26 by Treatment Sequence and Responder/Non-Responder at Week 13 Displays

Displays will be presented at the Week 26 visit for the following 4 groups:

- filgotinib responder at Week 13 switched to open-label filgotinib
- filgotinib non-responder at Week 13 switched to standard of care
- placebo responder at Week 13 switched to standard of care
- placebo non-responder at Week 13 switched to standard of care

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

For all secondary continuous endpoints, 8-point summary statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) by treatment sequence and responder/non-responder status at Week 13 will be provided for the value at the visit, change from baseline, and percentage change from baseline in sperm/semen parameters at Week 26. A 95% distribution-free CI on the median change and percentage change from baseline at Week 26 in all sperm/semen parameters will be constructed by treatment sequence and responder/non-responder status at Week 13 using the CIPCTLDF option in PROC UNIVARIATE in SAS.

A cumulative distribution plot for percentage change from baseline in sperm concentration at Week 26 will be constructed by treatment sequence and responder/non-responder status at Week 13 for those who remain on study drug at the Week 26 visit.

6.5.2. Treatment Sequence Displays at/after Week 39

Displays will be presented by analysis visit at Week 39 and thereafter in the Extension Phase for the following 3 treatment sequences:

- filgotinib to open-label filgotinib
- filgotinib to standard of care
- placebo to standard of care

The number and proportion of subjects with a ≥ 50% decrease from baseline in sperm concentration will be presented by analysis visit.

For all secondary continuous endpoints, 8-point summary statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) by treatment sequence will be provided for the value at the visit, change from baseline, and percentage change from baseline in sperm/semen parameters.

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

Subjects who met a pre-specified sperm decrease threshold (a ≥ 50% decrease from baseline 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 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 Week 13 will be displayed by double-blind treatment group for the Semen Analysis Set.

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 will be displayed at Week 26 for the Week 26 Semen Analysis Set; and at or after the Week 39 visit for the At/After Week 39 Semen Analysis Set. Presentations will be displayed by treatment sequence and responder/non-responder at Week 13 and treatment sequence, respectively.

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 by treatment sequence and responder/non-responder at Week 13; and at/after Week 39 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 by monitoring visit number and double-blind treatment group for subjects with a ≥ 50% decrease at/prior to Week 13.

Table 6-3. Hypothetical Proportion of Subjects Who Met Reversibility Criteria for Sperm Concentration by Treatment Group and Monitoring Phase Visit (Subjects with a ≥ 50% Decrease in Sperm Concentration at/prior to 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 treatment (double-blind treatment group for Baseline, Week 4 and Week 13; by treatment sequence and responder/non-responder status at Week 13 for Baseline and Week 26) will be provided for the value at the visit, change from baseline, and percentage change from baseline in hormone parameters for the Semen Analysis Set and Week 26 Semen Analysis Set, respectively. A 95% distribution-free CI on the median change and percentage change from baseline in all hormone parameters will be constructed by double-blind treatment group at Weeks 4 and 13; and by double-blind treatment and responder/non-responder status at Week 13 for the Week 26 visit 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. Models will adjust for the baseline hormone value, concurrent MTX use (yes, no), and sperm concentration strata (15 - 50 M/mL, > 50 M/mL). The CI=RANK option will be specified in PROC QUANTREG in SAS.

For Week 26, the change and percentage change from baseline in hormone parameters will be displayed by treatment sequence, responder/non-responder status at Week 13, and visit for subjects in the Week 26 Semen Analysis Set.

For visits at/after Week 39, the change and percentage change from baseline in hormone parameters will be displayed by treatment sequence and visit for subjects in the At/After Week 39 Semen Analysis Set.

Shift tables from baseline to Week 4 and Week 13 for LH, FSH, inhibin B, and total testosterone will be summarized by double-blind treatment group for the Semen Analysis Set. Shift tables from baseline to Week 26 for LH, FSH, inhibin B, and total testosterone will be summarized by treatment sequence and responder/non-responder status at Week 13 for the Week 26 analysis visit (using the closest value to target days) and the Week 26 Semen Analysis Set. The shift tables include shift analysis for hormone parameters below, within, and above the central laboratory defined normal range at baseline to below, within and above the normal range at Week 4, Week 13 and Week 26.

