STATISTICAL ANALYSIS PLAN

Project Number:	GLPG0634
Study Number:	GS-US-417-0304 (GLPG0634-CL-304)
Study Title:	A Multicenter, Open-label, Long Term Extension Study to Assess the Safety and Efficacy of Filgotinib in Subjects with Rheumatoid Arthritis

Development Phase:	3		
Status:	Final		
Version:	4.0	Date:	13-JUN-2025
EudraCT:	2016-003630-25	IND:	115,510
Sponsor:	Alfasigma S.p.A		
Biostatistician:	[redacted] and [redacted]		
Medical Leader:	[redacted]		

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF ABBREVIATIONS

5L	5 Levels
ACR	American College of Rheumatology
ACR20/50/70	American College of Rheumatology 20/50/70% improvement
AEI	adverse events of interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical drug class
bDMARDs	biologic disease modifying antirheumatic drugs
BLQ	below the limit of quantitation
BMI	body mass index
CBC	complete blood count
CDAI	clinical disease activity index
COX-2	cyclooxygenase-2
CPK	creatinine phosphokinase
cPYE	censored patient years of exposure
csDMARDs	conventional synthetic (cs) disease modifying antirheumatic drugs
CSR	clinical study report
CTCAE	Common Toxicity Criteria for Adverse Events
CVEAC	cardiovascular safety endpoint adjudication committee
DAS28	disease activity score for 28 joint count
DMC	data monitoring committee
DVT	deep vein thrombosis
EAIR	exposure adjusted incidence rate
EAER	exposure adjusted event rate
ECG	electrocardiogram
eCRF	electronic case report form
[redacted]	
ET	early termination
EULAR	European League Against Rheumatism
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-
Fatigue GI	gastrointestinal
HLGT	high-level group term
HLT	high-level term
[redacted]	
hsCRP	high-sensitivity C-reactive protein
IAC	independent adjudication committee
ID	identification
IWRS	interactive web response system
LDL	low density lipoprotein
LLOQ	lower limit of quantitation
LLT	lower-level term
LOQ	limit of quantitation
LTE	long term extension

MACE	major cardiovascular adverse event
MCS	mental component summary
MedDRA	Medical Dictionary for Regulatory Activities
MST	MedDRA search term
MTX	methotrexate
NLP	Natural Language Processing
NRI	Non-responder imputation
NSAIDs	Nonsteroidal anti-inflammatory drugs
OC	observed case
PCS	physical component summary
PE	pulmonary embolism
PGA	physician's global assessment
PT	preferred term
PTM	placebo-to-match
PYE	patient years of exposure
Q1, Q3	first quartile, third quartile
QD	quaque die; once daily
RA	rheumatoid arthritis
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SDAI	simplified disease activity index
SE	standard error
SF-36	36-item short form survey
SGA	subject's global assessment
SJC	swollen joint count
SMQs	standardized MedDRA queries
SOC	system organ class
TE	treatment-emergent
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TJC	tender joint count
TNFα	tumor necrosis factor-alpha
ULN	upper limit of normal
VAS	visual analog scale
WBC	white blood cell
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) in the final analyses for Study GS-US-417-0304. This SAP is based on the clinical study protocol (CSP) Amendment 7 dated 21 January 2022 and the electronic case report form (eCRF).

1.1. Study Objectives

The primary objective of this study is as follows:

- To evaluate the long-term safety and tolerability of filgotinib in subjects who have completed one of the parent studies of filgotinib in rheumatoid arthritis (RA)

The secondary objective of this study is as follows:

- To evaluate the long-term efficacy of filgotinib in subjects with RA

The exploratory objectives of this study are as follows:

- To evaluate the long-term effects of filgotinib on subject-reported outcomes, such as disability, fatigue, and quality of life
- [redacted]

1.2. Study Design

This study was originally designed and initiated as a dose-blinded, long term extension (LTE) study of safety and efficacy of filgotinib in subjects with RA. Subjects were enrolled in the study after having completed one of the 3 parent RA studies (GS-US-417-0301, GS-US-417-0302, or GS-US-417-0303). The study design changed to open-label following implementation of protocol amendment 4.

The subject's first dose of LTE study drug defines Day 1 of participation on the LTE, and should not occur sooner than the day after the final visit for the parent protocol. The Day -1 visit for this study will be performed at either the last visit for the parent protocol (Week 52 visit for GS-US-417-0301 and GS-US-417-303, Week 24 visit for GS-US-417-0302) or within 4 weeks after the final study visit for the parent protocol.

The subject could be consented and eligibility confirmed on Day -1 (the same day as the last visit for the parent protocol), with study drug dispensed to start on Day 1 of the LTE study.

Eligibility for participation in the study will be confirmed on completion of the parent protocol. Subjects who wish to enroll in the extension study are to review and sign the study ICF prior to any procedures or tests being performed.

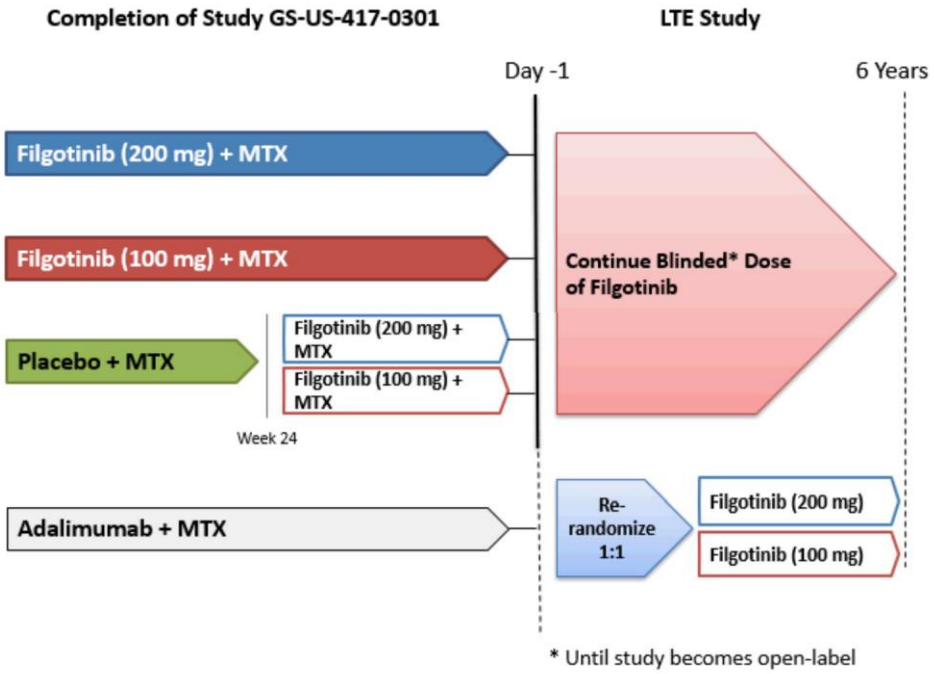
After eligibility for the study is confirmed, the subject will be assigned to one of the two filgotinib dosing arms in a blinded fashion:

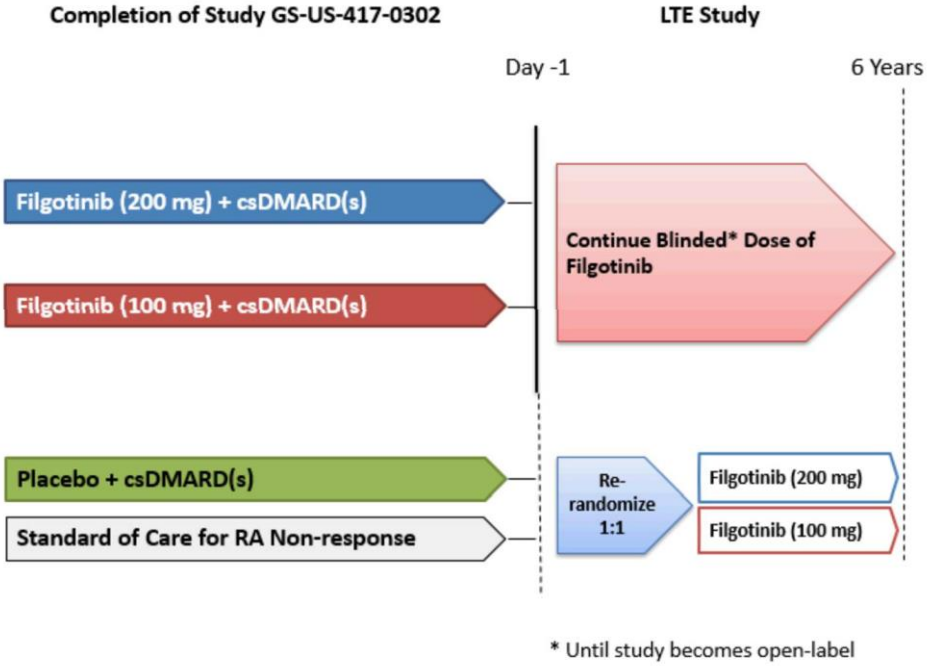
- Subjects who are on blinded filgotinib at the final visit of the parent study continued on their same dose of filgotinib in a blinded fashion (100 mg or 200 mg QD) until study unblinding.
- Subjects who are receiving adalimumab (GS-US-417-0301), placebo (GS-US-417-0302), or methotrexate monotherapy (GS-US-417-0303) at their final visit of the parent study, will be re-randomized on Day -1 of the LTE in a 1:1 ratio to either 100 mg or 200 mg filgotinib QD in a blinded fashion until study unblinding.
- Subjects from study GS-US-417-0302, who discontinue blinded study drug due to inadequate response of their RA, are eligible for the LTE, and will be re-randomized on Day -1 of the LTE in a 1:1 ratio to either 100 mg or 200 mg filgotinib QD in a blinded fashion until study unblinding.
- Subjects from studies GS-US-417-0301 and GS-US-417-0303, who complete the studies after being transitioned to standard of care therapy, will not be eligible for the LTE study.
- Subjects will be provided filgotinib for up to 6 years, until the sponsor terminates clinical development of filgotinib, or until filgotinib becomes commercially available; whichever comes first.
- All subjects will remain blinded to their filgotinib daily dosing regimen until the study becomes open-label.

All subjects may continue their stable dose of permitted csDMARD(s) as indicated in the parent protocol, with the exception of subjects from study GS-US-417-0303, who will discontinue their study drug doses of weekly MTX/PTM. These subjects may resume/start MTX and/or other csDMARDs (per investigator judgment) after at least 4 weeks of their first dose of study drugs in the LTE.

