STATISTICAL ANALYSIS PLAN

Study Title:	A multicenter, open-label, long-term follow-up safety and efficacy study of filgotinib treatment in subjects with moderately to severely active rheumatoid arthritis
Name of Test Drug:	Filgotinib
Study Number:	GLPG0634-CL-205
Protocol Version (Date):	Amendment 3: 15 March 2018
Analysis Type:	Final Analysis
Analysis Plan Version:	2.0
Analysis Plan Date:	24 January 2019
Analysis Plan Author(s):	[redacted]

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF ABBREVIATIONS

ACR	American College of Rheumatology
ACR20/50/70	American College of Rheumatology 20/50/70% improvement
AE	adverse event
AESIs	adverse events of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical drug class
bDMARD	biological disease modifying anti-rheumatic drug
b.i.d.	bis in die (twice a day)
BMI	body mass index
CBC	complete blood count
CCP	cyclic citrullinated peptide
CDAI	clinical disease activity index
CRF	case report form
CRP	c-reactive protein
CSR	clinical study report
CTCAE	Common Toxicity Criteria for Adverse Events
DAS28	disease activity score for 28 joint count
DMC	data monitoring committee
ECG	electrocardiogram
EULAR	European League Against Rheumatism
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
FAS	full analysis set
GI	Gastrointestinal
HAQ-DI	health assessment questionnaire-disability index
HLGT	high-level group term
HLT	high-level term
HRQoL	health-related quality of life
ID	identification
LLT	lower-level term
LOQ	limit of quantitation
MedDRA	Medical Dictionary for Regulatory Activities
MST	MedDRA search listing
MTX	methotrexate
NRI	non-responder imputation
OC	observed case
PGA	physician’s global assessment

PT	preferred term
PTM	placebo to match
PVE	pharmacovigilance & epidemiology department
Q1, Q3	first quartile, third quartile
q.d.	quaque die (each day)
RA	rheumatoid arthritis
RF	rheumatoid factor
SAE	serious adverse event
SAP	statistical analysis plan
SD	standard deviation
SDAI	simplified disease activity index
SF-36	36-item short form survey
SGA	subject's global assessment
SJC	swollen joint count
SMQ	standardized MedDRA query
SOC	system organ class
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
TJC	tender joint count
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 clinical study report (CSR) for study GLPG0634-CL-205. This SAP is based on the study protocol amendment 3 dated 15 March 2018 and the electronic case report form (eCRF). The SAP will be finalized before database finalization. Any changes made after the finalization of the SAP will be documented in the CSR.

1.1. Study Objectives

The primary objective of this study is as follows:

- To evaluate the long term safety and tolerability of filgotinib for the treatment of rheumatoid arthritis (RA)

The secondary objectives of this study are as follows:

- To evaluate the long-term efficacy of filgotinib
- To evaluate the long-term effects of filgotinib administration on subjects' disability, fatigue, and quality of life

GLPG0634-CL-205 data and the most current GLPG0634-CL-203 and GLPG0634-CL-204 data will be used in the analyses.

1.2. Study Design

This is a multicenter, open-label, long-term follow-up study in subjects with RA. An approximate total of 600 subjects will be enrolled in the study after they have completed one of the previous 2 core studies: (GLPG0634-CL-203 or GLPG0634-CL-204). All subjects will start the study at the same dose level (filgotinib 200 mg daily, administered either as 200 mg q.d. or as 100 mg b.i.d., depending on the dosing regimen (frequency of intake of active IMP) that was administered during the preceding core study). US males were limited to dosing with filgotinib 100 mg q.d. due to an FDA requirement based on a non-clinical testicular toxicity signal.

The investigator may decide to decrease the daily dose of filgotinib to 100 mg q.d. in case of intolerance or safety reasons. Subjects will return to the 200 mg filgotinib daily dose after the reasons for decreasing the dose have resolved, and at the investigator's discretion. Subjects can also continue on filgotinib 100 mg q.d. when deemed necessary by the investigator. Subjects coming from the GLPG0634-CL-204 study will be allowed to start MTX treatment, if deemed necessary according to the investigator's clinical judgment.

Subjects will enter the study at Entry Visit on Day -1, which will occur on the same day as the last visit from the previous core study (GLPG0634-CL-203 or GLPG0634-CL-204). Subjects

participating in the study will be requested to attend Quarterly Evaluation Visits (QEVs) every 12 weeks. If necessary, unscheduled visits can occur in between the scheduled study visits, for instance, in case the dose of study medication is to be adapted. In case of early withdrawal, an Early Discontinuation Visit (EDV) will be scheduled.

The duration of the participation and treatment for any subject in this study is expected to last for several years (from Entry visit to Follow-up Visit) and will be determined by the sponsor based on the following:

- Open label period (approximately 96 months).
- Marketing Application Approval/Withdrawal in Country of Residence.
- Applicable local reimbursement procedures are in place.

Sites will be notified once these criteria are met.