No sensitivity analysis will be performed. No formal testing is planned.

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 continuous semen parameter 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.

Disease type was not included as a stratification factor 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 only a few RA 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

Physician Global Assessment of Disease Activity (PhGADA) was used to determine whether a subject was an "Arthritis Responder" (defined as an improvement in PhGADA of at least 20% compared to baseline at the specified assessment time point) or "Arthritis Non-responder" (defined as a subject who does not fulfill the definition of responder at the specified assessment timepoint). The physician (who had access to disease assessments described in Protocol Sections 6.18 through 6.23 [as appropriate]) drew a perpendicular line on a 100-millimeter (mm) visual analogue scale (VAS) representing the subject's disease severity and measured the distance (in mm) between the "no disease activity" anchor (at 0) and the mark on the 100-mm line. The PhGADA ranged from 0-100.

A listing of all observed data based on the All Randomized Analysis Set will display the PhGADA at Screening, Day 1, and Week 13; and the percent change from baseline in PhGADA and associated arthritis responder/non-responder status at Week 13.

8. SAFETY ANALYSES

Safety analyses will be conducted on the Safety Analysis Set and Extension Phase Safety Analysis Set, defined in Section 3.1.3.

Safety data during the double-blind phase will be summarized based on actual treatment while the subject is on double-blind study drug. Subjects will be included in analysis after first dose date of double-blind study drug up to:

- First dose date of open-label filgotinib or first date in extension phase (minus 1 day for adverse events) for subjects who receive open label filgotinib or standard of care in extension phase
- Last dose date of double-blind study drug + 30 days (subjects meeting pre-specified sperm decrease thresholds at/prior to Week 13 who were switched to Monitoring Phase to assess reversibility of all sperm parameters meeting decrease thresholds; subjects permanently discontinued from study drug at/prior to Week 13)

Safety data during the extension phase will be summarized based on actual treatment sequence as described in Section 3.2 of this SAP. Safety data collected on (AEs only) or after first dose date of open-label filgotinib/first date in extension phase to:

- Last dose date of open-label filgotinib/standard of care in extension phase + 30 days will be included for subjects who have stopped or completed all study drug/SOC in the extension phase
- All available data (after start of extension phase) at time of data cut for subjects who are ongoing on open-label filgotinib/standard of care in the extension phase at the time of an interim analysis

8.1. Adverse Events and Deaths

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." For the extension phase, AEs marked as "Related to SOC" will be considered as related for subjects who switch to SOC in the extension phase. 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) during the double-blind phase are defined as 1 or both of the following:

- Any AEs with an onset date on or after the double-blind study drug start date through first dose date of open-label filgotinib/start date in extension phase minus 1 day (if subject enrolled in extension phase) or to last dose date of double-blind study drug + 30 days for subjects who permanently discontinue double-blind drug and do not enroll into extension phase (including subjects who meet a sperm decrease threshold who stop study drug and continue to be followed in the monitoring phase for reversibility)

- Any AEs leading to premature discontinuation of double-blind study drug

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

- Any AEs with an onset date on or after first dose date of open-label filgotinib/first dose date in extension phase to last dose date of open-label filgotinib + 30 days for subjects who permanently discontinue open-label filgotinib (including subjects who meet a sperm decrease threshold who stop study drug and continue to be followed in the monitoring phase for reversibility) or all AEs for subjects who are ongoing on open-label filgotinib/standard of care in extension phase at the time of an interim analysis

- Any AEs leading to premature discontinuation of open-label filgotinib study drug

For example, for a subject who took double-blind placebo and switched to standard of care in the extension phase, TEAEs with an onset date on/after first dose date of double-blind filgotinib and prior to first date in extension phase will be TE for double-blind filgotinib (summarized with double-blind phase); from first date in extension phase up to last dose date of standard of care in extension phase + 30 days will be TE for standard of care (summarized by treatment sequence [double-blind filgotinib to standard of care] for the extension phase).

