STATISTICAL ANALYSIS PLAN

Study Title:	A Phase 3, Randomized, Double-blind, Placebo and Adalimumab-controlled Study to Evaluate the Efficacy and Safety of Filgotinib in Subjects with Active Psoriatic Arthritis Who Are Naïve to Biologic DMARD Therapy
Name of Test Drug:	Filgotinib
Study Number:	GS-US-431-4566
Protocol Version (Date):	Amendment 2: 17 April 2020
Analysis Type:	Final Analysis
Analysis Plan Version:	1.0
Analysis Plan Date:	16 July 2021
Analysis Plan Author(s):	PPD

CONFIDENTIAL AND PROPRIETARY INFORMATION

TABLE OF CONTENTS

LIST OF IN-TEXT TABLES

LIST OF IN-TEXT FIGURES

LIST OF ABBREVIATIONS

ACR	American College of Rheumatology
ACR20/50/70	American College of Rheumatology 20/50/70% improvement
ADA	adalimumab
AE	adverse event
AESI	adverse events of special interest
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical drug class
bioDMARD	biologic disease modifying anti-rheumatic drug
BLQ	below the limit of quantitation
BMI	body mass index
CCI	[redacted]
BSA	body surface area
CASPAR	Classification Criteria for Psoriatic Arthritis
CBC	complete blood count
CI	confidence interval
eCRF	electronic case report form
COVID-19	coronavirus (2019 nCOV)
CRP	C-reactive protein
csDMARDs	conventional synthetic disease-modifying anti-rheumatic drugs
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DAPSA	Disease Activity in Psoriatic Arthritis
DAS28	disease activity score for 28 joint count
DMC	data monitoring committee
ECG	electrocardiogram
EDC	electronic data capture
eCRF	electronic case report form
EQ-5D-5L	EuroQoL 5 Dimensions with 5 Levels
ET	early termination
FACIT-Fatigue	Functional Assessment of Chronic Illness Therapy-Fatigue
FAS	full analysis set
GI	gastrointestinal
GSI	Gilead Sciences
HAQ-DI	Health Assessment Questionnaire-Disability Index
HLGT	high-level group term
HLT	high-level term

HRQoL	health-related quality of life
HCRU	healthcare resource utilization questionnaire
hsCRP	high-sensitivity C-Reactive Protein
ID	identification
IPD	important protocol deviation
IXRS	interactive voice/web response system
LDA	low disease activity
LDI	Leeds Dactylitis Index
LLOQ	lower limit of quantitation
LLT	lower-level term
LOQ	limit of quantitation
LS	least squares
LTE	long-term extension
MACEs	major adverse cardiovascular events
MCS	mental component score
MDA	Minimal Disease Activity
MedDRA	Medical Dictionary for Regulatory Activities
MI	multiple imputation
MMRM	mixed model repeated measures
MNAR	missing not at random
MRI	Magnetic Resonance Imaging
MST	MedDRA search term
MTX	methotrexate
NRI	non-responder imputation
NRS	numerical rating scale
NSAIDs	non-steroidal anti-inflammatory drugs
OC	observed case
PASDAS	Psoriatic Arthritis Disease Activity Score
PASI including BSA	Psoriasis Area and Severity Index including Body Surface Area
PASI75	Psoriasis Area and Severity Index 75% improvement
PCS	physical component summary
PD	pharmacodynamics
PGADA	Patient's Global Assessment of Disease Activity
PGAPI	Patient's Global Assessment of PsA Pain Intensity
PhGADA	Physician's Global Assessment of Disease Activity
PhGAP	Physician's Global Assessment of Psoriasis
PK	pharmacokinetic
PP	per-protocol
PROs	patient reported outcomes
PsA	psoriatic arthritis

PsAID-12	12-item Psoriatic Arthritis Impact of Disease
PsAMRIS	Psoriatic Arthritis Magnetic Resonance Imaging Score
PsARC	Psoriatic Arthritis Response Criteria
PT	preferred term
PTM	placebo to match
Q1, Q3	first quartile, third quartile
q2w	once every two weeks
q.d.	quaque die (each day)
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SE	standard error
SF-36v2	36-item short form survey version 2
SJC/TJC	Swollen and Tender Joint Count
SMQ	standardized MedDRA query
SOC	system organ class
SPARCC Enthesitis Index and LEI	Spondyloarthritis Research Consortium of Canada Enthesitis Index and Leeds Enthesitis Index
TDC	tender dactylitis count
TEAE	treatment-emergent adverse event
TFLs	tables, figures, and listings
ULN	upper limit of normal
VAS	visual analog scale
VLDA	Very Low Disease Activity
WBC	white blood cell
WHO	World Health Organization
WPAI-PsA	Work Productivity and Activity Impairment for Psoriatic Arthritis

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 GS-US-431-4566. This SAP is based on the study protocol amendment 2 dated 17 April 2020 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 effect of filgotinib compared to placebo in active PsA as assessed by the American College of Rheumatology 20% improvement (ACR20) response at Week 12

The secondary objectives of this study are as follows:

- To evaluate the effect of filgotinib on core domains of psoriatic arthritis (PsA, e.g. peripheral arthritis, psoriatic skin disease, enthesitis and dactylitis) as assessed by Minimal Disease Activity (MDA), Very Low Disease Activity (VLDA), ACR responses, Psoriasis Area and Severity Index including Body Surface Area (PASI including BSA) responses, Spondyloarthritis Research Consortium of Canada Enthesitis Index and Leeds Enthesitis Index (SPARCC Enthesitis Index and LEI), Leeds Dactylitis Index (LDI), Psoriatic Arthritis Disease Activity Score (PASDAS), Disease Activity Index for Psoriatic Arthritis (DAPSA), Modified Nail Psoriasis Area and Severity Index (mNAPSI), and Physician's Global Assessment of Psoriasis (PhGAP)

- To evaluate the effect of filgotinib on physical function in active PsA as assessed by Health Assessment Questionnaire-Disability Index (HAQ-DI)

- To evaluate the effect of filgotinib on fatigue and quality of life in active PsA as assessed by Functional Assessment of Chronic Illness Therapy - Fatigue Scale (FACIT-Fatigue), 36-item Short-Form Health Survey Version 2 (SF-36v2), and 12-item Psoriatic Arthritis Impact of Disease (PsAID-12)

- To evaluate the efficacy of filgotinib versus adalimumab in active PsA as assessed by ACR20 response

- To evaluate the safety and tolerability of filgotinib

CCI

[redacted]

1.2. Study Design

This is a randomized, double-blind, active-controlled, Phase 3 study in adult male and female subjects with active PsA who have had an inadequate response or intolerance to 1 or more therapies for PsA, such as conventional synthetic disease-modifying anti-rheumatic drugs (csDMARDs), apremilast and / or non-steroidal anti-inflammatory drugs (NSAIDs), but have never received a biologic DMARD (bioDMARD) for PsA or psoriasis.

The study consists of two parts, the Main Study: Screening through Week 16 (inclusive), and the Long Term Extension (LTE): after Week 16 up to approximately 2.25 years.

Part 1 – Main Study (Screening through Week 16):

Approximately 854 subjects will be randomized in a 2:2:1:2 ratio to one of 4 dosing groups as outlined below.