Following marketing, local (if applicable) regulatory and/or pertinent local reimbursement approval of the study medication, subjects should return to their next quarterly scheduled study visit as specified in the protocol. This visit will serve as a Final Visit and the subjects will be considered as having completed the treatment part of the study. At the end of the study, subjects will be requested to attend a Follow-up Visit approximately 2 weeks after the Final Visit.

The design of the study is presented in the following table:

Study 203		
Week 1-12	**Week 13-24**	**Roll over to Study 205**
Randomized to placebo	Responders(*) remain on placebo	Half of them are **rerandomized** to 200 mg q.d.
		Half of them are **rerandomized** to 100 mg b.i.d.
	Nonresponders: half of them are **rerandomized** to 100 mg q.d.	**Assigned** to 200 mg q.d.
	Nonresponders: half of them are **rerandomized** to 50 mg b.i.d.	**Assigned** to 100 mg b.i.d.
Randomized to 50 mg q.d.	Responders **remain** on 50 mg q.d.	**Assigned** to 200 mg q.d.
	Nonresponders **assigned** to 100 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 25 mg b.i.d.	Responders **remain** on 25 mg b.i.d.	**Assigned** to 100 mg b.i.d.
	Nonresponders **assigned** to 50 mg b.i.d.	**Assigned** to 100 mg b.i.d.
Randomized to 100 mg q.d.	**Remain** on 100 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 50 mg b.i.d.	**Remain** on 50 mg b.i.d.	**Assigned** to 100 mg b.i.d.
Randomized to 200 mg q.d.	**Remain** on 200 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 100 mg b.i.d.	**Remain** on 100 mg b.i.d.	**Assigned** to 100 mg b.i.d.
Study 204		Roll over to **Study 205**
Week 1-12	Week 13-24	
Randomized to placebo q.d.	All **assigned** to 100 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 50 mg q.d.	Responders **remain** on 50 mg q.d.	**Assigned** to 200 mg q.d.
	Nonresponders **assigned** to 100 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 100 mg q.d.	**Remain** on 100 mg q.d.	**Assigned** to 200 mg q.d.
Randomized to 200 mg q.d.	**Remain** on 200 mg q.d.	**Remain** on 200 mg q.d.

(*) responder: having a decrease of at least 20% in tender joint count (TJC)68 and swollen joint count (SJC)66 versus Baseline.

1.3. Sample Size and Power

It is expected that around 70% of the subjects in the previous core studies GLPG0634-CL-203 (N=595, approximately) and GLPG0634-CL-204 (N=280, approximately) will roll over into this study, corresponding to a sample size of roughly 600 subjects: 0.7 x (595 subjects in study 203 + 280 subjects in study 204) = 612.

2. TYPE OF PLANNED ANALYSIS

2.1. Data Monitoring Committee Analyses

An external multidisciplinary Data Monitoring Committee (DMC) will review the progress of the study and perform interim reviews of safety data in order to protect subject welfare and preserve study integrity. To ensure the best interests of the participants, the DMC will recommend to the sponsor if the nature, frequency, and severity of adverse effects associated with the study treatment warrant the early termination of the study, the continuation of the study, or the continuation of the study with modifications.

The initial review has been conducted approximately 3 months after the DMC organizational meeting. The 2nd and 3rd DMC meetings have been conducted approximately 3 months after the previous meeting. Subsequent DMC meetings will be scheduled at approximately 12 months intervals following the 3rd DMC meeting.

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

2.2. Pharmacovigilance & Epidemiology (PVE) Analysis

The Pharmacovigilance & Epidemiology (PVE) analysis may be conducted periodically as needed in order to review emerging safety data from GLPG0634-CL-205.

2.3. Interim Analysis for Conference

A subset of efficacy and safety data may be analyzed periodically as needed for the purpose of possible publication.

2.4. Week 156 Analysis

A planned Week 156 analysis will be conducted after all subjects have either completed their Week 156 visit or prematurely discontinued from the study.

2.5. Final Analysis

After all subjects have completed the study, outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned, 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. 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 were excluded and the reasons for their exclusion, will be summarized by treatment group.

3.1.1. Full Analysis Set

The Full Analysis Set (FAS) includes all subjects who enrolled in GLPG0634-CL-205 and received at least one dose of study drug in GLPG0634-CL-205. This is the primary analysis set for efficacy analyses.

The study drug in GLPG0634-CL-205 is filgotinib. The study drugs in the core studies are filgotinib and placebo to match (PTM).

3.1.2. Safety Analysis Set

The Safety Analysis Set includes all subjects who enrolled in GLPG0634-CL-205 and received at least one dose of study drug in GLPG0634-CL-205. This is the primary analysis set for safety analyses.

3.2. Subject Grouping

For analyses based on the FAS subjects will be grouped according to the treatment to which they were randomized or assigned in GLPG0634-CL-205. For analyses based on the Safety Analysis Set, subjects will be grouped according to actual treatment received in GLPG0634-CL-205. The actual treatment received will differ from the randomized treatment only when their actual treatment differs from randomized treatment for the entire GLPG0634-CL-205 treatment duration.