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. For the double-blind phase, 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 double-blind study drug, and
- The AE onset date is the same as or before the month and year (or year) of the extension phase open-label filgotinib or extension phase start date (if non-missing) or before the date corresponding to 30 days after the date of the last dose of double-blind study drug (if permanently discontinued from double-blind study drug)

For the extension phase, 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 open-label filgotinib/start date in the extension phase, and
- The AE onset date is the same as or before the month and year (or year) corresponding to 30 days after the date of the last dose of open-label filgotinib/last date in extension phase (if permanently discontinued from extension phase drug or entered monitoring phase for meeting a pre-specified sperm decrease threshold) or all events after first dose date of drug for extension phase if subject is ongoing in extension phase at time of interim analysis

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 double-blind study drug, will be considered to be treatment emergent for the double-blind phase. 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 open-label filgotinib/first date in extension phase will be considered treatment emergent for extension phase.

8.1.5.3. Analysis of Exposure-Adjusted Incidence Rate

The Exposure-Adjusted Incidence Rate (EAIR) for a TEAE 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 and is defined as:

For subjects with an event:

PYE = (Event start date – first date in DB phase +1) / 365.25

For subjects with no event:

PYE = (Calculation end date – first date in DB phase +1) / 365.25

where calculation end date for the double blind phase = the earliest date among

(last dosing date of double blind phase +30, first date in extension phase of the study -1 day, death date, cutoff date). If the last study drug dosing date is missing and the subject has stopped drug, the latest date among study drug end date, clinical visit date, and laboratory sample collection date that occurred during the double-blind phase will be used.

where calculation end date for the entire study display = the earliest date among

(last dosing date of OL filgotinib+ 30 days [subjects in the filgotinib to OL filgotinib sequence who permanently discontinued OL filgotinib); last dosing date of double-blind filgotinib [subjects in the filgotinib to standard of care group], last date on standard of care in extension phase [subjects in placebo to standard of care sequence who switched to monitoring phase or early terminated from study after starting extension phase], last dosing date of double-blind study drug + 30 for subjects who did not start extension phase, data cutoff date, death date).

If event start date during the double-blind phase is partially missing and the non-missing portion matches the first dosing date (month and/or year) of double-blind study drug; or if the date is completely missing, AE onset is imputed as the date of first dose of double-blind study drug. Otherwise it is imputed as the first day of the month (if only day is missing) or January 1st of the year (if day and month are missing).

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

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

To estimate the 95% CI for the EAIR difference between two treatment groups for events during the double-blind phase, the Method of Variance Estimates Recovery (MOVER)-type confidence interval described in {Li 2011} will be used:

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

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 the 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 (by double-blind treatment for double-blind phase; by treatment sequence for extension phase):

- 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 (NOTE: category applicable only for double-blind phase)

- TEAE that led to premature discontinuation of open-label filgotinib study drug (NOTE: category applicable only for extension phase)
- TEAE that led to temporary interruption of double-blind study drug (NOTE: category applicable only for double-blind phase)
- TEAE that led to temporary interruption of open-label filgotinib (NOTE: category applicable only for extension phase)
- TEAE that led to premature discontinuation of study
- All deaths observed in the study (if a subject received ≥ 1 dose of open-label filgotinib or standard of care in extension phase, the death will be summarized in extension phase; if the last study drug received was double-blind drug, the death will be summarized with the double blind phase)

The number and percentage of subjects who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and treatment (double-blind treatment for double-blind phase; treatment sequence for extension phase).