Randomization will be stratified by geographic region, concurrent use of conventional synthetic DMARD(s) (csDMARDs) and / or apremilast at randomization (yes or no)

Dosing groups in the Main Study:

- Filgotinib 200 mg group: **filgotinib 200 mg once daily** + placebo to match (PTM) filgotinib 100 mg once daily + PTM adalimumab subcutaneous (SC) injection once every two weeks (q2w)
- Filgotinib 100 mg group: PTM filgotinib 200 mg once daily + **filgotinib 100 mg once daily** + PTM adalimumab SC injection q2w

- Active comparator group: PTM filgotinib 200 mg once daily + PTM filgotinib 100 mg once daily + **adalimumab 40 mg SC injection q2w**

- Placebo control group: PTM filgotinib 200 mg once daily + PTM filgotinib 100 mg once daily + PTM adalimumab SC injection q2w

NOTE: All subjects will discontinue adalimumab / PTM injections by the Week 16 Visit (the last injections should be at approximately Week 14).

Part 2 – LTE (After the Week 16 Visit for 2 years):

After completion of the Main Study, subjects who have not permanently discontinued study drug will continue on to the LTE as follows:

- Those who were assigned to the filgotinib groups will continue on the same study drug assignments

- Those who were assigned to the placebo or active comparator group will be reassigned 1:1 in a blinded fashion to filgotinib 200 mg or 100 mg once daily

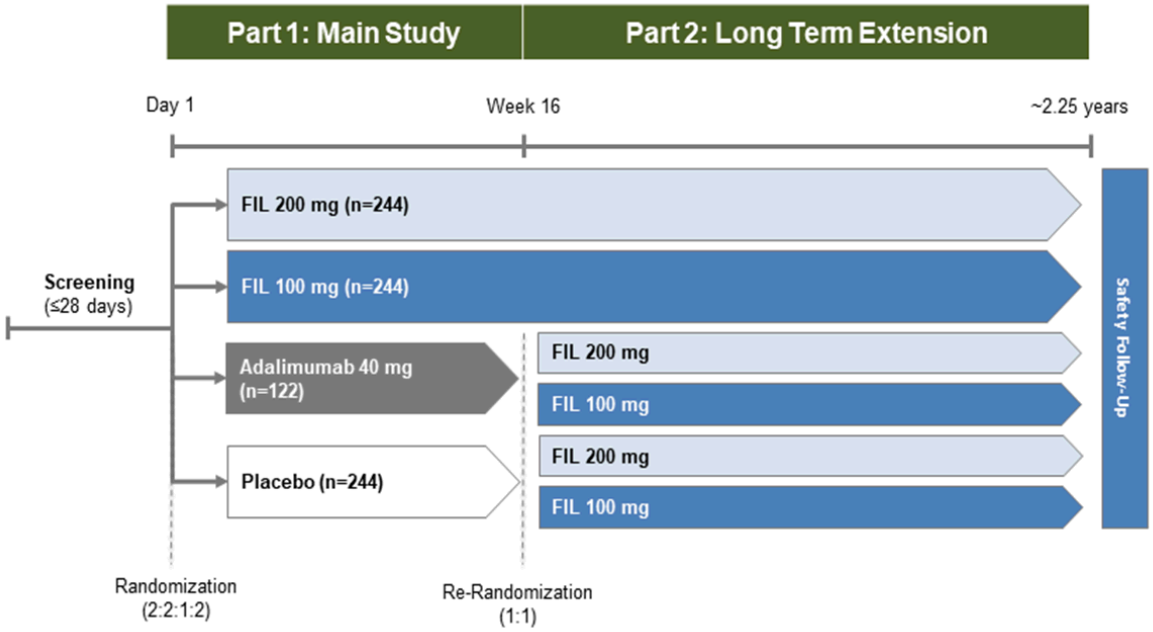
Dosing groups in the LTE:

- Filgotinib 200 mg group: **filgotinib 200 mg once daily** + PTM filgotinib 100 mg once daily

- Filgotinib 100 mg group: PTM filgotinib 200 mg once daily + **filgotinib 100 mg once daily**

For the first 16 weeks of study participation, subjects who temporarily interrupt or permanently discontinue study drug for any reason are to continue with study visits and assessments through the Week 16 Visit, per protocol Section 3.5 unless the subject withdraws consent, is lost to follow-up, and / or continued participation in the study is medically contraindicated, per investigator's judgment. Study drug interruption and discontinuation considerations are outlined in protocol Section 3.5. All subjects who permanently discontinue study drug should continue to receive standard of care treatment for their PsA including additional therapies, if required.

Figure 1-1. Study Design



1.3. Sample Size and Power

Sample size is determined based on the non-inferiority test of each filgotinib group compared to the adalimumab group on the ACR20 response rate at Week 12. When assuming the ACR20 response rate being 52% and 60% for the adalimumab and each filgotinib group, and 38.6% for the placebo group, 244 subjects in each filgotinib group and placebo group, and 122 subjects in the adalimumab group are required to obtain 90% power at a two-sided 0.025 significance level to demonstrate that filgotinib group preserves more than 50% of the effect of adalimumab with respect to the ACR20 response rate at Week 12.

A sample size of 244 subjects in each filgotinib group and the placebo group will provide over 95% power to detect a difference in ACR20 response rate of 21.4% at Week 12 (38.6% and 60% for the placebo group and each filgotinib group, respectively) using a two-sided 0.025 significance level superiority test.

In summary, the total sample size will be approximately 854 subjects per original planning.

The study was terminated per Sponsor's decision after 67 subjects had been enrolled. At study termination, 13 out of 67 subjects continued on to the LTE, and 7 of those 13 subjects were terminated immediately after entering LTE phase. All the other subjects were prematurely discontinued from or completed the Main Study visits.

2. TYPE OF PLANNED ANALYSIS

2.1. Final Analysis

After study early termination, the final analysis will be performed for all enrolled subjects. Final analysis will occur after outstanding data queries have been resolved or adjudicated as unresolvable, and the data have been cleaned and finalized.

3. GENERAL CONSIDERATIONS FOR DATA ANALYSES

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

By-subject listings will be presented for all subjects in the All Randomized Analysis Set 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 will be used in the listings. Age, sex at birth, race, and ethnicity will be included in the listings, as space permits.

Given the small number of subjects who entered LTE phase, for all the efficacy and safety analyses, unless specified otherwise, all the analyses and summary tables will be generated based on the data during Main Study phase (for subjects enrolled under Protocol Amendment (PA) 1 who stayed in Main Study beyond Week 16 visit, only the data up to Week 16 visit will be included). All the data will be presented in the listing.

3.1. Analysis Sets

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

For each analysis set, the number and percentage of subjects eligible for inclusion, will be summarized by treatment group.

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

3.1.1. All Randomized Analysis Set

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

3.1.2. Full Analysis Set

The Full Analysis Set (FAS) includes all randomized subjects who took at least 1 dose of study drug. The study drugs in this study are filgotinib 200 mg, filgotinib 100 mg, adalimumab and PTMs. This is the primary analysis set for efficacy analyses.

3.1.3. Safety Analysis Set

The Safety Analysis Set includes all subjects who took at least 1 dose of study drug. This is the primary analysis set for safety analyses.