Subjects will be presented by the following groups:

- Filgotinib with MTX
- Filgotinib monotherapy

The filgotinib with MTX group includes subjects who were from GLPG0634-CL-203 and randomized or assigned to filgotinib 100 mg q.d., filgotinib 100 mg b.i.d. or filgotinib 200 mg q.d. in GLPG0634-CL-205. The filgotinib monotherapy group includes subjects who were from GLPG0634-CL-204 and randomized or assigned to filgotinib 100 mg q.d., filgotinib 100 mg b.i.d. or filgotinib 200 mg q.d. in GLPG0634-CL-205.

3.3. Missing Data and Outliers

3.3.1. 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. 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. The handling of missing or incomplete dates for prior medications is described in Section 5.2. and for concomitant medications in Section 7.4. Imputation rules adopted in the efficacy analyses are specified in Section 6.1.

3.3.2. Outliers

Outliers will be identified during the data management and data analysis process, but no sensitivity analyses will be conducted. All data will be included in the data analyses.

3.4. Data Handling Conventions and Transformations

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

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).

3.5. Analysis Visit Window

3.5.1. Definition of Study Day

The first dose date of individual study drug will be calculated separately for each study drug in a treatment group. Study Day 1 is defined as the first dose date of any study drug, which is the minimum of the first dose dates of individual study drugs in a treatment group.

The last dose date of individual study drug will be calculated separately for each study drug in a treatment group. 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.

The first dose date of individual study drug in GLPG0634-CL-205 is identified by the first drug dispense date + 1 day. The last dose date of individual study drug in GLPG0634-CL-205 is identified by the Final Visit date or the Early Discontinuation Visit date.

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

- For postdose study days: Assessment date – Reference Date + 1

- For days prior to the first dose: Assessment date – Reference Date

Reference Date is defined as follows:

- For efficacy data in the core studies: Study Day 1 in the core studies

- For efficacy data in GLPG0634-CL-205: Study Day 1 in GLPG0634-CL-205

- For safety data: The first dose date of filgotinib in the core studies or GLPG0634-CL-205

3.5.2. Analysis Visit Windows

Subject visits may not occur on protocol-specified days. Therefore, for the purpose of analysis, observations will be assigned to analysis windows for efficacy data in the core studies (Table 3-1 and Table 3-2) and GLPG0634-CL-205 (Table 3-3 and Table 3-4), and for non-efficacy data (Table 3-5). The analysis visit windows for non-efficacy data will be applied up to the last dose date of any study drug.

The baseline value is defined as follows:

Efficacy data:

- Core baseline: The last nonmissing value on or prior to the first dose date of study drug in the core studies
- Extension baseline: The last nonmissing value on or prior to the first dose date of study drug in GLPG0634-CL-205, and no earlier than the core study nominal Week 24 visit date

Safety data:

- Filgotinib baseline: The last nonmissing value on or prior to the first dose date of filgotinib in the core studies or GLPG0634-CL-205

Table 3-1. Analysis Visit Windows for Efficacy Data (SF-36 and FACIT-Fatigue Excluded) in the Core Studies

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Core Baseline	1	(none)	1
Core Week 1	8	2	11
Core Week 2	15	12	22
Core Week 4	29	23	43
Core Week 8	57	44	71
Core Week 12	85	72	99
Core Week 16	113	100	127
Core Week 20	141	128	155
Core Week 24	169	156	≥169

Table 3-2. Analysis Visit Windows for SF-36 and FACIT-Fatigue in the Core Studies

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Core Baseline	1	(none)	1
Core Week 4	29	2	57
Core Week 12	85	58	127
Core Week 24	169	128	≥169

Table 3-3. Analysis Visit Windows for Efficacy Data (SF-36 and FACIT-Fatigue Excluded) in GLPG0634-CL-205

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Extension Baseline	1	(none)	1
Extension Week 12	85	2	127
Extension Week 24	169	128	211
Extension Week 36	253	212	295
Extension Week 48	337	296	379
Extension Week 60	421	380	463
Extension Week 72	505	464	547
Extension Week 84	589	548	631
Extension Week W	7xW+1	7xW–40	7xW+43

Table 3-4. Analysis Visit Windows for SF-36 and FACIT-Fatigue in GLPG0634-CL-205

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Extension Baseline	1	(none)	1
Extension Week 48	337	2	505
Extension Week 96	673	506	841
Extension Week W	7xW+1	7xW–166	7xW+169

Table 3-5. Analysis Visit Windows for Non-Efficacy Data

Analysis Visit	Nominal Study Day	Lower Limit	Upper Limit
Filgotinib Baseline	1	(none)	1
Week 12	85	2	127
Week 24	169	128	211
Week 36	253	212	295
Week 48	337	296	379
Week 60	421	380	463
Week 72	505	464	547
Week 84	589	548	631
Week W	7xW+1	7xW–40	7xW+43

CRP measured on or after nominal follow up visit will not be assigned analysis window and summarized, but will be included in the listing.