For other AEs described below, summaries will be provided by SOC, PT, and treatment (double-blind treatment for double-blind phase; treatment sequence for extension phase):

- 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
- All 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 temporary interruption of double-blind study drug (NOTE: table applicable only for double-blind phase)
- TEAEs leading to temporary interruption of open-label filgotinib study drug (NOTE: table applicable only for extension phase)

- TEAEs leading to premature discontinuation of double-blind study drug
  (NOTE: table applicable only for double-blind phase)
- TEAEs leading to premature discontinuation of open-label filgotinib study drug
  (NOTE: table applicable only for extension phase)
- 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 by PT only and treatment (double-blind treatment for double-blind phase; treatment sequence for extension phase), in descending order of total frequency for PT.

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 AEs
- 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

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 a MST list developed by Gilead
  - Hepatitis B or C infections, using a MST list developed by Gilead
- AEs of malignancies, excluding non-melanoma skin cancers, utilizing a MST list developed by Gilead
- AEs of non-melanoma skin cancers, utilizing a MST list developed by Gilead
- AEs of gastrointestinal perforation, utilizing a MST list developed by Gilead
- AEs of 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 SMQ
  - AEs of cerebrovascular events, utilizing the ischaemic central nervous system vascular conditions SMQ

For the double-blind phase, the number of subjects with TE AEIs and 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 by PT. The EAIR difference (Filgotinib 200 mg – Placebo) and corresponding 95% CIs for the difference between treatment groups (specified in section 8.1.5.3) will also be presented for the double-blind phase of the study.

For the overall study, the number of subjects with TE AEIs and 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 (filgotinib and placebo) by PT for each AEI category and separately for SAEs. Filgotinib group will include filgotinib to open-label filgotinib treatment sequence plus the filgotinib exposure for filgotinib to SOC treatment sequence and those randomized to filgotinib who did not continue to extension phase. Placebo will include the placebo to SOC treatment sequence and those randomized to placebo who did not continue to extension phase.

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 whether the event was positively adjudicated by CVEAC will be provided for each interim analysis and for the final analysis 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.

8.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data during the double-blind phase of the study will be provided by double-blind treatment group for the Safety Analysis Set. Summaries of laboratory data during the extension phase of the study will be provided by treatment sequence for the Extension Phase Safety Analysis Set.

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. Hemolized 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 (double-blind treatment group for double-blind phase and by treatment sequence of extension phase) 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 date of first dose of double-blind study drug. Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

Median (Q1, Q3) change from baseline at each timepoint for selected laboratory tests will be plotted using a line plot by treatment (double-blind treatment for double-blind phase and treatment sequence for extension phase) 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 during the double-blind phase are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point, up to the first dose date of open-label filgotinib/first date in extension phase (if non-missing) or to the date of last dose of double-blind study drug plus 30 days for subjects who permanently discontinued double-blind study drug (and did not continue to extension phase).
Treatment-emergent laboratory abnormalities during the extension phase are defined as values that increase at least 1 toxicity grade from baseline at any postbaseline time point after first dose date of open-label filgotinib/first date in extension phase to last dose date of open-label filgotinib/standard of care in extension plus 30 days or to the last available date in the database snapshot for subjects who were on extension phase treatment at the time of 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 during the double-blind phase are defined as values that increase from baseline by at least 3 toxicity grades at any postbaseline time point, during double-blind phase. Treatment-emergent marked laboratory abnormalities during the extension phase are defined as values that increase from baseline by at least 3 toxicity grades at any postbaseline time point during extension phase.

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 (double-blind treatment for double-blind phase; treatment sequence for extension phase). 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 double-blind phase summaries of laboratory abnormalities, the denominator is the number of subjects with ≥ 1 nonmissing postbaseline value during the double-blind phase. For the extension phase the denominator is the number of subjects with ≥ 1 nonmissing postbaseline value during the extension phase.

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 (by double-blind treatment during double-blind phase; by treatment sequence during extension phase) using the number and percentage of subjects with ≥ 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 during the double-blind phase for double-blind phase summaries and most severe postbaseline value during the extension phase for extension phase summaries. 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 ≥ 1 nonmissing postbaseline value of all relevant tests at the same postbaseline visit date for visits during double-blind phase; and the number of subjects with ≥ 1 nonmissing postbaseline value of all relevant tests at the same post baseline visit date for visits during the extension phase for the Extension Phase Safety Analysis Set.