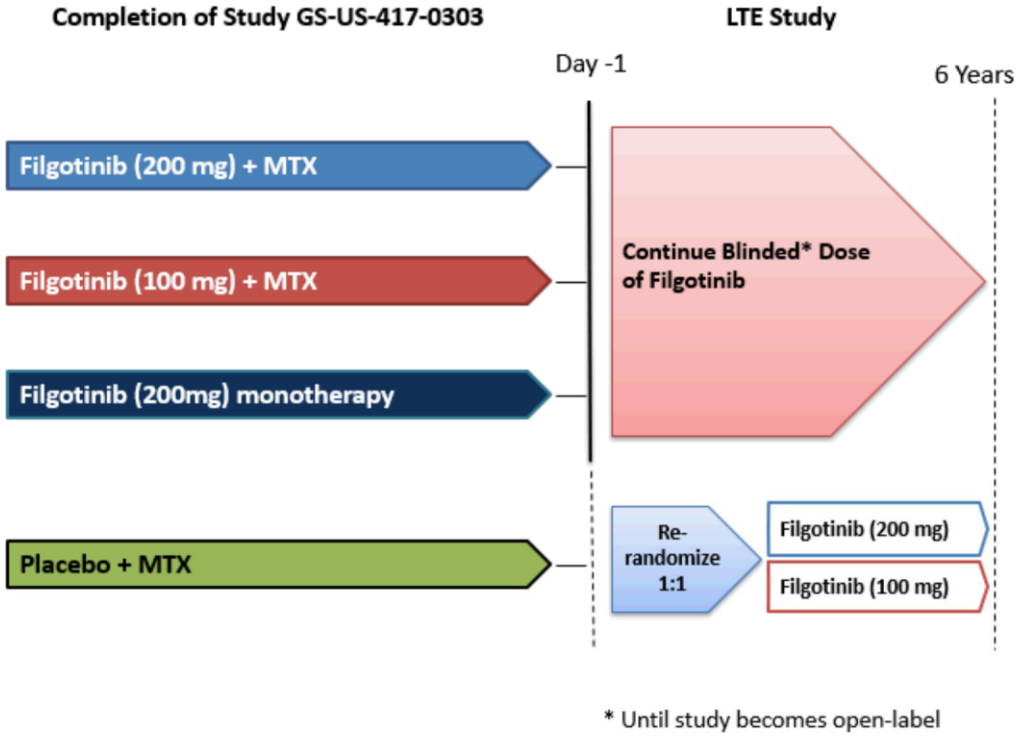
All subjects who are re-randomized on Day -1 of the LTE will be stratified by geographic region and parent study.

Figure 1-1. Study Design





2



1.3. Sample Size and Power

No formal hypothesis testing is planned for this study and sample size calculation is not conducted. All subjects who completed or met protocol specified discontinuation criteria in a prior filgotinib phase 3 parent study in RA could enroll in this long-term extension study.

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee Analyses

An external multidisciplinary data monitoring committee (DMC) appointed to monitor the parent studies also reviewed the progress of the study and performed interim reviews of safety data in order to protect subject welfare and preserve study integrity. To ensure the best interests of the participants, the DMC was to recommend to the sponsor if the nature, frequency, and severity of adverse effects associated with the study treatment warranted the early termination of the study, the continuation of the study, or the continuation of the study with modifications.

Regular DMC review of safety data was scheduled until the study was unblinded. The DMC was dissolved following global unblinding of the study, when protocol amendment 4 was globally approved.

The DMC's role and responsibilities and the scope of analysis provided to the DMC are described in the parent studies' mutually agreed upon charters, which define the DMC membership, meeting logistics, and meeting frequency. {Liu 2017, Ye 2019, Yin 2019}

2.2. Interim Analysis

Two unblinded interim safety analyses were performed for regulatory submission purposes. The initial interim safety analysis with data cut off on 12 March 2019 was performed after all randomized subjects completed their week 24 visit in parent studies (or prematurely discontinued from the parent studies prior to week 24). The second interim safety analysis with data cut off on 16 September 2019 was performed after all randomized subjects completed their week 52 visit in parent studies (or prematurely discontinued from the parent studies prior to week 52). The results of both interim analyses were included in submission to regulatory authorities. The first interim efficacy and safety analysis was performed with data cut-off on 01 June 2020. Additional interim analyses of safety and/or efficacy were performed as needed.

2.3. Final Analysis

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

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

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

By-subject listings will be presented for all subjects in the Safety Analysis Set and sorted by subject identification (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 listings. The treatment group to which subjects were randomized or assigned will be used in the listings, as well as age, sex at birth, race, and ethnicity.

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 TFL.

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

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

3.1.1. All Enrolled Analysis Set

All Enrolled Analysis Set includes all subjects who received a study subject identification number in the LTE study after screening.

3.1.2. Safety Analysis Set

The Safety Analysis Set includes all subjects who enrolled in the LTE and received at least 1 dose of filgotinib in the LTE. This is the primary analysis set for safety and efficacy analyses.

3.1.3. Parent Studies Analysis Set

The Parent Studies Analysis Set includes all subjects who enrolled in either of the 3 parent studies and received at least one dose of study treatment in the parent studies. The Safety Analysis Set is therefore a subset of this analysis set.

3.2. Subject Grouping

For analyses based on the Safety Analysis Set, subjects will be grouped according to actual treatment received. The actual treatment received will differ from the randomized or assigned treatment only when their actual treatment differs from the randomized or assigned treatment for

the entire treatment duration.

Subjects will be grouped by prior exposure to filgotinib in parent studies and the treatment received in the LTE. The treatment groups are: 1) filgotinib 200 mg QD with prior filgotinib exposure; 2) filgotinib 200 mg QD with no prior filgotinib exposure; 3) total filgotinib 200 mg QD; 4) filgotinib 100 mg QD with prior filgotinib exposure;
5) filgotinib 100 mg QD with no prior filgotinib exposure; 6) total filgotinib 100 mg QD.

3.3. Strata and Covariates

All subjects that require re-randomization will be randomly assigned to treatment groups via the interactive web response system (IWRS) in a 1:1 ratio using a stratified randomization schedule. Stratification will be based on the following variables:

- Geographic region (**Group A** includes the following countries: Australia, Belgium, Canada, France, Germany, Ireland, Israel, Italy, Netherlands, New Zealand, South Korea, Singapore, South Africa, Spain, Sweden, Switzerland, United Kingdom, and USA; **Group B** includes the following countries: Bulgaria, Croatia, Czech Republic, Estonia, Georgia, Hungary, India, Latvia, Moldova, Poland, Romania, Russia, Serbia, Slovakia, and Ukraine; **Group C** includes the following countries: Argentina, Brazil, Chile, Colombia, Mexico, Peru and Puerto Rico; **Group D** includes the following countries: China, Hong Kong, Malaysia, Philippines, Taiwan, Thailand, and Vietnam; **Group E** includes Japan;

- Parent Study (GS-US-417-0301, GS-US-417-0302, or GS-US-417-0303)

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 analyses. Sensitivity analysis using the IWRS stratification factor values may be performed as appropriate.

3.4. Missing Data and Outliers

3.4.1. 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. For partial date of initial RA diagnosis, imputation rules are described in Section 5.2. The handling of missing or incomplete dates for AE onset is described in Section 7.1.5.2, and for prior and concomitant medications in Section 7.4. For the missing or incomplete end of study date and missing of last contact date, the rules is in Section 4.2.1.

3.4.2. Outliers

Outliers will be identified during the data management and data analysis process, but no sensitivity analyses will be conducted. Outliers will not be excluded from data analysis.

3.5. Data Handling Conventions and Transformations

In general, age (in years) on the date of the first dose of study drug in LTE will be used for analyses and presentation in listings. Age (in years) on the date of the first dose of study drug in parent study will be present in the parent baseline demographics. If only birth year is collected on the eCRF, "01 July" will be used for the unknown birth day and month for the purpose of age calculation. If only birth year and month are collected, "01" 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 for 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).

3.6. Analysis Visit Windows

3.6.1. Definition of Study Day

The first dose date of individual study drug will be calculated separately for each study drug (for example, filgotinib 200 mg, filgotinib 100 mg and PTMs) in a treatment group. Study Day 1 is defined as the first dose date of any study drug administered during the LTE study, which is the earliest of the first dose dates of individual study drugs in a treatment group. For analyses including data from both the parent study and the LTE: Study Day 1 is defined as the first dose date of any study drug administered during the parent study.

The last dose date of individual study drug will be calculated separately for each study drug in a treatment group. The last dose date for an individual study drug will be the end date on study drug administration CRF for the record where the "study drug was permanently withdrawn" flag is 'Yes'. The last dose date of any study drug will be defined as the maximum of the last dose dates of individual study drugs in a treatment group.

Study Day will be calculated from the Study Day 1 and derived as follows:

- For postdose study days: Assessment Date − Study Day 1 + 1

- For days prior to the first dose: Assessment Date − Study Day 1

3.6.2. Analysis Visit Windows

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

In general, the parent study baseline will be the last nonmissing value on or prior to the first dose date of study drug in the parent study.

The LTE baseline will be the last nonmissing value on or prior to the first dose date of study drug in the LTE study.

- For laboratory tests, if there are no values available on or prior to the first dose date of study drug in the LTE study, the value on last study visit in parent study (including unscheduled visits after the last study visit) will be carried forward as LTE baseline. The last study visit is Week 52 for studies GS-US-417-0301 and GS-US-417-0303, and Week 24 for study GS-US-417-0302. No further values will be carried over. If the last study visit in parent study still has missing values, the LTE baseline will be set as missing. This is applied to all laboratory values including high-sensitivity C-reactive protein (hsCRP).

- For efficacy endpoints, if there are no values available on or prior to the first dose date of study drug in the LTE study, the value on last study visit in parent study (including unscheduled visits after the last study visit) will be carried forward as LTE baseline. The last study visit is Week 52 for studies GS-US-417-0301 and GS-US-417-0303, and Week 24 for study GS-US-417-0302. No further value will be carried over. If the last study visit in parent study still has missing values, the LTE baseline will be set as missing.

The parent study baseline will be the value that was used as baseline in the parent study.

For vital signs, ECG, weight, lipids and safety laboratory data, the analysis visit windows will be applied to data collected during the ‘on-treatment’ period. The ‘on treatment’ period is defined as the last dose date of any study drug + 7 days.

Analysis windows are constructed around the protocol pre-planned visits as defined in Protocol amendment 7.

The analysis windows for joint count assessment, HAQ-DI (including subject’s pain assessment), subject’s global assessment (SGA), physician’s global assessment (PGA), serum CRP, vital signs, weight, and safety laboratory data (lipids data excluded) are provided in Table 3-1.

Table 3-1. Analysis Visit Windows for Joint Count Assessment, HAQ-DI, SGA, PGA, CRP, Vital Signs, Weight, and Safety Laboratory Data (Lipid Data Excluded)

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
LTE Baseline/Day 1	1	(none)	1
LTE Week 2	15	2	29
LTE Week 6	43	30	64
LTE Week 12	85	65	127
LTE Week 24	169	128	211
LTE Week 36	253	212	295
LTE Week 48	337	296	379
LTE Week 60	421	380	463
LTE Week 72	505	464	547
LTE Week 84	589	548	631
LTE Week 96	673	632	715
LTE Week 108	757	716	799
LTE Week 120	841	800	883
LTE Week 132	925	884	967
LTE Week 144	1009	968	1093
LTE Week 168	1177	1094	1261
LTE Week 192	1345	1262	1429
LTE Week 216	1513	1430	1597
LTE Week 240	1681	1598	1765
LTE Week 264	1849	1766	1933
LTE Week 288	2017	1934	2101
LTE Week 312	2185	2102	≥ 2185

The analysis windows for 36-Item Short Form Survey (SF-36), Functional Assessment of Chronic Illness Therapy – Fatigue (FACIT-Fatigue), [redacted] and electrocardiogram (ECG) data are provided in Table 3-2.