Vital signs, ECG, and safety laboratory data collected in the follow up period will also be summarized. The analysis window for the post-treatment follow-up period is defined as from (last dose date of filgotinib + 1 day) to (last dose date of filgotinib + 30 days). If multiple valid, nonmissing measurements exist in the post treatment follow-up visit window, the latest record will be selected for analysis. If the chronological order cannot be determined, 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 last dose date plus 30 days will be excluded from the summaries, but will be included in the listings.

3.5.3. 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 1 value per analysis window.

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

- In general, the filgotinib baseline value will be the last nonmissing value on or prior to the first dose date of filgotinib in the core studies or GLPG0634-CL-205, unless otherwise specified. If multiple measurements occur on the same day, the last nonmissing value prior to the time of first dose of filgotinib in the core studies or GLPG0634-CL-205 will be considered as the filgotinib 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 filgotinib baseline value.

- For postbaseline visits:

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

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

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

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

- For filgotinib baseline, the last available record on or prior to the date of the first dose of filgotinib in the core studies or GLPG0634-CL-205 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.

The rules for selecting efficacy data in the event of multiple records in an analysis visit window are specified in Section 6.1.

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, and investigator within a country. 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.

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

A summary of subject disposition will be provided by treatment group. This summary will present the number of subjects in each of the categories listed below:

- Safety Analysis Set
- Full Analysis Set
- Continuing study
- Completed study
- Did not complete study with reasons for premature discontinuation from the study

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

A by-subject listing of reasons for premature study discontinuation will be provided by subject identification (ID) number in ascending order to support the above summary tables.

4.2. Extent of Filgotinib Exposure

Total duration of exposure to filgotinib will be defined as (last dose date of filgotinib in GLPG0634-CL-205 - first dose date of filgotinib in the core studies or GLPG0634-CL-205 + 1), regardless of any temporary interruptions in administration and will be expressed in weeks using up to 1 decimal place (eg, 4.5 weeks).

If the last filgotinib dosing date is missing, the latest date among the clinical visit date and laboratory sample collection date that occurred during the on-treatment period will be used.

If subjects are continuing on study drug, The data cutoff date will be used to impute the last dosing date for the calculation of the duration of exposure to filgotinib.

The total duration of exposure to filgotinib will be summarized using descriptive statistics and using the number (ie, cumulative counts) and percentage of subjects exposed for filgotinib baseline and every 12 weeks afterwards.

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

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. Demographics and Other Core Baseline Characteristics

Subject demographic and other core baseline characteristics 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 and other core baseline characteristics data will be provided for the Safety Analysis Set for the following:

- Age (on the first dose date of any study drug in the core studies)
- Sex at birth (male, female)
- Race
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Geographic region and country (Region 1: Latin America; Region 2: Central and Eastern Europe, EU; Region 3: Central and Eastern Europe, non-EU; Region 4: West and Asia Pac)
- Weight (kg)
- Height (cm)
- Body mass index (BMI; in kg/m^2)

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

5.2. Core Baseline Disease Characteristics

Core baseline disease characteristics include:

- Duration from RA diagnosis (years)

  Calculated as ((first dose date of study drug in the core studies) – (date of initial diagnosis) + 1 day) / 365.25. If the date of initial diagnosis is incomplete, then the following rules will be applied:

  — missing day: use the first of the month

  — missing month: use January

- Presence of RF (Yes/No)

  Positive presence of RF is defined as RF value > ULN

- Presence of anti-CCP (Yes/No)

  Positive presence of anti-CCP is defined as anti-CCP value > ULN

- Prior exposure to bDMARDs (Yes/No)

  Prior exposure to bDMARDs (Please refer to Appendix 1 for a list of bDMARDs) are defined as any bDMARDs taken before a subject took the first study drug in the core studies. A subject reporting the same bDMARDs more than once will be counted only once when calculating the number and percentage of subjects who received bDMARDs.

  For the purposes of analysis, any bDMARDs with a start date prior to the first dosing date of any study drug in the core studies will be counted regardless of when the stop date is. If a partial start date is entered the bDMARDs 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 in the core studies. bDMARDs with a completely missing start date will be counted, unless otherwise specified.

- Concurrent oral corticosteroids use on the first dosing date (Yes/No)

  — Oral corticosteroids dose, mg/day, expressed as prednisone-equivalent dose, mean (SD)

- Concurrent MTX use on the first dosing date (Yes/No)

  — MTX dose, mg/week, mean (SD)

These core baseline disease 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 core baseline characteristics will be provided for the Safety Analysis Set.

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

5.3. Medical History

Medical history collected at Screening in the core studies will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) Version 21.0.

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 summary. 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 FAS, defined in Section 3.1.1.

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 ACR/EULAR remission, we use the following steps unless otherwise specified:

- Step 1: Assign individual components to analysis visit windows defined in Section 3.5.2.
- Step 2: Within each analysis visit window, select the 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 missings will be set as non-responders.

If a subject only has core baseline measurements, OC analyses will not include this subject. But this subject will be treated as non-responder in NRI analyses.