3.2. Subject Grouping

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

The treatment groups during the Main Study phase are:

- Filgotinib 200 mg once daily
- Filgotinib 100 mg once daily
- Adalimumab q2w
- Placebo

The treatment groups during the LTE phase are:

- Filgotinib 200 mg once daily
- Filgotinib 100 mg once daily
- Adalimumab switch to filgotinib 200 mg once daily
- Adalimumab switch to filgotinib 100 mg once daily
- Placebo switch to filgotinib 200 mg once daily
- Placebo switch to filgotinib 100 mg once daily

3.3. Strata and Covariates

Subjects will be randomly assigned to treatment groups via the interactive voice or web response system (IXRS) in a 2:2:1:2 ratio to filgotinib 200 mg, filgotinib 100 mg, adalimumab or placebo control using a stratified randomization schedule. Stratification will be based on the following variables:

- Geographic region (Group A includes the following countries: Australia, Canada, New Zealand, Spain, and United States; Group B includes the following countries: Bulgaria, Poland, Group E includes Japan)
- Concurrent use of conventional synthetic DMARD(s) (csDMARDs) and / or apremilast at randomization (Yes or No)

List of csDMARD and/or apremilast is provided in Appendix 6.

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

For efficacy endpoints, stratification factors and the baseline value of the efficacy variable(s) will be included as a covariate in the efficacy analysis model, as specified in Section 6. Given the small number of subjects in geographic region Group E, Group B and E will be combined as a single group in all the efficacy modeling.

3.4. Multiple Comparisons

Due to study early termination and insufficient number of subjects enrolled, the hypothesis testings performed and correspondingly reported p-values for all the comparisons planned in this SAP will be nominal, and need to be interpreted with caution.

3.5. Missing Data and Outliers

3.5.1. Missing Data

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

For missing last dosing date of study drug, imputation rules are described in Section 4.2.1. For partial date of initial PsA 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. Imputation rules adopted in the efficacy analyses are specified in Section 6.

3.5.2. 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 analysis.

3.6. Data Handling Conventions and Transformations

The following conventions will be used for the imputation of date of birth:

- Since only the year of birth is collected in the study, “01 July” will be imputed as the day and month of birth
- If year of birth is missing, then date of birth will not be imputed.

In general, age collected at Day 1 (in years) will be used for analyses and presented in listings. If age at Day 1 is not available for a subject, then age derived based on date of birth and the Day 1 visit date will be used instead. If an enrolled subject was not dosed with any study drug, the randomization date will be used instead of the Day 1 visit date. For screen failures, the date the first informed consent was signed will be used for the age derivation.

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

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

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

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

3.7. Analysis Visit Windows

3.7.1. Definition of Study Day

The study consists of two study phases: Main Study phase and LTE phase. Main study is from screening through Week16 Visit. The subjects who complete the Main Study and have not prematurely discontinued study drug will continue on to the LTE. For the subjects who continue on to LTE phase, Main Study phase is up to Week 16 visit date (inclusive); LTE phase is from Week 16 visit date + 1 forward (for subjects enrolled under PA 1 who stayed in Main Study beyond Week 16 Visit, refer to Section 3.7.2 for exception).

The first dose date of individual study drug will be calculated separately for each study drug (ie, filgotinib 200mg, filgotinib 100mg, adalimumab and PTMs) 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 for an individual study drug will be the maximum 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

For subjects who continue on to the LTE, the LTE Study Day 1 is defined as the first dose date of any study drug for LTE. The LTE Study Day will be calculated from the LTE Study Day 1 and derived as follows:

- For days post LTE Study Day 1: Assessment Date – LTE Study Day 1 + 1
- For days prior to the LTE Study Day 1: Assessment Date – LTE Study Day 1

3.7.2. Analysis Visit Windows

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

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

For subjects who were re-randomized to LTE phase (placebo or adalimumab treated subjects), the re-randomized baseline will be the last nonmissing value obtained on or prior to the first dose date of any study drug for LTE.

For subjects enrolled under PA 1 who stayed in Main Study beyond Week 16 Visit, the analysis visit window will only be applied to the data up to Week 16 visit.

The analysis windows for joint count assessment, HAQ-DI including subject's pain assessment, Patient's Global Assessment of Disease Activity (PGADA), Patient's Global Assessment of Psoriatic Arthritis Pain Intensity (PGAPI), Physician's Global Assessment of Disease Activity (PhGADA), CRP, vital signs, weight, and safety laboratory data (lipid data excluded) are provided in Table 3-1.

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

Analysis Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	Week 16 Visit Date – Study Day 1 + 1*
Analysis Visit	**Nominal LTE Study Day**	**LTE Lower Limit**	**LTE Upper Limit**
Week 18	15	2	22
Week 20	29	23	43
Week 24	57	44	71
Week 28	85	72	113
Week 36	141	114	183
Week 48	225	184	267
Week 60	309	268	351
Week 72	393	352	435
Week 84	477	436	519
Week 96	561	520	603
Week 108	645	604	687
Week 120	729	688	≥ 729

* For subjects who did not enter LTE phase, upper limit is ">= 113"; for subjects enrolled under PA1 who stayed in Main Study beyond Week 16, upper limit is 127.

The analysis windows for lipid data are provided in Table 3-2:

Table 3-2. Analysis Visit Windows for Lipid Data

Analysis Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 16	113	2	Week 16 Visit Date – Study Day 1 + 1*
Analysis Visit	**Nominal LTE Study Day**	**LTE Lower Limit**	**LTE Upper Limit**
Week 48	225	2	393
Week 96	561	394	645
Week 120	729	646	≥ 729

* For subjects who did not enter LTE phase, upper limit is ">= 113"; for subjects enrolled under PA1 who stayed in Main Study beyond Week 16, upper limit is 141.

The analysis windows for PhGAP data are provided in Table 3-3.

Table 3-3. Analysis Visit Windows for PhGAP Data

Analysis Visit	Nominal Study Day	Visit Window Study Day	
		Lower Limit	Upper Limit
Baseline	1	(none)	1
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	Week 16 Visit Date– Study Day 1 + 1*
Analysis Visit	**Nominal LTE Study Day**	**LTE Lower Limit**	**LTE Upper Limit**
Week 18	15	2	22
Week 20	29	23	43
Week 24	57	44	71
Week 28	85	72	113
Week 36	141	114	183
Week 48	225	184	309
Week 72	393	310	477
Week 96	561	478	645
Week 120	729	646	≥ 729

* For subjects who did not enter LTE phase, upper limit is ">= 113"; for subjects enrolled under PA1 who stayed in Main Study beyond Week 16, upper limit is 127.

The analysis windows for PASI including BSA, mNAPSI, SPARCC Enthesitis Index and LEI, LDI data are provided in Table 3-4:

Table 3-4. Analysis Visit Windows for PASI including BSA, mNAPSI, SPARCC Enthesitis Index and LEI, LDI Data

		Visit Window Study Day	
Analysis Visit	**Nominal Study Day**	**Lower Limit**	**Upper Limit**
Baseline	1	(none)	1
Week 4	29	2	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	Week 16 Visit Date – Study Day 1 + 1*
Analysis Visit	**Nominal LTE Study Day**	**LTE Lower Limit**	**LTE Upper Limit**
Week 20	29	2	43
Week 24	57	44	71
Week 28	85	72	113
Week 36	141	114	183
Week 48	225	184	309
Week 72	393	310	477
Week 96	561	478	645
Week 120	729	646	≥ 729

* For subjects who did not enter LTE phase, upper limit is ">= 113"; for subjects enrolled under PA1 who stayed in Main Study beyond Week 16, upper limit is 127.