A listing of subjects who met at least 1 of the above criteria will be provided.

8.3. Body Weight and Vital Signs

Descriptive statistics will be provided by treatment (double-blind treatment for double-blind phase; treatment sequence for extension phase) 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. 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/start date in the relevant phase of the study. 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. 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 that are continued 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 during the double-blind phase will be by double-blind treatment group for subjects in the Safety Analysis Set. Summaries during the extension phase will be by treatment sequence for subjects in the Extension Phase 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 baseline versus Week 13 or Early Termination from double-blind phase will be presented by double-blind treatment group for the Safety Analysis Set. The shift tables of ECG results for baseline vs. Extension Week 156 or Early Termination will be presented for subjects in the filgotinib 200 mg to open-label filgotinib 200 mg group for the Extension Phase Safety Analysis Set.

The investigator will rate the ECG based on 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 13 will include nominal visits of "Week 13" or "Early Termination" if early termed from double-blind phase; and Week 156/early termination will have nominal visits of "Extension Week 156" or "Early Termination" if early terminated from extension phase and taking open-label filgotinib.

The number and percentage of subjects in each cross-classification group of the shift table will be presented. Subjects with a missing value at baseline or postbaseline will not be included in the denominator for the 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

Bottai M, Zhen H. Multiple imputation based on conditional quantile estimation. Epidemiology Biostatistics and Public Health 2013;10 (1).

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.

Li HQ, Tang ML, Poon WY, Tang NS. Confidence Intervals for Difference Between Two Poisson Rates. Communications in Statistics - Simulation and Computation 2011;40 (9):1478-93.

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

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.

Ulm K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). Am J Epidemiol 1990;131 (2):373-5.

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 – Double-blind Treatment Phase

EVENT								Safety follow-up[c]
Study days (D)/weeks (W)[a]	Screening	Day 1	W2 (±5 days)	W4 (±5 days)	W8 (±5 days)	W13** (+5 days)	ET[b]	30 Days after last dose of study drug (±5 days)
Informed Consent	X							
Demographics and Medical History(Disease Characteristics, RA/PsA/AS/nrAxSpA History, Number of Children Fathered, Smoking Habits, Average Weekly Alcohol Consumption, and Family History of Coronary Heart Disease)	X							
Inclusion/exclusion criteria review	X	X						
Complete Physical Exam[c]	X						X	X
Symptom-directed Physical Exam, as needed		X	X	X	X	X		
Vital Signs[c] and Weight	X	X	X	X	X	X	X	X
Height	X							
12-lead ECG	X					X	X	
Tender Joint Count 28 (TJC28) (for RA subjects only)	X							
Swollen Joint Count 28 (SJC28) (for RA subjects only)	X							
Tender Joint Count 68 (TJC68) (for PsA subjects only)	X							

EVENT								Safety follow-up[c]
Study days (D)/weeks (W)[a]	Screening	Day 1	W2 (±5 days)	W4 (±5 days)	W8 (±5 days)	W13** (+5 days)	ET[b]	30 Days after last dose of study drug (±5 days)
Swollen Joint Count 66 (SJC66) (for PsA subjects only)	X							
Patient Global Assessment of Disease Activity (PtGADA) (for RA and PsA subjects only)	X							
Physician's Global Assessment of Disease Activity (PhGADA)[d] (for all subjects)	X	X				X		
Patient's Global Assessment of Pain Intensity (PPain) (for PsA subjects only)	X							
Bath Ankylosing Spondyloarthritis Disease Activity Index (BASDAI) (for AS/nrAxSpA subjects only)	X							
TB (Quantiferon ®)Test[e]	X							
Chest X-ray[f]	X							
Urinalysis[g]	X	X				X	X	X
Urine drug screen[g]	X							
Hematology and Serum Chemistry[g]	X	X	X	X	X	X	X	X
Lipid profile (fasting) (Total cholesterol and subfractions)[g]		X				X		
Serum hsCRP	X	X	X	X	X	X		
Endocrine: TSH, HbA1c[g]	X							