For a specific visit if both regular CRP and high-sensitive CRP are available, only regular CRP will be used.

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
- Actual value and change from core baseline in individual components of the ACR response over time
- Actual value and change from core baseline in Disease Activity Score for 28 Joint Count using CRP (DAS28[CRP)]) over time
- The proportion of subjects who achieve DAS28(CRP) $\leq$ 3.2 over time
- The proportion of subjects who achieve DAS28(CRP) < 2.6 over time
- Actual value and change from core baseline in Clinical Disease Activity Index (CDAI) over time
- Actual value and change from core baseline in Simplifed Disease Acitivity Index (SDAI) over time
- ACR-N over time
- European League Against Rheumatism (EULAR) response over time.
- The proportion of subjects who achieve ACR/EULAR remission over time
- Actual value and change from core baseline in SF-36 over time
- Actual value and change from core baseline in FACIT-Fatigue over time

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 assessd 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 6-1 for 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 2.

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".

6.3.4. ACR20

A subject achieves ACR20 response when this subject has

- ≥ 20% improvement from core baseline in TJC68, AND
- ≥ 20% improvement from core baseline in SJC66, AND
- ≥ 20% improvement from core baseline in at least 3 of the following 5 items:
  1) PGA
  2) SGA
  3) Subject's pain assessment
  4) HAQ-DI score
  5) CRP

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

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

If the core baseline value is 0 then the percent improvement from core 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.

6.3.5. ACR50 and ACR70

ACR50 and ACR70 are similarly defined as ACR20 (see Section 6.3.4.), except the improvement threshold from core baseline is 50% and 70%, respectively

6.3.6. DAS28(CRP)

The DAS28(CRP) score is calculated as follows:

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

where

CRP = CRP measurement (mg/L)

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

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

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

6.3.7. SDAI and CDAI

Simplifed Disease Acitivity 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 CRP (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.

6.3.8. ACR-N

ACR-N is defined as the smallest percentage improvement from core 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.

6.3.9. 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) Improvement from Core Baseline		
DAS28(CRP) at Visit	**> 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

6.3.10. ACR/EULAR remission

ACR/EULAR remission is achineved when TJC28 ≤ 1 and SJC28 ≤ 1 and CRP ≤ 1 mg/dL and SGA ≤ 1 (on a 0-10 cm scale). ACR/EULAR remission is not achieved when at least 1 of the 4 components has a value > 1.

In case that some components are missing, ACR/EULAR 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 ACR/EULAR remission, then ACR/EULAR remission will be considered as missing.

6.3.11. Health-Related Quality of Life (HRQoL)

HRQoL questionnaires consist of Short-Form 36 Health Survey (SF-36), and FACIT-Fatigue.

SF-36

TThe 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, the physical component summary (PCS) score and the mental component summary (MCS) score will be evaluated based on the 8 SF-36 domains.

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.

6.4. Analysis of the Efficacy Endpoint

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. No formal statistical testing is planed.

Plots of proportions of subjects for categorical endpoints and mean ± SD for continuous endpoints will be presented over time by treatment and by visit.

6.5. Changes From Protocol-Specified Efficacy Analyses

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

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 the MedDRA Version 21.0. 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 Mild, Moderate and Severe 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 "Possible", "Probable" or "Certain" on the AE CRF to the question of "Causality". However, the investigator's original answer to the question will be showed in the by-subject data listings. Relatedness will always default to the investigator's choice, not that of the Medical Monitor. Events for which the investigator did not record causality will be considered related to study drug for summary purposes. However, by-subject data listings will show the relationship as missing.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if AEs met the definitions of SAE that were specified in the study protocol. SAEs captured and stored in the clinical database will be reconciled with the SAE database from the Gilead Pharmacovigilance & Epidemiology (PVE) 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 any AEs with an onset date on or after the filgotinib start date in the core studies or GLPG0634-CL-205, and no later than 30 days after permanent discontinuation of filgotinib in GLPG0634-CL-205.

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 filgotinib in the core studies or GLPG0634-CL-205, 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 filgotinib in the core studies or GLPG0634-CL-205, and
- The AE onset date is the same as or before the month and year (or year) of the date corresponding to 30 days after the date of the last dose of filgotinib in GLPG0634-CL-205

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 filgotinib in the core studies or GLPG0634-CL-205, 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 filgotinib in the core studies or GLPG0634-CL-205 will be considered treatment emergent.

7.1.6. Summaries of Adverse Events and Deaths

Treatment-emergent AEs will be summarized based on the Safety Analysis Set.

7.1.6.1. Summaries of AE incidence in Combined Severity Grade Subsets

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

- TEAEs of Grade severe
- TEAEs of Grade moderate or higher
- All TE treatment-related AEs
- TE Treatment-related AEs of Grade severe
- TE Treatment-related AEs of Grade moderate or higher
- 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 TEAEs leading to temporary interruption of any study drug
- All TE SAEs leading to death (ie, outcome of death)

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

Multiple events will be counted only once per subject in each summary. Adverse events will be summarized and listed first in alphabetic order of SOC and HLT within each SOC (if applicable), and then by PT in descending order of total frequency within each SOC. For summaries by

severity grade, the most severe grade will be used for those AEs that occurred more than once in an individual subject during the study.