The analysis windows for FACIT-Fatigue, SF-36v2, PsAID-12 data are provided in Table 3-5

Table 3-5. Analysis Visit Windows for FACIT-Fatigue, SF-36v2, PsAID-12 Data

		Visit Window Study Day	
Analysis Visit	**Nominal Study Day**	**Lower Limit**	**Upper Limit**
Baseline	1	(none)	1
Week 4	29	2	71
Week 16	113	72	Week 16 Visit Date – Study Day 1 + 1*
Analysis Visit	**Nominal LTE Study Day**	**LTE Lower Limit**	**LTE Upper Limit**
Week 48	225	2	393
Week 96	561	394	645
Week 120	729	646	≥ 729

* For subjects who did not enter LTE phase, upper limit is ">= 113"; for subjects enrolled under PA1 who stayed in Main Study beyond Week 16, upper limit is 141.

Treatment Policy estimand is defined for efficacy analyses, and for efficacy measurements, the analysis visit windows will be applied to all available data observed on study.

While on Treatment estimand is defined for safety analysis, and for vital signs, weight, and laboratory data, the analysis visit windows will be applied to data collected during the on-treatment period. The on-treatment period is defined up to the last dose date of any study drug + 7 days.

The safety data collected in the post-treatment safety 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). Data obtained after last dose date plus 30 days will be excluded from the summaries, but will still be included in the listing.

3.7.3. 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 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 nonmissing value on or prior to the first dosing date of study drug will be selected, unless specified differently. If there are multiple records with the same time or no time recorded on the same day, the baseline value will be the average of the measurements for continuous data, or the measurement with the lowest severity (e.g., normal will be selected over abnormal) for categorical data.

- For postbaseline values:

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

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

  — If there is more than 1 record on the selected day, the average will be taken for continuous data and the worse severity will be taken for categorical data, unless otherwise specified.

4. SUBJECT DISPOSITION

4.1. Subject Enrollment and Disposition

A summary of subject enrollment will be provided by treatment group and overall 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.

A similar enrollment table will be provided by randomization stratum. The denominator for the percentage of subjects in the stratum will be the total number of enrolled subjects. If there are discrepancies in the value used for stratification assignment between the IXRS and the clinical database, the value collected in the clinical database will be used for the summary. A listing of subjects with discrepancies in the value used for stratification assignment between the IXRS and the clinical database at the time of data finalization will be provided.

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

A summary of subject disposition will be provided by treatment group and study phase (Main Study and LTE). This summary will present the number of subjects screened, the number of subjects who met all eligibility criteria but were not randomized with reasons subjects not randomized, the number of subjects randomized, and the number of subjects in each of the categories listed below:

- Safety Analysis Set
- Full Analysis Set
- Completed study drug (Main Study phase)
- Did not complete study drug with reasons for premature discontinuation of study drug (Main Study phase)
- Entered LTE phase
- Did not complete study drug with reasons for premature discontinuation of study drug (LTE phase)
- Did not complete the study with reasons for premature discontinuation of study in Main Study and LTE phase respectively

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

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
- Reasons for screen failure (will be provided by screening ID number in ascending order)
- Lot number and kit ID of assigned study drugs

4.2. Extent of Study Drug Exposure and Adherence

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

4.2.1. Duration of Exposure to Study Drug

Total duration of exposure to any study drug for Main Study will be defined as last dosing date of any study drug for Main Study minus first dosing date of any study drug plus 1, regardless of any temporary interruptions in study drug administration, and will be expressed in weeks using up to 1 decimal place (e.g., 4.5 weeks).

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

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

The total duration of exposure to any study drug for Main Study will be summarized using descriptive statistics and using the number (i.e., cumulative counts) and percentage of subjects exposed through the following time periods: Baseline (Day 1), Week 2 (Day 15), Week 4 (Day 29), Week 8 (Day 57), Week 12 (Day 85), Week 16 (Day 113). Summaries will be provided by treatment group for the Safety Analysis Set.

No formal statistical testing is planned.

4.2.2. Adherence to Study Drug

Adherence will be calculated separately for individual study drugs, including filgotinib 200 mg)/PTM (tablets), filgotinib 100 mg/PTM (tablets), and adalimumab/PTM (syringes).

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

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

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

$$\text{Total Number of Syringes Administered} = \left(\sum \text{No. of Syringes Dispensed}\right) - \left(\sum \text{No. of Syringes Returned}\right)$$

If a bottle or a packaging kit is dispensed and the bottle or packaging kit is returned empty, then the number of tablets/syringes returned will be entered as zero. If a bottle or a packaging kit is dispensed but not returned (missing), then the number of tablets/syringes administered will be counted as zero.

On-Treatment Adherence will be calculated to assess the level of adherence to study drugs.

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

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

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

Study drug expected to be administered for filgotinib 200 mg/PTM (tablets) = 1× total duration of exposure to any study drug for Main Study (days).

Study drug expected to be administered for filgotinib 100 mg/PTM (tablets) = 1× total duration of exposure to any study drug for Main Study (days).

Study drug expected to be administered for adalimumab/PTM (syringes) = 1×ceil[(min(Last dose date of any study drug for Main Study, First dose date of any study drug + 14×7) – First dose date of any study drug + 1) /14].

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

No formal statistical testing is planned.

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

4.3. Protocol Deviations

Subjects who did not meet the eligibility criteria for study entry but randomized 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 All Randomized Analysis Set. A by-subject listing will be provided for those subjects who did not meet at least 1 eligibility (inclusion or exclusion) criterion. The listing will present the eligibility criterion (or criteria if more than 1 deviation) that subjects did not meet and related comments, if collected.

Protocol deviations occurring after subjects entered the study are documented during routine monitoring. The number and percentage of subjects with important protocol deviations (IPDs) by deviation reason (as specified in the IPD plan) will be summarized by treatment group for the All Randomized Analysis Set. A by-subject listing will be provided for those subjects with IPD.

4.4. Assessment of COVID-19 Impact

This study was ongoing during the novel coronavirus (2019 nCOV [COVID-19]) pandemic which has caused a disruption in the regular visit schedules for this study. Some subjects were unable to attend onsite visits due to shelter in place guidelines, site closures, or other reasons. This section provides how special situations due to COVID-19 will be handled in the analysis.

4.4.1. Study Drug or Study Discontinuation Due to COVID-19

A by-subject listing of reasons for premature study drug or study discontinuation due to COVID-19 will be created.

4.4.2. Protocol Deviations Due to COVID-19

A by-subject listing of important protocol deviation related to COVID-19 and non-important protocol deviation related to COVID-19 will be provided.

4.4.3. Missed and Virtual Visits due to COVID-19

A by-subject listing of subjects with missed or virtual visits due to COVID-19 will be provided by subject ID number in ascending order.