EVENT								Safety follow-up[c]
Study days (D)/weeks (W)[a]	Screening	Day 1	W2 (±5 days)	W4 (±5 days)	W8 (±5 days)	W13** (+5 days)	ET[b]	30 Days after last dose of study drug (±5 days)
LH, FSH, inhibin B, total Testosterone collection time between 07:00-11:00 in the morning[g]	X	X		X		X	X	X
PK collection (sparse)[h]			X	X		X		
Concomitant Medications	X	X	X	X	X	X	X	X
Assessment of Adverse Events	X	X	X	X	X	X	X	X
Semen Collection (2 samples as per collection instructions in Protocol Section 6.11)[i]	X					X	X[j]	
Date and time of most recent ejaculation[k]	X					X	X	
HIV, Hepatitis B, and Hepatitis C[l]	X							
HBV DNA[l, m]						X		
Randomization		X						
Study Drug Accountability				X	X	X	X	
Study Drug Dispensation[n]		X		X	X	X[o]		
In-clinic Dosing[n]		X	X			X		

a Visits correlate with the number of days/weeks on drug.
b Any time a subject discontinues participation in the study prior to Week 13 an Early Termination (ET) Visit is required. Subjects who are Responders at Week 13 and choose not to continue in the Extension Phase are also required to complete ET visit assessments.
c All subjects will have a 30-day safety follow-up visit after discontinuing blinded study drug, except Arthritis Responders at Week 13 who are continuing in the Extension Phase of the study.
c As detailed in Protocol Appendix 11
d For definitions of Arthritis Responder/Non-responder status please refer to protocol Definition of Terms.
e Proof of no active or untreated latent tuberculosis (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 Sponsor Medical Leader or designee.
f 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.
g As detailed in Protocol Appendix 11. For visits that require fasting, subjects should not have any food or drink (except water) for at least 8 hours before the visit.
h Sparse plasma pharmacokinetic (PK) samples at Week 2 are collected at 30 minutes post dose. The PK sample at Week 4 can be collected at any time without regard to dosing. The PK sample at Week 13 is collected prior to study drug administration. Please refer to protocol Section 6.14 for details on study drug administration.
i The Screening semen sample collection should coincide with the Day 1 (Baseline) visit as much as possible. Semen samples will be collected on the visit day and/or as soon as possible after the visit day.
j Semen collection will be completed at ET visit only if previous semen samples were not collected within 2 weeks of ET.
k This question will be asked prior to each semen collection.
l 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.
m Subjects with positive HBV core Ab and negative HBV DNA at Screening require ongoing monitoring with blood tests for HBV DNA every 3 months.
n The subject will take the first dose (Day 1) in the clinical study center. For PK collection purposes, at 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 6.14.
o Only for subjects continuing on open-label filgotinib in the Extension Phase.

** Week 13 visit must occur after 13 weeks of drug exposure. Therefore, visit window for this visit is +5 days. The end of the Double-Blind Treatment Phase is when the Week 13 semen results are available to the investigator and have been evaluated. Based on this evaluation, subjects are assigned to enter the Monitoring Phase or the Extension Phase.

Monitoring Phase (Protocol Section 6.6)

EVENT	Monitoring Phase (±5 days)	ET
Study days (D)/weeks (W)	**Visits occur every 13 weeks starting from their entry into the Monitoring Phase (for up to 52 weeks or until Reversibility is met, whichever is sooner)**	
Symptom-directed Physical Exam, as needed	X	X
Vital Signs[a] and Weight	X	X
Urinalysis[a]	X	X
Hematology and Chemistry[a]	X	X
Serum hsCRP	X	
LH, FSH, inhibin B, total Testosterone[a] collection time between 07:00-11:00 in the morning	X	X
Concomitant Medications	X	X
Assessment of Adverse Events[b]	X	X
Semen Collection[c] (2 samples as per collection instructions in Protocol Section 6.11)	X	X[e]
Date and time of most recent ejaculation[d]	X	