Table 3-2. Analysis Visit Windows for SF-36, FACIT-Fatigue, [redacted], and ECG Data

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
LTE Baseline/Day [redacted]	[redacted]		
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			
LTE Week [redacted]			

LTE Week [redacted]		[redacted]
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		
LTE Week		

The analysis windows for [redacted] and lipid data are provided in Table 3-3.

Table 3-3. Analysis Visit Windows for [redacted] and Lipid Data

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
LTE Baseline/Day [redacted]	[redacted]		
LTE Week [redacted]			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			
LTE Week			

Vital signs, weight, ECGs, lipids and safety laboratory data collected in the post-treatment follow-up period will also be summarized. The analysis window for the post-treatment follow-up period is defined as from (the last dose date of any study drug + 8 days) to (the last dose date of any study drug + 30 days). If multiple valid, nonmissing measurements exist in the post-treatment follow-up visit window, then the latest record will be selected for analysis. If chronological order cannot be determined (eg, more than 1 record on the same day with time missing), for any given subject, the value with the worst severity will be selected for categorical variable, and the average will be taken for continuous variable. Data obtained after the last dose date + 30 days will be excluded from the summaries, but it will be included in the listings.

3.6.3. Selection of Non-Efficacy Data in the Event of Multiple Records in an Analysis Visit Window

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

Continuous Measurements:

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:

- The baseline value will be the last nonmissing value on or prior to the first dose date of study drug in LTE for LTE baseline or the first dose date in the parent study for parent baseline, unless otherwise specified. If multiple measurements occur on the same day, the last nonmissing value prior to the time of first dose of study drug will be considered as the baseline value. If these multiple measurements occur at the same time or the time is not available, the average of these measurements (for continuous data) will be considered the baseline value.
- For postbaseline visits:
  - The record closest to the nominal day for that visit will be selected.
  - If there are 2 records that are equidistant from the nominal day, the later record will be selected.
  - If there is more than 1 record on the selected day, the average will be taken, unless otherwise specified.

Categorical Measurements:

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 first dose date of study drug in LTE for LTE baseline or the first dose date in the parent study for parent study baseline will be selected. If there are multiple records with the same time or no time recorded on the same day, the value with the lowest severity will be selected (eg, normal will be selected over abnormal for safety ECG findings).
- For postbaseline visits:
  - The record closest to the nominal day for that visit will be selected.
  - If there are 2 records that are equidistant from the nominal day, the later record will be selected.
  - If there is more than 1 record on the selected day, the value with the worst severity will be selected (eg, abnormal will be selected over normal for safety ECG findings), unless otherwise specified.

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

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

For subjects who are re-randomized, the enrollment will be summarized by stratification factor stratum (i.e., parent study protocol and geographic region). The denominator for the percentage of subjects in the stratum will be the total number of subjects re-randomized. If there are discrepancies in the value used for stratification assignment between the IWRS and the clinical database for re-randomized subjects, 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 re-randomization in the study will be provided in a listing and as an appendix to the CSR.

A summary of subject disposition will be provided by treatment group, once based on the Safety Analysis Set and another based on the Parent Studies Analysis Set. This summary will present the number of subjects in each of the categories listed below:

- Safety Analysis Set
- Continuing study drug
- Did not complete study drug with reasons for premature discontinuation of study drug
- Continuing study
- Did not complete study with reasons for premature discontinuation from the study for the reasons for premature discontinuation of study drug and 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 corresponding to that column.

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

- Reasons for premature study drug or study discontinuation
- Lot number and kit ID of assigned study drugs

4.2. Extent of Study Drug Exposure

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

4.2.1. Duration of Exposure to Study Drug

For analyses based on the Safety Analysis Set, total duration of exposure to study drug will be defined as (last dose date of study drug in the LTE study - first dose date of study drug in the LTE study + 1), minus all temporary interruptions in study drug administration. and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks). The total duration of exposure (in years) is defined as:

$$\text{Total duration of exposure in years} = \text{Total duration of exposure to study drug in days} / 365.25$$

For analyses based on the Parent Studies Analysis Set including exposure data from both the parent and LTE studies, the total duration of exposure to study drug will be defined as (last dose date of study drug (over both parent and LTE study combined) – first dose date of study drug in the parent study + 1).

For subjects who completed the study, discontinued the study or die, if the last dose date is incomplete or missing, the following rules for deriving the last dose date will be applied:

- If day is missing, it will be imputed with the end of the month. If this imputed date falls after the date of last contact, then it will be imputed with the date of last contact. The date of last contact is the maximum date in the database per subject, based on visit/assessment dates, adverse event dates, and subject disposition dates.
- If day and month are missing, these will be imputed with the end of the year (31-Dec), if the available year does not match the year of last contact (from the date of last contact). If the available year matches the year of last contact, then it will be imputed with the date of last contact.
- If completely missing, it will be imputed with the date of last contact.

Similarly, if end of study date is missing or incomplete, it will be imputed with the date of last contact.

The total duration of exposure to any 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 in the LTE:

LTE Baseline (Day 1)
LTE Week 2 (Day 15) – Year 0.04
LTE Week 6 (Day 43) – Year 0.12
LTE Week 12 (Day 85) – Year 0.23

LTE Week 24 (Day 169) – Year 0.46
LTE Week 36 (Day 253) – Year 0.69
LTE Week 48 (Day 337) – Year 0.92
LTE Week 60 (Day 421) – Year 1.15
LTE Week 72 (Day 505) – Year 1.38
LTE Week 84 (Day 589) – Year 1.61
LTE Week 96 (Day 673) – Year 1.84
LTE Week 108 (Day 757) – Year 2.07
LTE Week 120 (Day 841) – Year 2.30
LTE Week 144 (Day 1009) – Year 2.76
LTE Week 168 (Day 1177) – Year 3.22
LTE Week 192 (Day 1345) – Year 3.68
LTE Week 216 (Day 1513) – Year 4.14
LTE Week 240 (Day 1681) – Year 4.60
LTE Week 264 (Day 1849) – Year 5.06
LTE Week 288 (Day 2017) – Year 5.52
LTE Week 312 (Day 2185) – Year 5.98

If no subject has reached a time period, the total duration of exposure won't be summarized at this time period.

Summaries will be provided by treatment group for the Safety Analysis Set. No formal statistical testing is planned.

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

4.2.2. Dose Reduction

Event of dose reduction from filgotinib 200 mg to filgotinib 100 mg is identified from drug dispensations in IWRS data. IWRS for clinical trials are used to randomize patients and manage the drug supply.

If the subject with event of dose reduction also has AE's with action taken as "Dose Reduced", then the reason of dose reduction is categorized as due to AE. If the dose reduction is due to mistake, sites are instructed to enter descriptive texts in general comment CRF page. Thus for subject with event of dose reduction, if the general comment contains key word of "dose reduction" or its synonyms, and also key word "error" or its synonyms, the reason of dose reduction is categorized as due to error. Any dose reduction events not categorized as due to AE or error is assigned with the reason of others.

4.3. Protocol Deviations

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

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

5. BASELINE CHARACTERISTICS

5.1. LTE Demographics and Other Baseline Characteristics

The LTE baseline will be the last nonmissing value on or prior to the first dose date of study drug in the LTE study.

Subject demographic variables will be summarized by treatment group and overall using descriptive statistics for continuous variables, and using number and percentage of subjects for categorical variables. The summary of demographic data will be provided for the Safety Analysis Set for the following:

- Age (on the first dose date of any study drug in the LTE)
- Age group (<65 years, ≥65 years)
- Age group (<75 years, ≥75 years)
- Sex at birth (male, female)
- Race
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Geographic region and country
- Weight (<60kg, 60kg≤ to <100kg, ≥100kg)
- Height
- Body mass index (BMI; in kg/m^2) (< 25 or ≥ 25) (< 30 or ≥ 30)
- Smoking status

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

5.2. LTE Baseline Disease Characteristics

LTE baseline disease characteristics include:

- Duration of RA from diagnosis (years)

Calculated as ([first dose date in the LTE] – [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 day of the month

  - missing month: use January
- Parent study
- Prior exposure to biologic disease modifying antirheumatic drugs (bDMARDs) (Yes/No): n (%)
  - Among Yes: Adalimumab from subjects in study GS-US-417-0301
- Concurrent oral corticosteroids use on the first dosing date (Yes/No): n (%)
  - Oral corticosteroids dose, mg/day, expressed as prednisone-equivalent dose, mean (SD)
- Concurrent MTX use on the first dosing date (Yes/No): n (%)
  - MTX dose, mg/week, mean (SD)
- Number of concurrent csDMARDs use on the first dosing date: n (%)
  - 0
  - 1
  - ≥ 2
- Swollen joint count based on 66 joints (SJC66)
- Tender joint count based on 68 joints (TJC68)
- Swollen joint count based on 28 joints (SJC28)
- Tender joint count based on 28 joints (TJC28)
- Health assessment questionnaire-disability index (HAQ-DI) total score
- Disease Activity Score for 28 joint count using CRP (DAS28[CRP])
- Subject's pain assessment (by visual analog scale [VAS] in mm)
- Subject's global assessment of disease activity (SGA) (by VAS in mm)
- Physician's global assessment (PGA) (by VAS in mm)
- Simplified Disease Activity Index (SDAI)
- Clinical Disease Activity Index (CDAI)
- High-Sensitivity C-reactive Protein (hsCRP [mg/L])

Selection of csDMARD and bDMARD will be based on WHODrug Dictionary Preferred Terms, while Oral Corticosteroid will be selected based on ATC Class 4 (Appendix 1).

These LTE baseline characteristics will be summarized by treatment group and overall using descriptive statistics for continuous variables and using number and percentage of subjects for categorical variables. The summary of LTE baseline characteristics will be provided for the Safety Analysis Set.

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

5.3. Parent Study Demographics and Baseline Disease Characteristics

The parent study demographics and baseline disease characteristics will be summarized similar to LTE study. For each efficacy endpoint, its values on both LTE and parent study baseline will be summarized in its own summary table.

5.4. Medical History

Any ongoing or unresolved adverse events from the preceding parent studies will be documented in the medical history eCRF.

Medical history collected at screening will be coded using the Medical Dictionary for Regulatory Activities (MedDRA).

Medical history will be summarized by system organ class (SOC), preferred term (PT), treatment group, and overall. Subjects who report 2 or more medical history items that are coded to the same SOC and/or PT will be counted only once by the unique coded term in the total subject summary.

In addition, numbers and percentages of subjects who have any first degree relatives that had experienced myocardial infarction or stroke before the age of 50 years, experienced myocardial infarction or stroke before the age of 50 years, or have diabetes (type I or II) will be summarized.

The summary will be provided for the Safety Analysis Set. No formal statistical testing is planned.

A by-subject listing of medical history will be provided by subject ID number.