Information regarding missed or virtual visits due to COVID-19 was collected as free text in the CRF comment fields. The determination of missing or virtual visits due to COVID-19 will be done using Natural Language Processing (NLP) to search the CRF comment fields. A detailed explanation of the algorithm is given in Appendix 5.

4.4.4. Adverse Events due to COVID-19

AEs due to COVID-19 will be included in analyses of AE if applicable, which will be determined through COVID-19 standardized MedDRA queries (SMQ) broad search. A by-subject listing of AEs due to COVID-19 will be provided.

5. BASELINE CHARACTERISTICS

5.1. Demographics and Other Baseline Characteristics

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

- Age (on the first dose date of any study drug)
- Age group (< 65 years, ≥ 65 years)
- Sex at birth (male, female)
- Race
- Ethnicity (Hispanic or Latino, not Hispanic or Latino)
- Geographic region and country
- Weight (kg)
- Height (cm)
- Body mass index (BMI; in kg/m^2)
- Smoking Status (non-smoker, former smoker or current smoker)

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

5.2. Baseline Disease Characteristics

Baseline disease characteristics include:

- Concurrent use of csDMARD and/or apremilast at randomization (Yes or No) and daily or weekly dose taken at randomization by medication
  - — Only csDMARD and/or apremilast medication records with a start date prior to randomization date and an end date on or after randomization date or missing will be used for counting the number of subjects with csDMARD and/or apremilast use at randomization.
  - — Only csDMARD and/or apremilast medication records with a start date prior to randomization date and an end date on or after randomization date or missing that have a valid dose, dose unit, and dosing frequency will be used for summarizing daily or weekly dose taken at randomization.

- Oral Corticosteroids use on Study Day 1 (Yes or No) and daily dose taken on Study Day 1
  - Only oral Corticosteroid medication records with a start date prior to Study Day 1 and an end date on or after Study Day 1 or missing will be used for counting the number of subjects with oral Corticosteroid use on Study Day 1.
  - Only oral Corticosteroid medication records with a start date prior to Study Day 1 and an end date on or after Study Day 1 or missing that have a valid dose, dose unit, and dosing frequency will be used for summarizing daily dose taken on Study Day 1.
  - All oral Corticosteroid medications' doses will be converted to prednisone equivalent dose (conversion chart is in Appendix 3).
  - If multiple oral Corticosteroid medications are taken on Study Day 1, the total daily dose will be the sum of those medications' doses at Study Day 1 after conversion.
- Topical Corticosteroids use on Study Day 1 (Yes or No)
  - Only topical Corticosteroid medication records with a start date prior to Study Day 1 and an end date on or after Study Day 1 or missing will be used for counting the number of subjects with topical Corticosteroid use on Study Day 1.
- NSAIDs use on Study Day 1 (Yes or No)
  - Only NSAIDs medication records with a start date prior to Study Day 1 and an end date on or after Study Day 1 or missing will be used for counting the number of subjects with NSAIDs use on Study Day 1.
- Duration of PsA from diagnosis (years)
  - Calculated as ((first dose date) – (date of initial diagnosis) + 1 day) / 365.25. If the date of initial diagnosis is incomplete, then the following rules will be applied:
    - missing day: use the first of the month
    - missing month: use January
- 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)
- HAQ-DI total score

- SF-36v2 physical component summary (PCS) score
- SF-36v2 mental component summary (MCS) score
- FACIT-Fatigue
- LDI: Leeds Dactylitis Index in subjects with dactylitis at Baseline
- SPARCC Enthesitis Index: Spondyloarthritis Research Consortium of Canada Enthesitis Index in subjects with enthesitis at Baseline
- LEI: Leeds Enthesitis Index in subjects with enthesitis at Baseline
- Patient's Global Assessment of Psoriatic Arthritis Pain Intensity (PGAPI) (by VAS)
- PASI: Psoriasis Area and Severity Index in subjects with psoriasis covering >=3% of the BSA at Baseline
- PhGAP: Physician's Global Assessment of Psoriasis in subjects with psoriasis covering >=3% of the BSA at Baseline
- mNAPSI: Modified Nail Psoriasis Severity Index in subjects with psoriatic nail involvement at Baseline
- PsAID-12: 12-item Psoriatic Arthritis Impact of Disease
- DAS28 (CRP): Disease Activity Score 28 by CRP
- DAPSA: Disease Activity in Psoriatic Arthritis
- PASDAS: Psoriatic Arthritis Disease Activity Score
- Tender dactylitis count (TDC) in subjects with dactylitis at Baseline
- Patients' global assessment of disease activity (PGADA)
- Physician's global assessment of disease activity (PhGADA)
- High-Sensitivity C-reactive Protein (hsCRP)
- hsCRP (>= 6mg/L or <6mg/L)

These 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 these baseline characteristics will be provided for the Safety Analysis Set. No formal statistical testing is planned.

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

5.3. Medical History

Medical history collected at screening will be coded using the current version of 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 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 in ascending order.

6. EFFICACY ANALYSES

6.1. General Considerations

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

Due to study early termination and insufficient number of subjects enrolled, all the hypothesis testings performed and p-values reported will be nominal, and, unless specified otherwise, all the confidence intervals will be calculated based on the nominal level of 95%. The results need to be interpreted with caution.

Efficacy analysis will be conducted on all available data during the Main Study phase (including data collected through the Week 16 visit after study drug discontinuation).

Estimands:

Treatment policy estimand is defined for the primary and key secondary as well as other secondary efficacy endpoints and will be the primary estimand for all the corresponding efficacy analyses:

1) Population: Subjects in the FAS.

2) Variable / endpoint: Primary and key secondary endpoints as well as other secondary endpoints (see details in subsequent Sections).

3) Strategy to address intercurrent events: include all on-study data regardless of protocol violations, use of rescue medication, change in background medication, and study drug discontinuation.

4) Population-level summary: treatment difference in the percentage of subjects meeting specific criteria for response defined by binary efficacy endpoints or mean treatment difference for continuous efficacy endpoints between each filgotinib group and placebo group or adalimumab group if applicable.

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

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 (e.g., more than 1 record on the same day with time missing), for any given subject, the worst outcome will be selected.

Calculation of Composite Endpoints

For the calculation of composite endpoints including DAS28(CRP), ACR20/50/70, MDA, VLDA, PASDAS, DAPSA, PsARC, we use the following steps unless otherwise specified:

- Step 1: Assign individual components to analysis visit windows defined in Section 3.7.2
- 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.

Missing Data Consideration

Unless specified otherwise, below are the consideration for the missing data 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 nonmissing components. If nonmissing components are not sufficient to determine final composite endpoint, then the composite endpoint will be set as missing. For continuous composite endpoints, if any components are missing, the composite endpoints will be set as missing.

If a subject only has baseline measurements, OC analyses will not include this subject.

For subjects who were re-randomized, the baseline value for the LTE phase will be the last nonmissing measurement obtained on or prior to the first dose date of filgotinib in the LTE. The change from baseline at analyses visits during LTE phase will be assessed from the re-randomized baseline of the LTE phase. The evaluation of response at analysis visits during LTE phase will be assessed from the re-randomized baseline of the LTE phase.