In addition to the above summary tables, all TEAEs and TE treatment-related AEs will be summarized by PT only, in descending order of total frequency.

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

- All AEs, indicating whether the event is treatment emergent
- All AEs of Grade severe
- All AEs of Grade moderate 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

7.1.7. Adverse Events of Special Interest

Events of interest will be identified by the use of either SMQs or MSTs. However, should additional cases not detected by the predefined search term listings be identified during the clinical review process, these cases will also be reported by respective category.

7.1.7.1. Adjudication Committee for MACE

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

To identify the MACE event, the following potential cases will be adjudicated:

- All deaths
- Cardiovascular events (meeting serious criteria)
- Myocardial infarction (Narrow)

- Unstable angina (meeting serious or hospitalization criteria)
- Transient ischemic attack
- Stroke
- Cardiac failure (meeting serious or hospitalization criteria)
- Percutaneous coronary intervention

Potential MACE will be identified using MedDRA search listing (MST) searches.

The CVEAC will review those potential MACE and related clinical data to determine whether a MACE has occurred.

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 term.

A by-subject listing for all patients who have potential MACE at any time will be provided.

7.1.7.2. Other Adverse Events of Special Interest

In addition to general safety parameters and MACE, safety information on other adverse events of special interest (AESIs) will also be analyzed. AESIs will be identified by either laboratory results, standardized MedDRA queries (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 special interest as defined below
    a) Herpes zoster
    b) Active tuberculosis
    c) Opportunistic infections
    d) Hepatitis B or C infections
- Deep vein thrombosis (DVT) and pulmonary embolism (PE)

- Malignancy (including lymphoma; not including nonmelanoma skin cancer)
- Nonmelanoma skin cancer
- Gastrointestinal (GI) perforations

The number and percentage of patients with aforementioned events of special interest will be provided by the PT term for each AE of special interests.

A by-subject listing for all patients having AE of special interests at any time will be provided for each AE of special interests.

7.2. Laboratory Evaluations

Laboratory 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 filgotinib plus 30 days for subjects who have permanently discontinued filgotinib in GLPG0634-CL-205, or all available data at the time of the database snapshot for subjects who were ongoing at the time of an interim analysis. 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 3.4.

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.

7.2.1. Summaries of Numeric Laboratory Results

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

- Hematology
  - Hematocrit
  - Hemoglobin
  - Platelet count
  - Red blood cell count
  - 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
  - Glucose
- Lipid
  - Triglycerides
  - Total cholesterol
  - HDL
  - LDL
  - LDL/HDL ratio
- Hormone in male subjects
  - Follicle stimulating hormone
  - Inhibin B
  - Luteinizing hormone
  - Testosterone

A filgotinib baseline laboratory value will be defined as the last measurement obtained on or prior to the date/time of first dose of filgotinib in the core studies or GLPG0634-CL-205. Change from filgotinib baseline to a postbaseline visit will be defined as the visit value minus the filgotinib baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

Median (Q1, Q3) of the observed 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 3.5.3.

7.2.2. Graded Laboratory Value

The CTCAE Version 4.03 will be used to assign toxicity grades 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 filgotinib baseline at any postbaseline time point, up to and including the date of last dose of filgotinib plus 30 days for subjects who permanently discontinued filgotinib in GLPG0634-CL-205, or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

If the relevant filgotinib 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 filgotinib baseline by at least 3 toxicity grades at any postbaseline time point, up to and including the date of the last dose of filgotinib plus 30 days for subjects who permanently discontinued filgotinib in GLPG0634-CL-205 or the last available date in the database snapshot for subjects who were still on treatment at the time of an interim analysis.

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

7.2.2.3. Summaries of Laboratory Abnormalities

Laboratory data that are categorical will be summarized using the number and percentage of subjects in the study with the given response at filgotinib 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 laboratory test:

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

For all summaries of laboratory abnormalities, the denominator is the number of subjects with nonmissing postbaseline values up to 30 days after last dosing date of filgotinib in GLPG0634-CL-205.

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 laboratory test of interest, with all applicable severity grades or abnormal flags displayed.