6. EFFICACY ANALYSES

6.1. General Considerations

The primary analysis set for efficacy analyses will be the Safety Analysis Set, defined in Section 3.1.2.

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

- 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, or more than 1 record (with time known) on the selected day, the latest record will be taken
- If chronological order cannot be determined (eg, more than 1 record on the same day with time missing), for any given subject, the worst outcome will be selected.

For the calculation of composite endpoints including DAS28(CRP), ACR20/50/70, ACR-N, SDAI, CDAI and Boolean remission, we use the following steps unless otherwise specified:

- Step 1: Assign individual components to analysis visit windows defined in Section 3.6.2.
- Step 2: Within each analysis visit window, select component-level data based on the rules for selecting efficacy data as above
- Step 3: Calculate the composite endpoint based on the selected component-level data in Step 2.

Below are the descriptions for the imputation methods that will be used throughout the efficacy analyses:

- Observed case (OC): Missing values remain missing. For the categorical composite endpoints, in the case that some components are missing, the composite endpoint assessment will be derived based on the non-missing components. If non-missing components are not sufficient to determine final composite endpoint, then the composite endpoint will be set as missing. For continuous composite endpoints (including ACR-N), if any components are missing, the composite endpoints will be set as missing.
- Non-responder imputation (NRI): For all binary response measurements, starting from OC, all missing values will be set as non-responders, except missing values that are due to study termination in a country where filgotinib became commercially available (as per CSP amendment 7) or due to site termination (if any).

6.2. Efficacy Endpoints

The efficacy endpoints include:

- The proportion of subjects who achieve an American College of Rheumatology 20% improvement response (ACR20) over time
- The proportion of subjects who achieve an American College of Rheumatology 50% improvement response (ACR50) over time
- The proportion of subjects who achieve an American College of Rheumatology 70% improvement response (ACR70) over time
- Change from parent study baseline in individual components of the ACR response over time
- Change from parent study baseline in Disease Activity Score for 28 Joint Count using CRP (DAS28[CRP)]) over time
- The proportion of subjects who achieve DAS28(CRP) ≤ 3.2 over time
- The proportion of subjects who achieve DAS28(CRP) < 2.6 over time
- Change from parent study baseline in Clinical Disease Activity Index (CDAI) over time
- Change from parent study baseline in Simplified Disease Activity Index (SDAI) over time
- The proportion of subjects who achieve low disease activity (LDA) defined as CDAI≤10 over time
- The proportion of subjects who achieve clinical remission defined as CDAI ≤2.8 over time
- The proportion of subjects who achieve LDA defined as SDAI≤11 over time
- The proportion of subjects who achieve clinical remission defined as SDAI ≤3.3 over time
- ACR-N over time
- European League Against Rheumatism (EULAR) response over time.
- Boolean Remission over time
- Change from parent study baseline in SF-36 transformed scores, physical component summary (PCS) and mental component summary (MCS) over time
- Change from parent study baseline in Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-Fatigue) over time
- Change from parent study baseline in [redacted] over

time

- Change from parent study baseline in [redacted]
- [redacted]

6.3. Definition of Efficacy Endpoints

6.3.1. Tender/Swollen Joint Counts (TJC/SJC)

Tender joint count based on 68 joints (TJC68) and swollen joint count based on 66 joints (SJC66) will be collected during the course of the study. The assessment for each joint will be from the following selections: Tender Only, Swollen Only, Tender and Swollen, Joint Non-Evaluable or Missing, or Not Tender or Swollen.

Individual joint with missing assessment will not be imputed. If at least half of the joints are assessed at a given visit, the prorated tender and swollen joint counts will be calculated using the following formula:

$$TJC68 = \frac{Total\ number\ of\ tender\ joints}{68 - (Number\ of\ nonevaluable\ or\ missing\ joints\ out\ of\ 68\ joints)} \times 68$$

$$SJC66 = \frac{Total\ number\ of\ swollen\ joints}{66 - (Number\ of\ nonevaluable\ or\ missing\ joints\ out\ of\ 66\ joints)} \times 66$$

If less than half of joints are assessed at a given visit, joint counts are treated as missing for that visit.

A more abbreviated assessment considering 28 joints as listed in Table 6-1for both tenderness and swelling will also be conducted (as part of the TJC68 and SJC66 assessment), denoted as TJC28 and SJC28, respectively.

Table 6-1. Composition of the 28 Joints

Joints	Number
Shoulder Joints (Left and Right)	2
Elbow Joints (Left and Right)	2
Wrist Joints (Left and Right)	2
Metacarpophalangeal Joints I-V (Left and Right)	10
Hand Proximal Interphalangeal Joints I-V (Left and Right)	10
Knee Joints (Left and Right)	2

If there exist non-evaluable or missing joints among the 28 joints, similar prorated tender and swollen joint counts will be calculated as follows:

$$TJC28 = \frac{Total\ number\ of\ tender\ joints}{28 - (Number\ of\ nonevaluable\ or\ missing\ joints\ out\ of\ 28\ joints)} \times 28$$

$$SJC28 = \frac{Total\ number\ of\ swollen\ joints}{28 - (Number\ of\ nonevaluable\ or\ missing\ joints\ out\ of\ 28\ joints)} \times 28$$

If less than half of the 28 joints are assessed at a given visit, TJC28 and SJC28 are treated as missing for that visit.

6.3.2. Global Assessment of Disease Activity

Subject's Global Assessment of Disease Activity (SGA) and Physician's Global Assessment of Disease Activity (PGA) based on a 0-100 mm visual analog scale (VAS) will be recorded during the study, with 0 indicating "no disease activity" and 100 indicating "maximum disease activity" (or similar description of disease severity).

6.3.3. Health Assessment Questionnaire Disability Index (HAQ-DI)

The HAQ-DI score is defined as the average of the scores of eight functional categories (dressing and grooming, arising, eating, walking, hygiene, reach, grip, and other activities), reported by the patient. Responses in each functional category are collected as: without any difficulty; with some difficulty; with much difficulty; unable to do a task in that area and with or without aids or devices. The HAQ-DI score ranges from 0 (no disability) to 3 (completely disabled), when 6 or more categories are non-missing. Detailed algorithm for calculating HAQ-DI score is described in Appendix 3.

6.3.4. Subject's Pain Assessment

HAQ-DI also includes a separate pain assessment and subject will be requested to mark the severity of the pain in the past week on a 0-100 mm VAS, with 0 indicating "no pain" and 100 indicating "severe pain".

6.3.5. High-Sensitivity C-reactive Protein (hsCRP [mg/L])

High-Sensitivity C-reactive Protein (hsCRP) is a very important laboratory value. It's an element of many efficacy endpoints, including ACR20/50/70, ACR-N, DAS28(CRP), SDAI, EULAR response and Boolean remission.

6.3.6. ACR20

A subject achieves ACR20 response when this subject has

- ≥ 20% improvement from parent study baseline in TJC68, AND
- ≥ 20% improvement from parent study baseline in SJC66, AND
- ≥ 20% improvement from parent study baseline in at least 3 of the following 5 items:

- PGA
- SGA
- Subject's pain assessment
- HAQ-DI score
- hsCRP

Percent improvement from parent study baseline at a postbaseline visit is calculated as follows for all 7 components mentioned above:

$$\%improvement = \frac{parent\ study\ baseline\ value - postbaseline\ value}{parent\ study\ baseline\ value} \times 100$$

If the parent study baseline value is 0 then the percent improvement from parent study baseline is set to missing.

In the case that some ACR20 components are missing, the ACR20 assessment will be based on the non-missing components. If non-missing components are not sufficient to determine ACR20 response, then the ACR20 response will be considered as missing.

6.3.7. ACR50 and ACR70

ACR50 and ACR70 are similarly defined as ACR20 (see section 6.3.6.), except the improvement threshold from parent study baseline is 50% and 70%, respectively

6.3.8. DAS28(CRP)

The DAS28(CRP) score is calculated as follows:

$$DAS28(CRP) = 0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.36\ln(CRP + 1) + 0.014 \times SGA + 0.96,$$

Where;

CRP = CRP measurement (mg/L)

SGA = subject's global assessment of disease activity on a 0-100 mm VAS

TJC28 = tender joint count based on 28 joints

SJC28 = swollen joint count based on 28 joints

Higher DAS28(CRP) value indicates more severe disease activity.

No component-level imputation will be performed for the calculation of DAS28(CRP) for the analyses. If any components are missing, the DAS28(CRP) will be set as missing.

6.3.9. SDAI and CDAI

Simplified Disease Activity Index (SDAI) is a composite measure that sums the TJC28, SJC28, the SGA on a 0-10 cm scale, the PGA on a 0-10 cm scale, and the hsCRP (in mg/dL). SDAI is scored as follow {Aletaha 2005}:

SDAI = TJC28 + SJC28 + SGA + PGA + CRP

Higher SDAI score indicates more severe disease activity status.

Clinical Disease Activity Index (CDAI) is a further simplification of the SDAI that excludes the CRP, which is calculated using the following formula {Aletaha 2005}:

CDAI = TJC28 + SJC28 + SGA + PGA

CDAI can range from 0 to 76, with higher score indicating more severe disease activity status.

No component-level imputation will be performed for the calculation of both SDAI and CDAI. If any components are missing, the SDAI and CDAI will be set as missing.

6.3.10. ACR-N

ACR-N is defined as the smallest percentage improvement from parent study baseline in swollen joints, tender joints and the median of the following 5 items (PGA, SGA, Patient pain assessment, HAQ-DI score and CRP). It has a range between 0 and 100%. In particular,

ACR-N = min {improvement in TJC68 (%), improvement in SJC66 (%), median [improvement in SGA (%), improvement in PGA (%), improvement in pain assessment (%), improvement in HAQ-DI (%), improvement in CRP (%)]}.

If this calculation results in a negative value, then the ACR-N is set to 0. If any components are missing, the ACR-N will be set as missing.

6.3.11. EULAR Response

Subject's response will be categorized according to the following table based on the DAS28(CRP).

Table 6-2. EULAR Response Criteria

DAS28(CRP) at Visit	DAS28(CRP) Improvement from Parent Study Baseline		
	> 1.2	**> 0.6 and ≤ 1.2**	**≤ 0.6**
≤ 3.2	Good Response	Moderate Response	No Response
> 3.2 and ≤ 5.1	Moderate Response	Moderate Response	No Response
> 5.1	Moderate Response	No Response	No Response

6.3.12. Boolean Remission 1.0

Boolean remission is achieved when TJC28 ≤ 1 and SJC28 ≤ 1 and CRP ≤ 1 mg/dL and SGA ≤ 1 (on a 0-10 cm scale). Boolean remission is not achieved when at least 1 of the 4 components has a value > 1. In case some components are missing, Boolean remission is not achieved if at least 1 of the non-missing components has a value > 1. If non-missing components are not sufficient to determine Boolean remission, then Boolean remission will be considered as

missing.

6.3.13. Boolean Remission 2.0

Boolean remission is achieved when TJC28 ≤ 1 and SJC28 ≤ 1 and CRP ≤ 1 mg/dL and SGA ≤ 2 (on a 0-10 cm scale). Boolean remission is not achieved when at least 1 of the 4 components has a value greater than the threshold. In case some components are missing, Boolean remission is not achieved if at least 1 of the non-missing components has a value above the threshold. If non-missing components are not sufficient to determine Boolean remission, then Boolean remission will be considered as missing.