6.2. Primary Efficacy Endpoint

The primary efficacy endpoint is the proportion of subjects who achieve ACR20 response at Week 12.

6.2.1. Definition of the Primary Efficacy Endpoint

6.2.1.1. ACR20

A subject achieves ACR20 response when this subject has

- ≥ 20% improvement from baseline in TJC68, AND
- ≥ 20% improvement from baseline in SJC66, AND

- ≥ 20% improvement from baseline in at least 3 of the following 5 items:
  - PGADA (0 – 100 visual analog scale (VAS))
  - PhGADA (0 – 100 VAS)
  - HAQ-DI's pain assessment (0 – 100 VAS)
  - Patient's assessment of physical function as measured by HAQ-DI score
  - hsCRP

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

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

If the baseline value is 0 then the percent improvement from baseline is set to missing. In the case that some ACR20 components are missing, the ACR20 assessment will be based on the nonmissing components. If nonmissing components are not sufficient to determine ACR20 response, then the ACR20 response will be considered as missing.

6.2.1.2. ACR Components

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

TJC68 and SJC66 will be collected during the course of the study. The assessment for each joint will be from the following selections: Normal, Tender, Swollen, Tender and Swollen or Not able to evaluate.

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

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

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

If less than half of joints are assessed and evaluable 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) – hands only	10
Proximal Interphalangeal Joints I-V (Left and Right) – hands only	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:

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

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

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

6.2.1.2.2. Patient's Global Assessment of Disease Activity (PGADA)

PGADA is performed by patients. It is based on a 0 – 100 VAS, with a score of 0 indicating "very well" and 100 indicating "very poor" to the question "Considering all the ways psoriatic arthritis affects you, how well are you doing today?".

6.2.1.2.3. Physician's Global Assessment of Disease Activity (PhGADA)

PhGADA is performed by physicians. It is based on a 0 – 100 VAS with 0 indicates "no disease activity" and 100 indicates "maximum disease activity".

6.2.1.2.4. 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), administered by the subject. 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 nonmissing. Detailed algorithm for calculating HAQ-DI score is described in Appendix 1.

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 VAS, with 0 indicating "no pain" and 100 indicating "severe pain".

6.2.2. Primary Analysis of the Primary Efficacy Endpoint

To test for superiority of filgotinib 200 mg or filgotinib 100 mg group versus placebo control group in proportion of subjects who achieve ACR20 at Week 12, a logistic regression analysis with treatment groups and stratification factors in the model will be used. The model will include all treatment groups. Missing data on the ACR20 response rate at Week 12 will be imputed using multiple imputation (MI) assuming missing at random (MAR):

- The MI procedure replaces each missing binary ACR20 value with a set of plausible values that represent the uncertainty about the right value to impute. Twenty imputed datasets will be generated with starting seed 12345. These multiple imputed data sets are then analyzed by using the same method for the primary analysis for complete data as specified above. The results from each set of imputed data sets will then be combined using Rubin's rule {Rubin 1987}. The stratification factors and treatment groups will be included in the imputation model as covariates and all available data at post-baseline visits up to Week 16 visit will be included in the longitudinal model.

The p-value from the logistic regression model with MI for testing the superiority of filgotinib 200 mg or filgotinib 100 mg as compared to placebo will be provided. The 2-sided 95% confidence interval (CI) of the ACR20 response rate at Week 12 based on MI method will be provided for each treatment group. In addition, treatment difference in ACR20 response rate at Week 12 between each filgotinib dose group and placebo group along with its 95% CI calculated based on MI method will be provided.

6.3. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints are:

- The proportion of subjects who achieve ACR50 response at Week 12
- Change from Baseline in HAQ-DI at Week 12
- Change from Baseline in SF-36v2 physical component summary (PCS) at Week 16
- Change from Baseline in LEI at Week 16, in subjects with enthesitis at Baseline
- The proportion of subjects who achieve PASI75 response at Week 16, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- The proportion of subjects who achieve MDA response at Week 16
- Change from Baseline in FACIT-Fatigue at Week 16
- Change from Baseline in LDI at Week 16, in subjects with dactylitis at Baseline

6.3.1. Definition of Key Secondary Efficacy Endpoints

6.3.1.1. ACR50

ACR50 is similarly defined as ACR20 (see Section 6.2.1), except that the improvement threshold from Baseline is 50%.

6.3.1.2. 36-Item Short-form Health Survey (SF36v2)

The SF-36 Version 2 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.

6.3.1.3. Enthesitis Assessments: Leeds Enthesitis Index (LEI)

The enthesitis examination by LEI is based on 6 anatomical sites: medial femoral condyle (left and right), lateral epicondyle (left and right) and the achilles tendon insertion (left and right). Enthesitis at each site was scored on a dichotomous basis: 0 means Enthesitis Absent and 1 means Enthesitis Present.

LEI is derived as the sum of the enthesitis score over the 6 sites mentioned above.

If there are missing items, then available data are used to derive LEI.

The summaries and analysis of LEI will be conducted on the FAS subjects with enthesitis at Baseline.

6.3.1.4. Psoriasis Area and Severity Index (PASI)

PASI score is used to measure the severity and extent of psoriasis. The representative site of psoriasis is selected for each body region (head and neck (h), upper limb (u), trunk (t) and lower limbs (l)), and are separately scored by using 3 parameters, erythema (redness) (E), induration (thickeness) (I) and desquamation (scaling) (D), each of which is graded on a severity scale of 0 to 4 (0 = none, 1 = mild, 2 = moderate, 3 = severe and 4 = very severe). The area-wise percentage involvement of the involved sites (head and neck (Ah), upper limbs (Au), trunk (At), and lower limbs (Al)) is calculated as: 0 = no involvement, 1 = >0–< 10%; 2 = 10 – < 30%; 3 = 30 – < 50%; 4 = 50 – < 70%; 5 = 70 – < 90%; and 6 = 90 – 100%.

The final formula for PASI score is:

$$\underline{\text{PASI}} = 0.1 \times (\text{Eh} + \text{Ih} + \text{Dh}) \times \text{Ah} + 0.2 \times (\text{Eu} + \text{Iu} + \text{Du}) \times \text{Au} + 0.3 \times (\text{Et} + \text{It} + \text{Dt}) \times \text{At} + 0.4 \times (\text{El} + \text{Il} + \text{Dl}) \times \text{Al}$$

The PASI score ranges between 0 (no disease) and 72 (maximal disease), but is considered unreliable when BSA < 3%.

If any of the components is missing, then PASI score remains missing.

Percent improvement from baseline at a postbaseline visit is defined the same as in section 6.2.1.1. If the baseline value is 0 then the percent improvement from baseline is set to missing.

PASI75 response is defined for subjects achieved 75% improvement from Baseline in PASI score.

The summaries and analysis of PASI score will be conducted on the FAS subjects with psoriasis covering at least 3% of the BSA at Baseline.

6.3.1.5. Minimal Disease Activity (MDA)

MDA is a measure to indicate disease remission, and is based on a composite score of 7 domains.

A subject will be classified as having achieved MDA when at least 5 out of 7 of the following criteria are met.