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
- AST or ALT > 3 x ULN and total bilirubin > 2 x ULN

The summary will include data from all postbaseline visits up to 30 days after the last dose of filgotinib in GLPG0634-CL-205. For individual laboratory tests, subjects will be counted once based on the most severe postbaseline values. For both the composite endpoints 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 filgotinib baseline; (b) filgotinib baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) filgotinib baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline
- WBC count: (a) any postbaseline worsening CTCAE grade from filgotinib baseline; (b) filgotinib baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) filgotinib baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline
- Absolute neutrophil count: (a) any postbaseline worsening CTCAE grade from filgotinib baseline; (b) filgotinib baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) filgotinib baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline
- Lymphocyte count: (a) any postbaseline worsening CTCAE grade from filgotinib baseline; (b) filgotinib baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) filgotinib baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline
- Platelet count: (a) any postbaseline worsening CTCAE grade from filgotinib baseline; (b) filgotinib baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) filgotinib baseline value of less than Grade 3 and increase to Grade 4 at worst postbaseline

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

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]) as follows:

- Filgotinib baseline value
- Values at each postbaseline visit
- Change from filgotinib baseline at each postbaseline visit

A filgotinib baseline value will be defined as the last available value collected on or prior to the date/time of first dose of filgotinib in the core studies or GLPG0634-CL-205. Change from filgotinib baseline to a postbaseline visit will be defined as the postbaseline value minus the filgotinib baseline value. Body weight and vital signs measured at unscheduled visits will be included for the filgotinib baseline value selection.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.5.3. No formal statistical testing is planned.

A by-subject listing of vital signs (systolic and diastolic blood pressure [mmHg], pulse [beats/min]) 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. Concomitant Medications

Medications collected during the study will be coded using the World Health Organization (WHO) Drug dictionary version BSEP17.

Concomitant medications are defined as medications taken while a subject took study drug in GLPG0634-CL-205. Use of concomitant medications will be summarized by ATC drug class 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 filgotinib in GLPG0634-CL-205 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 filgotinib in GLPG0634-CL-205 will be considered concomitant medications. Medications started and stopped on the same day as the first dosing date or the last dosing date of filgotinib in GLPG0634-CL-205 will also be considered concomitant. Medications stopped on the same day as the first dosing date of filgotinib in GLPG0634-CL-205 will be considered concomitant. Medications with a stop date prior to the date of first dosing date of filgotinib in GLPG0634-CL-205 or a start date after the last dosing date of filgotinib in GLPG0634-CL-205 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 filgotinib administration in GLPG0634-CL-205 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 filgotinib stop date in GLPG0634-CL-205 will be excluded from the concomitant medication summary. GLPG0634-CL-205 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 GLPG0634-CL-205 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 shift table of the investigators' assessment of ECG results at each visit compared with filgotinib baseline values will be presented by treatment group using the following categories: normal; abnormal (not clinically significant); abnormal (clinically significant); or missing. The number and percentage of subjects in each cross-classification group of the shift table will be presented. Subjects with a missing value at filgotinib baseline or postbaseline will not be included in the denominator for percentage calculation.

No formal statistical testing is planned.

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

7.6. Other Safety Measures

A data listing will be provided for subjects who have positive or borderline urine or serum pregnancy test result during the study.

7.7. Changes From Protocol-Specified Safety Analyses

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

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.

9. SOFTWARE

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

10. SAP REVISION

Revision Date (DD MMM YYYY)	Section	Summary of Revision	Reason for Revision
24 JAN 2019	Section 2.2 and 2.3	Update analysis frequency	Administrative change
24 JAN 2019	Section 6.3.11	Update physical component score (PCS) to physical component summary (PCS), and mental component score (MCS) to mental component summary (MCS)	Administrative change
24 JAN 2019	Section 5.2	Update Oral systemic corticosteroids to Oral corticosteroids	Administrative change

11. APPENDIX

Appendix 1. Lists of RA Medications

1. List of of bDMARDs and Investigational bDMARDs (WHO Preferred Terms)

- ABATACEPT
- ADALIMUMAB
- ANAKINRA
- BAMINERCEPT
- BELIMUMAB
- BIMEKIZUMAB
- CABIRALIZUMAB
- CERTOLIZUMAB
- CERTOLIZUMAB PEGOL
- CLAZAKIZUMAB
- DENOSUMAB
- EPRATUZUMAB
- ETANERCEPT
- GOLIMUMAB
- INFLIXIMABINTERLEUKIN-2
- OCRELIZUMAB
- REMTOLUMAB
- RITUXIMAB
- SARILUMAB
- SECUKINUMAB
- SIRUKUMAB
- TABALUMAB
- TILDRAKIZUMAB
- TOCILIZUMAB
- TREGALIZUMAB
- Other investigational bDMARDs

2. List of Oral Corticosteroid (WHO Preferred Terms)

- BETAMETHASONE
- BETAMETHASONE DIPROPIONATE
- BETAMETHASONE SODIUM PHOSPHATE
- CORTISONE
- DEXAMETHASONE
- DEXAMETHASONE PALMITATE
- DEXAMETHASONE SODIUM PHOSPHATE
- MEPREDNISONE
- METHYLPREDNISOLONE
- METHYLPREDNISOLONE ACETATE
- METHYLPREDNISOLONE SODIUM SUCCINATE
- MOMETASONE FUROATE
- PREDNISOLONE
- PREDNISOLONE FARNESYLATE
- PREDNISOLONE SODIUM SUCCINATE
- PREDNISONE

3. List of Methotrexate (WHO Preferred Terms)

- METHOTREXATE
- METHOTREXATE SODIUM

Appendix 2. 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), usually administered by the patient. Responses in each functional category are collected as 0 (without any difficulty) to 3 (unable to do a task in that area), with or without aids or devices.