6.3.14. 36-Item Short-form Health Survey (SF-36)

The SF-36 is a 36-item, self-reported, generic, comprehensive, and health-related quality of life questionnaire that yields an 8 health domains (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, and mental health). Each domain is scored by summing the individual items and transforming the scores into a 0 to 100 scale with higher scores indicting better health status or functioning, In addition, 2 summary scores, PCS score and the MCS score will be evaluated based on the 8 SF-36 domains.

6.3.15. FACIT-Fatigue

The FACIT-Fatigue scale is a brief, 13-item, symptom-specific questionnaire that specifically assesses the self-reported severity of fatigue and its impact upon daily activities and functioning in the past 7 days. The FACIT-Fatigue uses 0 (“not at all”) to 4 (“very much”) numeric rating scales. Negatively stated items are reversed by subtracting the response from “4” before being added to obtain a total score.

FACIT-Fatigue ranges from 0 to 52 with higher score indicating less fatigue. In the case of missing response for some items in the questionnaire, if at least half of the items (ie, ≥ 7 of 13 items) were answered at a given visit, the prorated score will be calculated and used in the analysis.

[redacted]

6.4. Analysis of the Efficacy Endpoints

6.4.1. Analysis of the Key Efficacy Endpoints

For binary efficacy endpoints, both OC and NRI data will be summarized using descriptive statistics (counts and proportions of subjects) by treatment and visit. For continuous efficacy endpoints, OC data will be summarized using descriptive statistics (sample size, mean, SD,

median, Q1, Q3, minimum and maximum) by treatment and visit.

For binary endpoint of ACR20/50/70, LDA/Remission based on DAS28(CRP)/CDAI/SDAI/ Boolean remission 1.0 and 2.0:

> To compare filgotinib 200 mg group (with or without prior filgotinib exposure, and combined) versus filgotinib 100 mg group (with or without prior filgotinib exposure, and combined) in proportion of subjects who achieve endpoint at all study visit, a logistic regression analysis with treatment groups and stratification factors in the model will be used. The model will include filgotinib 200 mg group and filgotinib 100 mg group in three scenarios: (1) with prior filgotinib exposure; (2) without prior filgotinib exposure; (3) scenarios (1) and (2) combined. The nominal p-value from the logistic regression model for comparison of filgotinib 200 mg to filgotinib 100 mg will be provided. In each of the three scenarios, when fitting the model, only data of the two comparison groups are included. Firth logistic regression is used when quasi-complete separation occurs.
>
> The 2-sided 95% confidence interval (CI) of the endpoint rate at study visit based on normal approximation with the sample variance will be provided for each treatment group. The 2-sided 95% confidence interval (CI) of the endpoint rate difference between filgotinib 200 mg (with or without prior filgotinib exposure, and combined) versus filgotinib 100 mg (with or without prior filgotinib exposure, and combined) at study visit based on normal approximation with the sample variance will be provided.

For continuous endpoint of change from parent baseline in HAQ-DI, SDAI, CDAI, DAS28(CRP), PGA, SGA, subject's pain score, TJC28/TJC68/SJC28/SJC66, FACIT-Fatigue, [redacted]:

> The change from parent study baseline in the continuous endpoints at a postbaseline visit will be analyzed using the mixed-effects model for repeated measures (MMRM) that includes data at postbaseline visits up to the maximum time point where there are still enough data guarantee the model convergence. Subjects that have a baseline value and at least 1 postbaseline value are included in the analysis. The MMRM models will be used to evaluate treatment effect on change score from parent study baseline, with parent study baseline value, stratification factors, treatment, visit, and treatment by visit interaction included as fixed effects and subject being the random effect. The MMRM model will include two comparison treatment groups of filgotinib 200 mg and filgotinib 100 mg in three scenarios: (1) with prior filgotinib exposure; (2) without prior filgotinib exposure; (3) scenarios (1) and (2) combined. In each of the three scenarios, when fitting the model, all data of all groups are included. An unstructured variance-covariance matrix will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Missing change scores will not be imputed using the MMRM approach.
>
> The least squares (LS) means, 95% CIs of the difference and nominal p-value of the comparison in mean change from parent study baseline between filgotinib 200 mg and filgotinib 100 mg in the three scenarios from MMRM will be provided up to the maximum visit where there are enough data to support MMRM convergency.

Plots of proportions of subjects for binary endpoints (ACR20/50/70, LDA/Remission based on

DAS28(CRP)/CDAI/SDAI/Boolean remission 1.0 and 2.0) and mean ± SD for continuous endpoints (ACR-N, DAS28(CRP), HAQ-DI, CDAI, SDAI, TJC28/TJC68/SJC28/SJC66, PGA, SGA, subject's pain score, hsCRP, FACIT-Fatigue, SF-36) both change from parent baseline and actual values, will be presented over time by treatment and by visit.

For efficacy endpoints, change from baseline refers to change from parent study baseline.

6.4.2. Subgroup Analysis

The subgroup defined for efficacy endpoints is also based on values on parent study baseline.

Subgroup :

- Age Group (< 65 or ≥ 65 years old)
- Age Group (< 75 or ≥ 75 years old)
- BMI Group (<25, 25≤ to <30, ≥30)
- Weight (<60kg,≥ 60kg to <100kg, ≥100kg)
- Baseline Disease Activity (DAS28(CRP) ≤5.1 or >5.1)
- Race
- Region
- Parent study (GLPG0634-cl-301, GLPG0634-cl-302 and GLPG0634-cl-303)

Efficacy Endpoints of interest to apply subgroup:

- ACR20/50/70
- DAS28(CRP) ≤ 3.2
- DAS28(CRP) < 2.6
- CDAI ≤ 2.8
- CDAI ≤ 10
- SDAI ≤ 3.3 (subgroup by parent study only)
- Boolean remission 1.0 (subgroup by parent study only)
- Boolean remission 2.0 (subgroup by parent study only)
- HAQ-DI ≤ 0.5

- Subject's pain assessment (VAS) ≤ 20 mm (subgroup by parent study only)

For binary endpoint of ACR20/50/70, DAS28(CRP) ≤3.2 and 2.6, CDAI ≤2.8, and CDAI ≤10, Boolean remission 1.0 and 2.0, HAQ-DI ≤ 0.5, Subjects Pain Assessment (VAS) ≤ 2.0:

Subgroup analyses comparing filgotinib 200 mg group (with or without prior filgotinib exposure, and combined) to filgotinib 100 mg group (with or without prior filgotinib exposure, and combined) will be performed at all study visit for the endpoints. Nominal p-value for treatment group comparison in proportion of subjects who achieve the endpoints will be obtained using the Fisher's exact test. The number and percentage of subjects who achieve the endpoints will also be provided for each treatment group within the subgoups.

The 2-sided 95% confidence interval (CI) of the endpoint rate at study visit based on normal approximation with the sample variance will be provided for each treatment group. The 2-sided 95% confidence interval (CI) of the endpoint rate difference between filgotinib 200 mg (with or without prior filgotinib exposure, and combined) versus filgotinib 100 mg (with or without prior filgotinib exposure, and combined) at study visit based on normal approximation with the sample variance will be provided.

The above subgroup analyses will be provided for the Safety Analysis Set.

The subgroup analyses by parent study will be provided for the Parent Studies Analysis Set only. For these subgroup analyses, no distinction will be made between filgotinib 100 and 200 mg. The analyses will be made on the pooled treatment group and hence no comparison between treatment arms will be performed.

6.4.3. SF-36

The transformed scores and change from parent study baseline in the 8 domains and 2 summary component scores (PCS and MCS) of the SF-36 will be summarized at each visit by treatment group using descriptive statistics (sample size, mean, SD, median, Q1, Q3, minimum and maximum).

6.4.4. FACIT-Fatigue

Summary statistics for the actual score and change from parent study baseline in FACIT-Fatigue scale score at each postbaseline visit will be presented by treatment group.

[redacted]

7. SAFETY ANALYSES

7.1. Adverse Events and Deaths

7.1.1. Adverse Event Dictionary

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

7.1.2. Adverse Event Severity

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

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE CRF to the question of "Related to Study Treatment." Relatedness will always default to the investigator's choice, not that of the Medical Monitor. Events for which the investigator did not record relationship to study drug will be considered related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing.

If the relationship to study drug is captured multiple times (for multiple study drugs) for the same AE, the worst-case approach will be used to determine the relationship to study drug. For examples if relationship to one or more study drug is 'related' for a certain AE, then derived relationship will be considered "related". The same approach will be applied when deriving premature discontinuation (permanent and temporal) of study drug due to AE, and premature discontinuation of the study due to AE.

7.1.4. Serious Adverse Events

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

7.1.5. Treatment-Emergent Adverse Events

7.1.5.1. Definition of Treatment-Emergent Adverse Events

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

- Any AEs with an onset date on or after the study drug start date and no later than the minimum of [30 days after the last dose date, the date of last contact, the database cut-off

date], or any AE with an onset date on or after the study drug start date. See handing of missing last dose date in Section 4.2.1.

- Any AEs leading to premature discontinuation of study drug

The above definition of TEAEs is used for all analyses on the Safety Analysis Set. However, for analyses using the Parent Studies Analysis Set, TEAEs are all AEs listed above plus those AEs considered as TEAEs during the parent studies.

7.1.5.2. Incomplete Dates

If the onset date of the AE is incomplete and the AE stop date is not prior to the first dosing date of study drug, then the month and year (or year alone if month is not recorded) of onset determine whether an AE is treatment emergent. The event is considered treatment emergent if both of the following 2 criteria are met:

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

- The AE onset date is the same as or before the month and year (or year) of the date corresponding to the minimum of [30 days after the date of the last dose of study drug, the date of last contact, the date of the database cut-off].

The death date is the minimum of the non-missing death dates in parent and extension studies. In case this date is incomplete, impute it to be as early as possible using available information and then take the maximum of this imputed date and the last contact date. Note that a death date should be derived for any subject with a fatal AE or death page in the CRF. In the case of death information available but no death date in SDTM, death date will be assigned as the last contact date.

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

All TEAEs will be allocated to the analysis treatment period by default. Non-TEAE that occurred after the analysis treatment period will be allocated in a post-treatment period. Only TEAE will be summarized, however, all AEs (including those in the post-treatment period) will be listed:

Table 7-1. Adverse Events Analysis Period

Analysis Period	Start Analysis Period	End Analysis Period
Treatment	First dose date.	Last dose date + 30 days

Post-Treatment#	End date of the preceding analysis period + 1 day.	Date of last contact

If the date of last study drug administration + 30 days is after date of last contact, then a post-treatment period will not be included for that participant. The last analysis period will end on the date of last contact.

7.1.6. Exposure Adjusted Incidence Rate and Exposure Adjusted Event Rate

Exposure adjusted incidence (EAIR) and exposure adjusted event rates (EAER) will be calculated for TEAEs.