- TJC68 ≤ 1,
- SJC66 ≤ 1,

- PASI ≤ 1 for subjects with a Baseline psoriasis covering at least 3% of the BSA (or for subjects with a Baseline psoriasis covering BSA < 3%, this criterion is considered met),
- Patent's Global Assessment of PsA pain intensity (PGAPI) score ≤ 15 (0 – 100 VAS),
- PGADA ≤ 20 (0 – 100 VAS),
- HAQ-DI ≤ 0.5,
- Leeds Enthesitis Index ≤ 1 for subjects with enthesitis at Baseline (or for subjects without enthesitis at Baseline, this criterion is considered met).

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

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

The FACIT-Fatigue scale and scoring guidelines are in Appendix 2.

6.3.1.7. Dactylitis Assessments: Leeds Dactylitis Index (LDI)

LDI measures the ratio of the circumference of the affected digit to the circumference of the digit on the contralateral hand or foot using a Leeds Dactylometer.

LDI score is calculated as follows:

$$\text{LDI score of a dactylitic finger/toe} = \left\{\left[\left(\frac{A}{B}\right) - 1\right] \times 100\right\} \times C$$

where

A = circumference of the dactylitic finger/toe (mm)

B = circumference of the contralateral digit (mm)

C = tenderness score (0 = no tenderness, 1 = tender)

If both ipsilateral and contralateral digits are thought to be dactylitic, then the reference range (given at the foot of the Dactylitis Score Sheet as in Protocol Appendix 25) will be used as the comparator (i.e., B in the above formula).

Binary tenderness score is used in LDI score calculation due to lack of reliability in the original Ritchie index. {Healy 2007}

Total LDI score equals to the sum of the non-negative individual LDI scores across all fingers and toes. Negative individual LDI scores will not be included in the total LDI score.

If there are missing fingers or toes, then the missing individual scores for those fingers or toes will not be included in the total LDI score. For subjects with dactylitis status absent for all the fingers and toes, their total LDI scores will be set as 0, and the tenderness score for each finger and toe will also be set as 0.

The summaries and analysis of LDI will be conducted on the FAS subjects with dactylitis at Baseline.

6.3.2. Analysis Methods for Key Secondary Efficacy Endpoints

For the superiority test of filgotinib 200 mg or filgotinib 100 mg group versus placebo control group in binary key secondary endpoints such as the ACR50 response rate at Week 12, the PASI75 response rate at Week 16 in subjects with psoriasis covering ≥3% of the BSA at Baseline and the MDA response rate at Week 16, the observed response rate and 2-sided 95% CI based on normal approximation with continuity correction will be provided by visit. Non-stratified response rate difference along with its 95% CI calculated based on the normal approximation with continuity correction will be provided. Appendix 4 provides sample SAS models statements for constructing the confidence interval for the proportion. In addition, the p-value based on the logistic regression model same as the one used in section 6.2.2 with observed cases for testing the superiority of filgotinib 200 mg or filgotinib 100 mg as compared to placebo will be provided as well. In case of limited data leading to improper use of logistic model, the p-value will not be provided.

For non-inferiority test of filgotinib 200 mg or filgotinib 100 mg group versus adalimumab group in the ACR20 response rate at Week 12, the approach proposed by {Liu 2014} will be used to demonstrate that each filgotinib dose preserves more than 50% of the effect of adalimumab on the response rate of ACR20 at Week 12. Let π_T, π_C and π_P denote the true response rates of ACR20 for filgotinib, adalimumab and placebo, respectively, at Week 12, with corresponding variances σ_T^2, σ_C^2 and σ_P^2. The non-inferiority hypotheses are

$$H_{0,NI}: \frac{\pi_T - \pi_P}{\pi_C - \pi_P} \leq 0.5$$

vs

$$H_{1,NI}: \frac{\pi_T - \pi_P}{\pi_C - \pi_P} > 0.5$$

According to {Liu 2014}, the test statistics for non-inferiority (filgotinib preserves more than 50% of the effect of adalimumab) is

$$Z_{NI} = \frac{\hat{\pi}_T - 0.5\hat{\pi}_C - (1-0.5)\hat{\pi}_P}{\sqrt{\hat{\sigma}_T^2 + 0.5^2\hat{\sigma}_C^2 + (1-0.5)^2\hat{\sigma}_P^2}}.$$

Note that the "hat" (^) denotes the estimated values of each parameter at Week 12. Missing data will be imputed using multiple imputation (MI) assuming that missing data are missing at random (MAR), which will follow the same method as described for the primary endpoint.

For the superiority test of filgotinib 200 mg or filgotinib 100 mg group versus placebo control group in continuous key secondary endpoints such as the change from baseline in HAQ-DI at Week 12, the change from Baseline in SF-36v2 PCS at Week 16, the change from Baseline in LEI at Week 16 for subjects with enthesitis at Baseline, the change from Baseline in FACIT-Fatigue at Week 16 and the change from Baseline in LDI at Week 16 for subjects with dactylitis at Baseline, a mixed-effect model repeated measures (MMRM) analysis that includes data at postbaseline visits through Week 16 will be applied. 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 Baseline, with 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 all treatment groups. An unstructured variance-covariance matrix will be used. The Kenward-Roger method will be used to estimate the degrees of freedom. Missing data will not be otherwise imputed using the MMRM approach. The least squares (LS) means and 95% CIs of the difference in mean change from Baseline between each filgotinib dose group and placebo group from MMRM will be provided at all the visits in Main study including Week 12 and Week 16.

In case of limited data leading to improper use of MMRM model, the descriptive statistics (sample size, mean, SD, median, Q1, Q3, minimum, and maximum) by treatment will be used for the continuous key secondary endpoints.

6.4. Other Secondary Efficacy Endpoint

Other secondary endpoints will be evaluated at all scheduled or applicable time points in the Main Study other than the time points that are already specified in the primary and key secondary endpoints.

CCI

The other secondary endpoints include:

- Change from Baseline in PASDAS

- MDA response
- VLDA response
- Change from Baseline in DAPSA
- Change from Baseline in PhGAP, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- Change from Baseline in mNAPSI, in subjects with psoriatic nail involvement at Baseline
- Change from Baseline in LEI, in subjects with enthesitis at Baseline
- Change from Baseline in PsAID-12
- PASDAS low disease activity (LDA, i.e. PASDAS ≤3.2)
- PASDAS remission (i.e. PASDAS ≤1.9)
- ACR20 response
- ACR50 response
- ACR70 response
- Change from Baseline in individual components of the ACR response criteria
- Change from Baseline in DAS28(CRP)
- DAS28(CRP) LDA
- DAS28(CRP) remission
- Time to achieve DAS28(CRP) LDA
- DAPSA LDA
- DAPSA remission
- Time to achieve DAPSA LDA
- PsARC response
- Change from Baseline in PASI, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- PASI50 response, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- PASI75 response, in subjects with psoriasis covering ≥3% of the BSA at Baseline

- PASI90 response, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- PASI100 response, in subjects with psoriasis covering ≥3% of the BSA at Baseline
- Change from Baseline in SPARCC Enthesitis Index, in subjects with enthesitis at Baseline
- Change from Baseline in LDI, in subjects with dactylitis at Baseline
- Change from Baseline in tender dactylitis count (TDC), in subjects with dactylitis at Baseline
- Change from Baseline in HAQ-DI
- Change from Baseline in FACIT-Fatigue
- Change from Baseline in SF-36v2 PCS and mental component score (MCS)

6.4.1. Definition of Other Secondary Endpoints

6.4.1.1. Tender Dactylitis Count (TDC)

Tender score (0 = no tenderness, 1 = tender, 2 = tender and wince, 3 = tender and withdraw) is collected for Dactylitis Assessments on the Dactylitis Score Sheet that was used for calculation of LDI total score.