The highest score for questions in each category (range 0 to 3) determines the score for the category, unless aids or devices are required. Dependence on equipment or physical assistance increases a lower score (ie, scores of 0 or 1) to the level of 2 to more accurately represent underlying disability. The eight category scores are averaged into an overall HAQ-DI score on a scale from 0 (no disability) to 3 (completely disabled) when 6 or more categories are non-missing. If more than 2 categories are missing, the HAQ-DI score is set to missing. The HAQ-DI can be treated as a continuous measure.

The HAQ-DI score using aids (and/or devices) is computed by taking the maximum score of the questions in each category (range: [0, 3]) and whether or not aids/devices are used (0 or 1):

A = max(dressing & grooming category questions, 2*aids indicator) +
max(rising category questions, 2*aids indicator) +
max(eating category questions, 2*aids indicator) +
max(walking category questions, 2*aids indicator) +
max(hygiene category questions, 2*aids indicator) +
max(reach category questions, 2*aids indicator) +
max(grip category questions, 2*aids indicator) +
max(usual activities category questions, 2*aids indicator)

HAQ-DI = A/(total number of categories with at least 6 non-missing)

The following table shows the contribution of the 43 questions used to calculate the HAQ-DI:

HAQ-DI Category:	Category questions: At least 6 categories must have scores to compute the HAQ-DI.		HAQ-DI Category Score with Aids/Devices Calculation:
	Category Questions	Aids/Devices Indicators	
Dressing/Grooming	HAQ0101, HAQ0102 (DRESS, HAIR)	HAQ0114, HAQ0119 (DRSG, GROOM)	Using each question with a scale of 0-3, calculate the category score as the maximum of the category questions. If the Aids/Devices indicator is "No", no need to adjust the category score. If the Aids/Devices indicator is "Yes" and the category score is <2, then the category score with the Aids/Devices is set to 2. If the Aids/Devices indicator is "Yes" and the category score is ≥2, then the category score with Aids/Devices is the calculated category score without adjustment. For example: The Dressing/Grooming category score is 2 if subject answered 1 for both questions 1 and 2 and "Yes" for both question 14 and 18. In the HAQ-DI score calculation, questions on other device/aids will not be used.
Arising	HAQ0103, HAQ0104 (STAND, BED)	HAQ0116, HAQ0120 (CHAIR, ARISING)	
Eating	HAQ0105,HAQ0106, HAQ0107 (MEAT, LIFT,MILK)	HAQ0115, HAQ0121 (UTENSIL, EAT)	
Walking	HAQ0108, HAQ0109 (WALK, STEPS)	HAQ0110, HAQ0111, HAQ0112, HAQ0113, HAQ0122 (CANE, WALKER, CRUTCH, WHEEL, WALKING)	
Hygiene	HAQ0123, HAQ0124, HAQ0125 (WASH, BATH, TOILET)	HAQ0134, HAQ0135, HAQ0137, HAQ0139, HAQ0142 (RAISEAT, BATHBAR, BATHSEAT, LONGBATH, HYGIENE)	
Reach	HAQ0126, HAQ0127 (REACH, BEND)	HAQ0138, HAQ0143 (LONGRCH, REACH)	
Grip	HAQ0128, HAQ0129, HAQ0130 (OPENCAR, JAR, FAUCET)	HAQ0136, HAQ0144 (JAROPEN, GRIP)	
Activity	HAQ0131, HAQ0132, HAQ0133 (SHOP, INCAR, CHORES)	HAQ0145 (ERRAND)	

Handling Missing Data: If no more than 2 categories have missing category scores, then the HAQ-DI is the mean of the non-missing category scores. Otherwise, the HAQ-DI score is set to missing.

If any of the category questions are missing, but the aids/device indicator is non-missing, the category score can still be computed. However, if all category questions and its aids/device indicators are missing, then the category score is considered missing.

Appendix 3. Corticosteroids

The following table will be used for converting non-prednisone medications to prednisone equivalent:

*Example: Subject is taking 8 mg of Methylprednisolone orally daily. To get the equivalent dose of prednisone: 8 mg Methylprednisolone = (5*8)/ 4 = 10 mg prednisone.*

Corticosteroids Name	Equivalent Dose (mg) to 5 mg Prednisone
Betamethasone	0.75
Betamethasone Dipropionate	0.75
Betamethasone Sodium Phosphate	0.75
Cortisone	20
Dexamethasone	0.75
Dexamethasone Palmitate	0.75
Dexamethasone Phosphate	0.75
Dexamethasone Sodium Phosphate	0.75
Hydrocortisone	20
Meprednisone	4
Methylprednisolone	4
Methylprednisolone Acetate	4
Methylprednisolone Sodium Succinate	4
Prednisone	5
Prednisolone	5
Prednisolone Farnesylate	5
Prednisolone Sodium Succinate	5
Triamcinolone	4