The EAIR for a TEAE is defined as:

Incidence rate per 100 censored patient years of exposure (cPYE) =
(Total number of subjects with an event/Total cPYE) x 100

For the calculation of EAIR, censored PYE (cPYE) will be calculated, which means that in case of the occurrence of an event, only exposure until the first occurrence of the event will be taken into account. This will be done by calculating for the event of interest, for each treatment the total censored patient years of exposure (cPYE) to a treatment by taking the sum of the individual subject's cPYE (censored in case of an event) within a treatment. Individual subject's cPYE per treatment is defined as:

- For subjects with an event within the treatment:
  cPYE = (Event start date – analysis period start date + 1) / 365.25

  To calculate cPYE in case of incomplete or missing start dates of the event:
  - Incomplete start date: Use the maximum date of the first day/month of the year and the start of the analysis period
  - Missing start date: Use the start of the analysis period.

- For subjects with no event within the treatment:
  cPYE = (analysis period end date – analysis period start date + 1) / 365.25

In addition, for the analyses based on the Parent Studies Analysis Set, EAIR will be calculated for all TEAEs from both the parent and the LTE study combined. cPYE will be calculated based on exposure to filgotinib (pooled over the doses) during both the parent study and the LTE.

The EAER for a TEAE is defined as:

Event rate per 100 patient years of exposure (PYE) =
(Total number of events/Total PYE) x 100

For the calculation of EAER, and the total patient years of exposure (PYE) in the column headers, the total patient years of exposure (PYE) to a treatment or treatment subgroup combination will be calculated as the sum of the individual subject's PYE within a treatment. Individual subject's PYE per treatment is defined as:

PYE = (analysis period end date – analysis period first dose date + 1) / 365.25

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

(X^2_2 (Total number of subjects with an event),0.025/ (2*Total cPYE), X^2_2 (1+Total number of subjects with an event)0.975/ (2*Total cPYE)) *100

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

(X^2_2 (Total number of events),0.025/ (2*Total PYE), X^2_2 (1+Total number of events),0.975/ (2*Total PYE)) *100

The EAIR or EAER will be calculated and exact CIs based on Poisson distribution will be calculated using the method described in {Ulm 1990}, for each treatment group: (1) with prior filgotinib exposure; (2) without prior filgotinib exposure; (3) scenarios (1) and (2) combined. In each of the three scenarios, only data of the two comparison groups are included.

7.1.7. Summaries of Treatment-emergent Adverse Events and Deaths

Treatment-emergent TEAEs will be summarized based on the Safety Analysis Set unless specified otherwise. The analyses will be based on data from GLPG0634-cl-304 only, unless specified otherwise.

7.1.7.1. Summaries of TEAEs

The number and percentage of subjects and EAIR who experienced at least 1 TEAE will be provided and summarized by SOC, HLT, PT, and treatment group. For other TEAEs described below, summaries will be provided by (1) SOC, PT, and treatment group and (2) PT and treatment group:

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

- All Deaths
- All TEAEs leading to temporary interruption of any study drug

The safety analysis of EAIR will be performed for all TEAEs. The safety analysis of EAER will be performed only for AEIs of all infections and serious infections.

A brief, high-level summary of TEAEs described above will be provided by treatment group and by the number and percentage of subjects who experienced the above TEAEs. All deaths occurred in the study regardless of TE will be also included in this summary. This summary will be based on the safety population. In addition, a similar summary will be provided based on the Parent Studies Analysis Set, combining exposure and AE data from both the parent and LTE studies.

Multiple events will be counted only once per subject in EAIR 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, treatment-related TEAEs and AEIs will be summarized by PT only, in descending order of total frequency.

The following summary statistics will be provided for all safety analyses:

A) The number of subjects within treatment group

B) The number of subjects with specific event within treatment group

C) Raw cumulative proportion of specific event within treatment group

D) The number of exposure years within treatment group as specified in Section 7.1.6

E) EAIRs or EAERs with 95% CI within treatment group as specified in Section 7.1.6

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

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

- AEs leading to premature discontinuation of study
- AEs leading to temporary interruption of any study drug

7.1.8. Adverse Events of Interest

AEIs will be identified by the use of either standardized MedDRA queries (SMQs) or MedDRA search term (MST) listings. However, should additional cases not detected by the predefined search term listings be identified during the review process, these cases will also be reported in the respective category. The safety analysis of EAIRs and EAER (for some specific recurrent AEIs) will be performed.

7.1.8.1. Adjudication Committee for MACE

An independent cardiovascular safety endpoint adjudication committee (CVEAC) was formed to periodically review and adjudicate all potential major adverse cardiovascular events (MACE). MACE is defined as cardiovascular death, non-fatal myocardial infarction and non-fatal stroke.

Potential MACE events for adjudication will be identified using these MSTs:

- All deaths
- Cardiovascular events (meeting serious criteria)
- Myocardial infarction (Narrow)
- Hospitalization for unstable angina
- Transient ischemic attack
- Stroke
- Hospitalization for cardiac failure
- Percutaneous coronary intervention

The CVEAC reviews those potential MACE and related clinical data to determine whether a MACE has developed. The CVEAC's role and responsibilities and the data to be provided to the CVEAC are described in a mutually agreed upon CVEAC charter. The CVEAC charter defines the CVEAC membership, adjudication process, meeting logistics, and meeting frequency. The number and percentage of subjects with positively adjudicated MACE will be summarized by treatment group using the PT. {Gilead Science Inc 2019} The categories of adjudicated MACE include: cardiovascular death/cardiac death, non-fatal myocardial infarction, and non-fatal stroke.

A by-subject listing for all patients who have a potential MACE, and have a positively adjudicated MACE at any time, will be provided.

7.1.8.2. Adjudication Committee for Thromboembolic Events

The same adjudication committee CVEAC to review MACE events also reviewed and adjudicated all potential thromboembolic events. Potential cases of both arterial and venous thromboembolic events, identified using the following SMQ searches, will be adjudicated.

- SMQ: Arterial thrombotic and embolic events (ATE)
- SMQ: Venous thrombotic and embolic events (VTE)
- SMQ: Mixed arterial and venous thrombotic and embolic events

The adjudication committee reviews those potential thromboembolic events and related clinical data to determine retrospectively whether a thromboembolic event (pulmonary embolism (PE), deep vein thrombosis (DVT) and arterial systemic thromboembolic [ASTE]) has occurred. The number and percentage of subjects with positively adjudicated treatment-emergent events will be summarized by treatment group and the category of ASTE or VTE respectively.

7.1.8.3. Other Adverse Events of Interest

In addition to general safety parameters and MACE, safety information on other AEIs will also be analyzed. AEIs will be identified by either laboratory results, SMQs, or sponsor defined MSTs, or a combination of these methods as indicated below.

- All infections (defined as all PTs in the Infections and Infestations SOC)
- Serious infections (defined as all PTs in the Infections and Infestations SOC that are SAEs)
- Infections of interest as defined below:
  a) Herpes zoster
  b) Active tuberculosis
  c) Opportunistic infections
  d) Hepatitis B or C infections
- Malignancy (excluding nonmelanoma skin cancer)
- Nonmelanoma skin cancer
- Gastrointestinal (GI) perforations

The number and percentage of patients with aforementioned AEIs will be provided by the PT for each AEI based on the Safety Analysis Set. In addition, a similar table for AEIs (not by PT) will be provided for the Parent Studies Analysis Set and using exposure and AEs from both the parent and LTE studies. As the MedDRA versions differ between studies no summaries by PT will be provided for this additional analysis. For AEIs, the analyses will be based on the flag for AEIs

created in the parent study analysis datasets.

A by-subject listing for all patients having AEIs at any time will be provided for each AEI.

7.1.9. Subgroup Analysis

The safety analysis of EAIRs will be performed. The following subgroups for AE's are defined based on values on LTE baseline, except that race is still based on parent baseline value.

Subgroup:

- Age Group (< 65 or ≥65 years old)
- Age Group (<75 or ≥75 years old)
- Concurrent Oral Corticosteroid (yes or no)
- Concurrent csDMARD (yes or no)
- Race
- Region

The above subgroup analysis will be conducted for the following summary of TEAEs:

- Overall Summary of TEAEs
- TEAEs by Preferred Term
- Overall Summary of AEIs

7.2. Laboratory Evaluations

Laboratory data collected during the study will be analyzed and summarized using both quantitative and qualitative methods. Summaries of laboratory data will be provided for the Safety Analysis Set and will include data collected up to the last dose of any study drug + 30 days for subjects who have completed the study or permanently discontinued study drug. The analysis will be based on values reported in conventional units. When values are below the LOQ, they will be listed as such, and the closest imputed value will be used for the purpose of calculating summary statistics as specified in Section 3.5.

A by-subject listing for laboratory test results will be provided by subject ID number and visit 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 on the CTCAE severity grade will be flagged in the data listings, as appropriate.

No formal statistical testing is planned.

For lab tests, table, figure and listing will contain all available data up to the last dose date plus 30 days.

7.2.1. Summaries of Numeric Laboratory Results

Descriptive statistics of LTE baseline values, values at each postbaseline visit and change from LTE baseline at each postbaseline visit will be provided by treatment group for the following laboratory tests:

- Hematology
  - Hematocrit
  - Hemoglobin
  - Platelet count
  - Red blood cell count
  - Mean corpuscular volume
  - White blood cell (WBC) count
  - Lymphocytes
  - Monocytes
  - Neutrophils
  - Eosinophils
  - Basophils
- Chemistry
  - Alanine aminotransferase (ALT)
  - Aspartate aminotransferase (AST)
  - Alkaline phosphatase (ALP)
  - Total bilirubin
  - Serum creatinine
  - Creatinine clearance by Cockcroft-Gault formula
  - Creatine phosphokinase (CPK)

  - Glucose

- Lipid
  - Triglycerides
  - Total cholesterol
  - High density lipoprotein (HDL)
  - Low density lipoprotein (LDL)
  - LDL/HDL ratio

An LTE baseline laboratory value will be defined as the last measurement obtained on or prior to the date of first dose of any study drug in the LTE. Change from LTE baseline to a postbaseline visit will be defined as the visit value minus the LTE 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 + 1.

Median (Q1, Q3) of the observed values for the laboratory tests specified above will be plotted using a line plot by treatment group and visit.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.6.3.

7.2.2. Graded Laboratory Value

The CTCAE Version 4.03 (with one modification specified in the protocol) 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.

7.2.2.1. Treatment-Emergent Laboratory Abnormalities

Treatment-emergent laboratory abnormalities are defined as values that increase at least 1 toxicity grade from LTE baseline at any postbaseline time point, up to and including the date of last dose of any study drug + 30 days for subjects who permanently discontinued study drug.

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

7.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

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

date of the last dose of any study drug + 30 days for subjects who permanently discontinued study drug.

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

7.2.2.3. Summaries of Laboratory Abnormalities

Laboratory data that are categorical will be summarized using the number and percentage of subjects and EAIR in the study with the given response at LTE baseline and each scheduled postbaseline visit.

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

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

For all summaries of laboratory abnormalities, the denominator is the number of subjects with non-missing postbaseline values up to 30 days after the last dosing date.