a As detailed in protocol Appendix 11.
b As detailed in protocol Section 7.3
c Semen samples will be collected every 13 weeks from the day of study drug discontinuation, for up to 52 weeks or until Reversibility is met, whichever is achieved sooner. Reversibility is met when all sperm parameter(s) qualifying the subject to enter the Monitoring Phase return(s) to greater than 50% of Baseline (ie, to greater than [0.50 x Baseline]).
d This question will be asked prior to semen collection not at the site visit.
e Semen collection will be completed at ET visit only if previous semen samples were not collected within 2 weeks of ET.

Extension Phase (Protocol Section 6.5)

EVENT	Extension Phase		Safety follow-up
Study days (D)/weeks (W)	**Week 26 up to 156 weeks[a] (±10 days)**	**ET[b]**	**30 Days after last dose of study drug (±5 days)**
Symptom-directed Physical Exam, as needed	X	X	X
Vital Signs and Weight	X	X	X
TB QuantiFERON[c]	X		
Urinalysis[d]	X	X	X
Hematology and Chemistry[d]	X	X	X
Lipid profile (fasting) [Total cholesterol and subfractions][d,e]	X		
Serum hsCRP	X		
LH, FSH, inhibin B, total Testosterone[d] collection time between 07:00-11:00 in the morning	X		
Concomitant Medications	X	X	X
Assessment of Adverse Events	X	X	X
Semen Collection (2 samples as per collection instructions in protocol Section 6.11)	X	X[l]	
Date and time of most recent ejaculation[f]	X		
Study Drug Accountability[g]	X	X	
Study Drug Dispensation[g]	X		
12-lead ECG[h]	X	X	
PK collection (sparse)	X[i]		
HBV DNA[j,k]	X		
In-clinic dosing	X (Week 26)		

a Subjects in the Extension Phase (EP) will complete EP Week 26 visit and every 13 weeks thereafter up to 156 weeks.
b Any time a subject discontinues participation in the study prior to Week 156 in the Extension Phase, an Early Termination (ET) Visit is required.
c In the EP 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.
d As detailed in protocol Appendix 12. For visits that require fasting, subjects should not have any food or drink (except water) for at least 8 hours before the visit.
e Lipid profile should be performed at Weeks 26, 39, 65, 91, 117 and 143.
f This question will be asked prior to semen collection not at the site visit.
g Only applicable for subjects on filgotinib.
h ECG will be performed at Week 156 or ET (for subjects on open-label filgotinib only).
i A sparse plasma PK sample is collected at Week 26 prior to study drug administration (for subjects on open label filgotinib only). Please refer to protocol Section 6.14 for details on study drug administration.
j 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 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.
k Subjects with positive HBV core Ab and negative HBV DNA at Screening require ongoing monitoring with blood tests for HBV DNA every 3 months.
l Semen collection will be completed at ET visit only if previous semen samples were not collected within 2 weeks of ET.

Appendix 2. Calculation of the 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 4). Strata are (1) MTX use = yes, sperm concentration 15-50 M/mL; (2) MTX use = no, sperm concentration 15-50 M/mL; (3) MTX use = yes, sperm concentration > 50 M/mL; (4) MTX use = no, 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.

SAP GLPG0634-CL-227_v1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Regulatory Affairs eSigned	18-Jan-2021 04:49:49
PPD	Clinical Research eSigned	19-Jan-2021 16:04:42
PPD	Biostatistics eSigned	19-Jan-2021 16:09:19

Signature Page for glpg0634-cl-227-sap 15548

Approval	PPD ian Biometrics 20-Jan-2021 17:55:59 GMT+0000

Approval	PPD ector Clinical Development 20-Jan-2021 18:00:57 GMT+0000

Signature Page for glpg0634-cl-227-sap43952_16577_36930