A by-subject listing of treatment-emergent Grade 3 or higher laboratory abnormalities and marked laboratory abnormalities will be provided separately by subject ID number and visit in chronological order. These listings will include all test results that were collected throughout the study for the lab test of interest, with all applicable severity grades or abnormal flags displayed.

A shift table of Creatinine from LTE baseline is also provided.

Fasting lipids (Triglyceride, Cholesterol, HDL, LDL, LDL/HDL ratio) will also be summarized by ATP III classification.

Table 7-2. ATP III Classification of LDL, HDL, Total Cholesterol, Triglyceride, and LDL/HDL ratio

Lab test	Classifications				
LDL (mg/dL)	< 100	[100, 130)	[130, 160)	[160, 190)	≥ 190
HDL (mg/dL)	< 40	[40, 60)	≥ 60		
Total cholesterol (mg/dL)	< 200	[200, 240)	≥ 240		
Triglyceride (mg/dL)	< 150	[150, 200)	[200, 500)	≥ 500	
LDL/HDL ratio	< 2.5	[2.5, 3.3]	> 3.3		

7.2.3. Laboratory Evaluations of Special Interest

7.2.3.1. Liver-related Laboratory Evaluations

Liver-related abnormalities after initial study drug dosing will be examined and summarized using the number and percentage of subjects who were reported to have the following laboratory test values for postbaseline measurements:

- AST: (a) > 3 times of the upper limit of reference 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
- ALT or AST: (a) > 3 x ULN; (b) > 5 x ULN; (c) > 10 x ULN; (d) > 20 x ULN
- Total bilirubin > 2 x ULN
- ALP > 1.5 x ULN
- AST or ALT > 3 x ULN
  - Total bilirubin > 1.5 x ULN
  - Total bilirubin > 2 x ULN

The summary will include data from all postbaseline visits up to 30 days after the last dose of any study drug. For individual laboratory tests, subjects will be counted once based on the most severe postbaseline values. For both the composite endpoint of AST or ALT and total bilirubin, subjects will be counted once when the criteria are met at the same postbaseline visit date. The denominator is the number of subjects in the Safety Analysis Set who have nonmissing postbaseline values of all relevant tests at the same postbaseline visit date. A listing of subjects who met at least 1 of the above criteria will be provided.

7.2.3.2. Complete Blood Count-Related Laboratory Evaluations

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

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

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

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

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

The summary will include data from all postbaseline visits up to 30 days after the last dose of any study drug.

7.3. Body Weight and Vital Signs

Descriptive statistics will be provided by treatment group for body weight, BMI and vital signs (systolic and diastolic blood pressures [mmHg], pulse [beats/min], respiration [breaths/min]) as follows:

- LTE baseline value

- Values at each post LTE baseline visit

- Change from LTE baseline at each postbaseline visit

A LTE baseline value will be defined as the last available value collected on or prior to the date of first dose of any study drug in LTE. Change from LTE baseline to a postbaseline visit will be defined as the postbaseline value minus the LTE baseline value. Body weight and vital signs measured at unscheduled visits will be included for the LTE baseline value selection.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.6.3. No formal statistical testing is planned.

A by-subject listing of vital signs (systolic and diastolic blood pressure [mmHg], pulse [beats/min], respiration [breaths/min], and body temperature [°C]) will be provided by subject ID number and visit in chronological order. In the same manner, a by-subject listing of body weight, height, and BMI will be provided separately.

7.4. Prior and Concomitant Medications

All continuing concomitant medications from the parent studies of participation are documented in the eCRF.

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

All the analyses in this section will be performed for general prior /concomitant medications and RA-specific prior /concomitant medications separately, unless otherwise specified.

For RA-specific prior / concomitant medications, below grouping is used:

- Nonsteroidal anti-inflammatory drugs (NSAIDs) (including cyclooxygenase-2 (COX-2) inhibitor)
- Opioids
- RA-specific medications, not NSAIDs and not Analgesic

These groupings are selected based on ATC Class 3 and 4 (see Appendix 1)

- Nonsteroidal anti-inflammatory drugs (NSAIDs)
- Drugs used in pain therapies: Analgesic drugs
- Analgesia producing opioids

7.4.1. Prior Medications

Prior medications are defined as any medication taken before a subject took the first study drug.

Prior medications will be summarized by Anatomical Therapeutic Chemical (ATC) drug class and preferred name using the number and percentage of subjects for each treatment group and overall. A subject reporting the same medication more than once will be counted only once when calculating the number and percentage of subjects who received that medication. The summary will be provided by preferred name in order of descending overall frequency. For drugs with the same frequency, sorting will be done alphabetically.

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

Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

7.4.2. Concomitant Medications

Concomitant medications are defined as medications taken while a subject took study drug. Use

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

For the purposes of analysis, any medication with a start date prior to or on the first dosing date of any LTE study drug and continued to take after the first dosing date, or started after the first dosing date but prior to or on the last dosing date of study drug will be considered concomitant medications. Medications started and stopped on the same day as the first dosing date or the last dosing date of any study drug will also be considered concomitant. Medications stopped on the same day as the first dosing date will be considered concomitant. Medications with a stop date prior to the date of first dosing date of any study drug or a start date after the last dosing date of any study drug will be excluded from the concomitant medication summary. If a partial stop date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the date of first study drug administration will be excluded from the concomitant medication summary. If a partial start date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing start and stop dates will be included in the concomitant medication summary, unless otherwise specified.

Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

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

7.5. Electrocardiogram Results

7.5.1. Investigator Electrocardiogram Assessment

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

7.6. Other Safety Measures

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

7.7. Changes From Protocol-Specified Safety Analyses

Depending on resource and timing, a by-subject listing may be provided by subject identification (ID) number in ascending order to show:

- Reasons for premature study drug or study discontinuation due to COVID-19
- Missing or virtual visits due to COVID-19

- Important protocol deviations related to COVID-19

Premature study drug or study discontinuation due to COVID-19 is identified either with reason of AE, and AE is COVID-19 related, or with reason of other than AE, and description contains key word “COVID”.

The determination of missing or virtual visits due to COVID-19 was done using Natural Language Processing (NLP) to search the CRF comment fields. A detailed explanation of the algorithm is given in Appendix 4.

IPD related to COVID-19 is identified by searching description with key word “COVID”.

AE’s with preferred term equal to one the following terms will be listed. “Suspected COVID-19”, “COVID-19”, “COVID-19 pneumonia” or “Asymptomatic COVID-19”.

8. REFERENCES

Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. Arthritis research & therapy 2005;7 (4):R796-806.

Buch MH, Silva-Fernandez L, Carmona L, et al. Development of EULAR Recommendations for the Reporting of Clinical Trial Extension Studies in Rheumatology. Ann. Rheum. Dis 2005; 74:963-969.

Gilead Science Inc. Cardiovascular Event Adjudication Committee (CVEAC) Charter. Version 2. August. 2019.

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.

Liu L, Gilead Science Inc. Data Monitoring Committee (DMC) Charter. A Randomized, Double-blind, Placebo-controlled, Multicenter, Phase 3 Study to Assess the Efficacy and Safety of Filgotinib Administered for 24 weeks in Combination with Conventional Synthetic Disease-modifying Anti-rheumatic Drug(s) (csDMARDs) to Subjects with Moderately to Severely Active Rheumatoid Arthritis Who Have an Inadequate Response to Biologic DMARD(s) Treatment. Study Number: GS- US-417-0302. Version 1.0. 31 March. 2017.

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

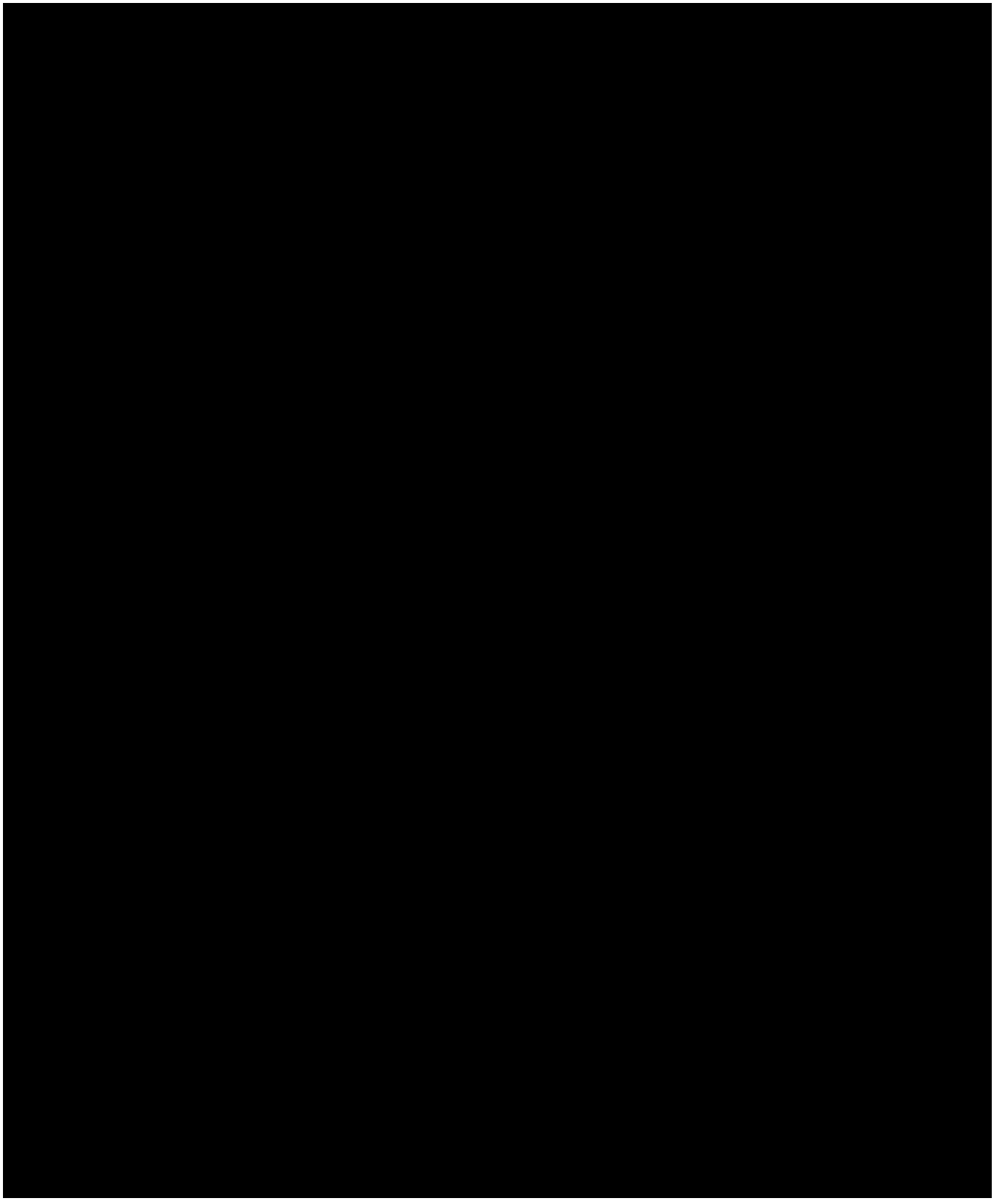
Ye L, Gilead Science Inc. Data Monitoring Committee (DMC) Charter. A Randomized, Double-blind, Placebo- and Activecontrolled, Multicenter, Phase 3 Study to Assess the Efficacy and Safety of Filgotinib Administered for 52 weeks in Combination with Methotrexate to Subjects with Moderately to Severely Active Rheumatoid Arthritis Who Have an Inadequate Response to Methotrexate. Study Number: GS- US-417-0301. Version 2.0. 02 April. 2019.

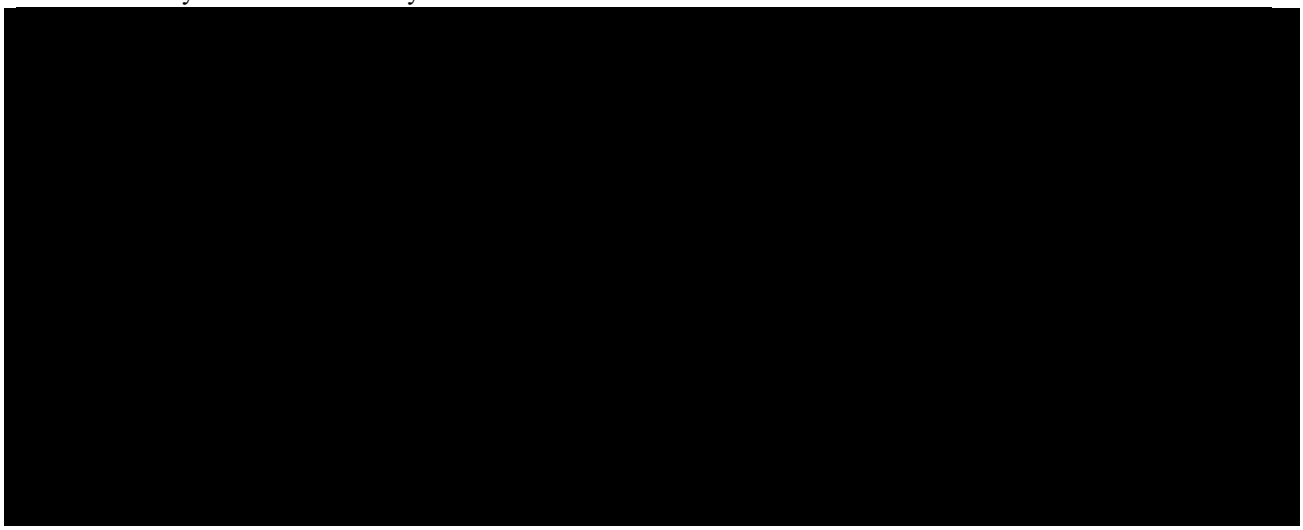
Yin Z, Gilead Sciences Inc. Data Monitoring Committee (DMC) Charter. A Randomized, Double-blind, Placebo- and Activecontrolled, Multicenter, Phase 3 Study to Assess the Efficacy and Safety of Filgotinib Administered for 52 Weeks Alone and in Combination with Methotrexate (MTX) to Subjects with Moderately to Severely Active Rheumatoid Arthritis Who Are Naïve to MTX Therapy. Study Number: GS-US-417-0303. Version 2.0. 02 April. 2019.

9. SOFTWARE

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

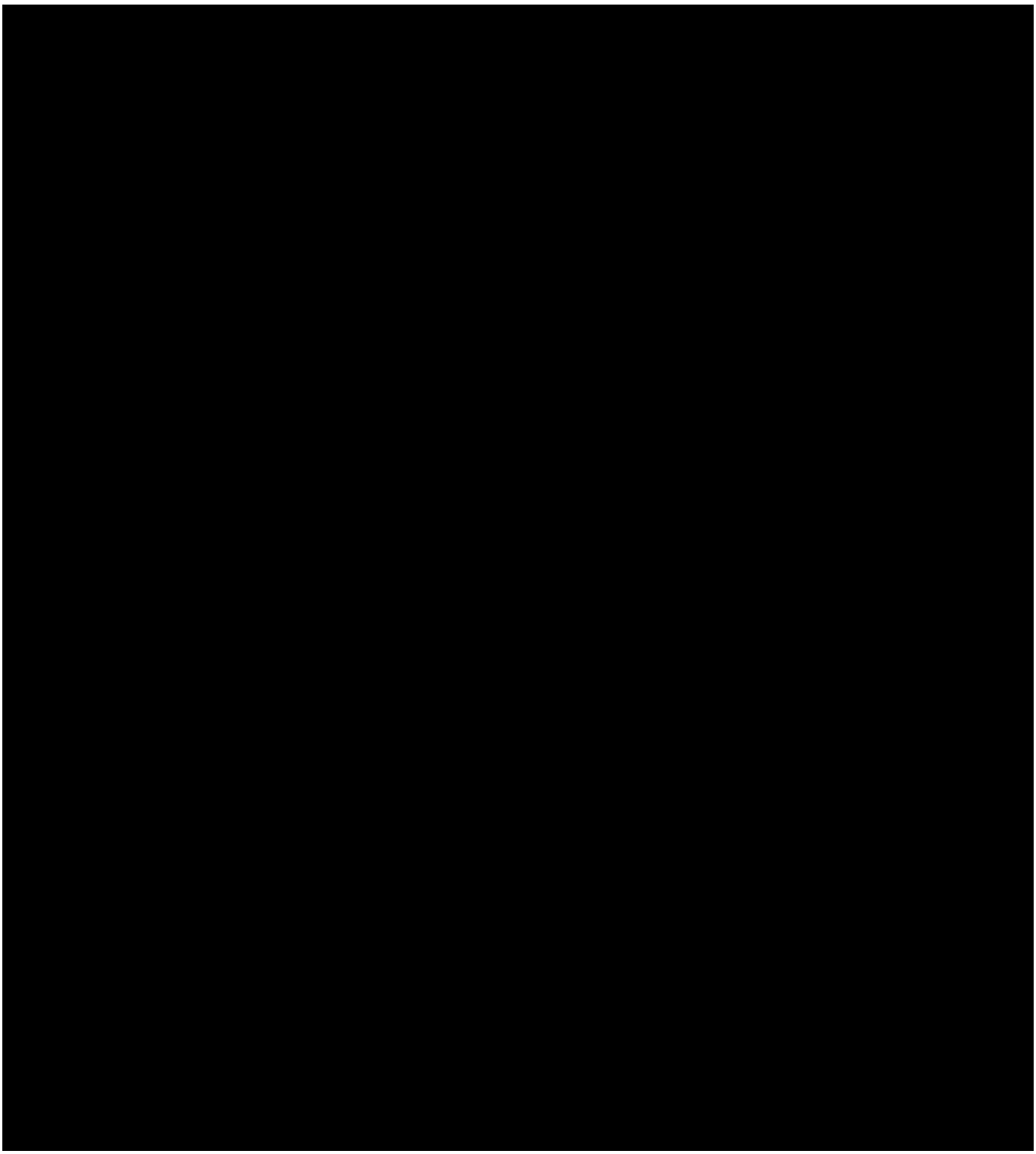
10. SAP REVISION





11. APPENDICES

Appendix 1



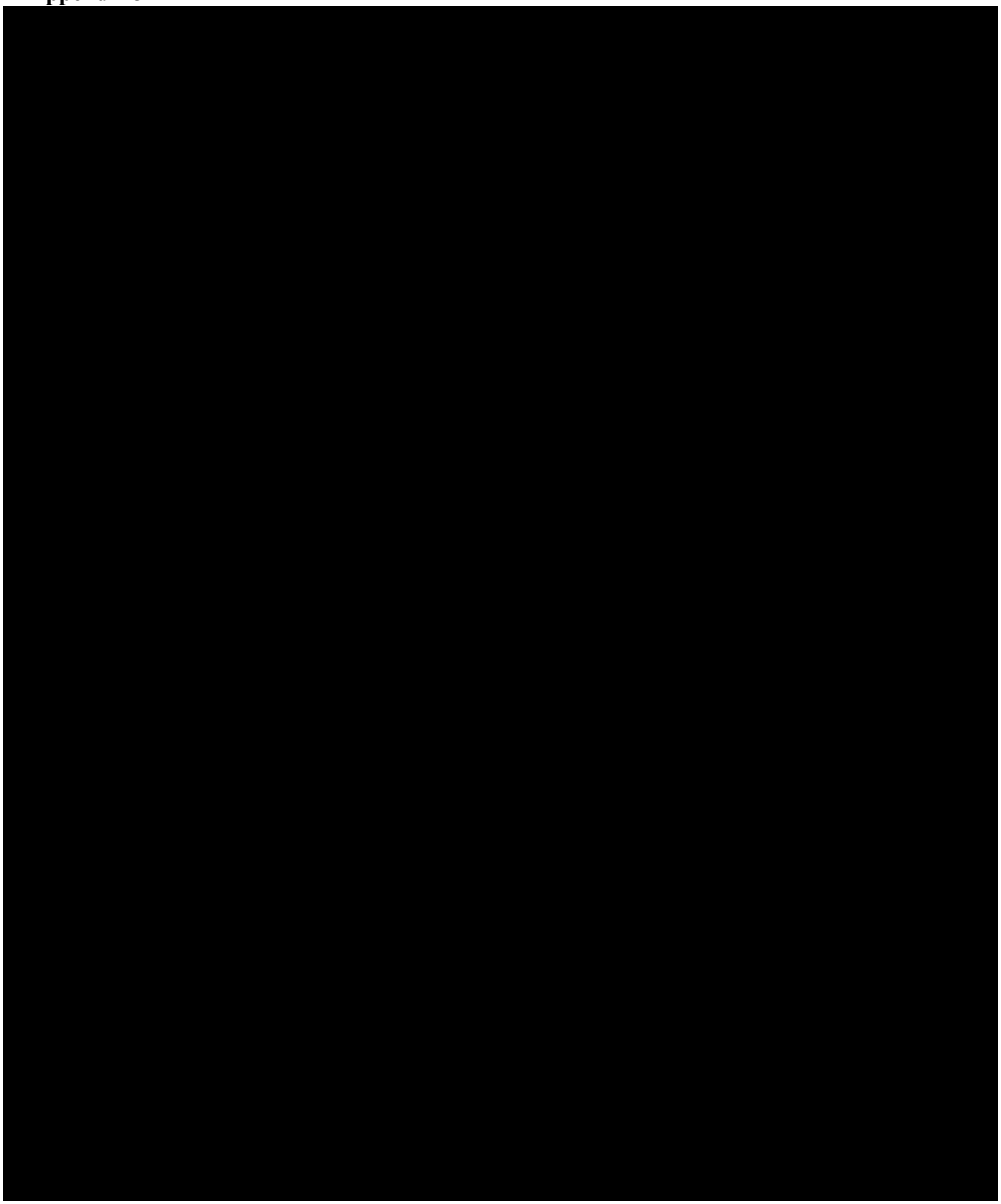
[redacted]

[redacted]

Appendix 2

[redacted]

Appendix 3



Appendix 4