Tender dactylitis count (TDC) equals the number of tender fingers and toes (tendor score >0). For subjects with dactylitis status absent for all the fingers and toes, the TDC will be set as 0. This score is used as a component of the PASDAS.

If there are missing fingers or toes, then the missing count for those fingers or toes will not be included in the TDC.

6.4.1.2. Psoriatic Arthritis Disease Activity Score (PASDAS), PASDAS Low Disease Activity (LDA) and PASDAS remission

$$\text{PASDAS} = \left[0.18 \times \sqrt{\text{PhGADA}} + 0.159 \times \sqrt{\text{PGADA}} - 0.253 \times \sqrt{\text{PCS}} + 0.101 \times \text{Ln}(\text{SJC66} + 1) + 0.048 \times \text{Ln}(\text{TJC68} + 1) + 0.23 \times \text{Ln}(\text{LEI} + 1) + 0.377 \times \text{Ln}(\text{TDC} + 1) + 0.102 \times \text{Ln}(\text{CRP} + 1) + 2\right] \times 1.5$$

where

PhGADA = Physician's Global Assessment of Disease Activity (0 – 100 VAS);

PGADA = Patient's Global Assessment of Disease Activity (0 – 100 VAS);

PCS = Physical Component Score of the SF-36;

LEI = Leeds Enthesitis Index, derived from the enthesitis assessments;

TDC = Tender Dactylitis Count from the LDI;

CRP is in the unit of mg/L.

A lower score indicates better function.

If a subject did not have dactylitis at baseline and hence no dactylitis assessments are collected post-baseline, then it is assumed that TDC = 0 at baseline and all post-baseline visits for calculating PASDAS.

If a subject did not have enthesitis at baseline and hence no enthesitis assessments are collected post-baseline, then it is assumed that LEI = 0 at baseline and all post-baseline visits for calculating PASDAS.

If any of the components is missing, then PASDAS remains missing.

PASDAS LDA is defined as PASDAS≤3.2; PASDAS remission is defined as PASDAS≤1.9.

6.4.1.3. Very Low Disease Activity (VLDA)

VLDA is defined based on the same 7 criteria as in MDA. A subject will be classified as having achieved VLDA when all 7 out of 7 of the criteria are met. In the case that some of the VLDA components are missing, the VLDA assessment will be based on the nonmissing components. If nonmissing components are not sufficient to determine the VLDA response, then the VLDA response will be considered as missing.

6.4.1.4. Disease Activity Index for Psoriatic Arthritis (DAPSA), DAPSA LDA, DAPSA remission and Time to Achieve DAPSA LDA

DAPSA = TJC68 + SJC66 + PGADA (0 – 100 VAS)/10 + PGAPI (0 – 100VAS)/10 + CRP (mg/dL).

where

PGADA = Patient's Global Assessment of Disease Activity (0 – 100 VAS);

PGAPI = Patient's Global Assessment of PsA Pain Intensity (0 – 100 VAS);

CRP is in unit of mg/dL.

If any of the components is missing, then DAPSA remains missing.

The derived DAPSA binary parameters include:

- DAPSA remission: DAPSA ≤ 4,
- DAPSA LDA: DAPSA ≤ 14.

The time to achieve DAPSA LDA is defined as:

Date of first (DAPSA ≤ 14) – First dose date of any study drug + 1

For each subject included in the analysis, all available measurement for DAPSA will be considered for the time to achieve DAPSA LDA based on the treatment policy estimand.

The time to achieve DAPSA LDA can be derived for the subject in each treatment group based the data during Main study phase. If the DAPSA LDA is not achieved during Main study phase, the time to achieve DAPSA LDA will be censored at the last nonmissing DAPSA assessment date during Main study phase. If the component scores of DAPSA are at different dates for a visit, the latest date will be used for the derivation of time to achieve DAPSA LDA.

6.4.1.5. Physician's Global Assessment of Psoriasis (PhGAP)

The physician will give a score on the subject's psoriasis disease activity, according to the following grades:

- Induration (I) (0 – 5 scale, averaged over all lesions),
- Erythema (E) (0 – 5 scale, averaged over all lesions),
- Scaling (S) (0 – 5 scale, averaged over all lesions).

The sum of the 3 grades will be divided by 3, i.e., = (I + E + S)/3, and rounded to the nearest integer (i.e., 0, 1, 2, 3, 4, 5) to obtain the total average score. Physician's Static Global Assessment (0 – 5 scale) will be based on the total average score: 0 = cleared; 1 = minimal; 2 = mild; 3 = moderate; 4 = marked; and 5 = severe.

If any of the grades above is missing, then the score remains missing.

Physician's Static Global Assessment (0-5 scale) will be the endpoint for analysis, which will be considered as continuous endpoint. The summaries and analyses will be conducted on the FAS subjects with psoriasis covering at least 3% of the BSA at Baseline.

6.4.1.6. Modified Nail Psoriasis Area and Severity Index (mNAPSI)

mNAPSI is used to assess each nail abnormality for each of subject's nail. Three features or groups of features (pitting, onycholysis together with oil-drop dyschromia, and crumbling) of each fingernail will be graded on a scale from 0 to 3. Four features (leukonychia, splinter hemorrhages, hyperkeratosis, and red spots in the lunula) will be graded as either present (1) or absent (0) for each fingernail. Each finger will have a score between 0 and 13. The total mNAPSI score is the sum of all abnormalities individual score across all fingers, and the total mNAPSI score ranges from 0 to 130.

In case of one or more missing digits, the available digits will be used to calculate mNAPSI.

The summaries and analyses of nail psoriasis will be conducted on the FAS subjects who had psoriatic nail involvement at Baseline.

6.4.1.7. 12-item Psoriatic Arthritis Impact of Disease (PsAID-12)

The PsAID is calculated based on 12 numerical rating scales (NRS) questions. Each NRS is assessed as number between 0 and 10.

$$\begin{aligned}\text{PsAID final value} = {} & (\text{PsAID1 (pain) NRS} \times 3) + (\text{PsAID2 (fatigue) NRS} \times 2) + \\ & (\text{PsAID3 (skin) NRS} \times 2) + (\text{PsAID4 (work } \textit{and / or} \text{ leisure activities) NRS} \\ & \times 2) + (\text{PsAID5 (function) NRS} \times 2) + (\text{PsAID6 (discomfort) NRS} \times 2) + (\\ & \text{PsAID7 (sleep) NRS} \times 2) + (\text{PsAID8 (coping) NRS} \times 1) + (\text{PsAID9 (anxiety)} \\ & \text{NRS} \times 1) + (\text{PsAID10 (embarrassment) NRS} \times 1) + (\text{PsAID11 (social life)} \\ & \text{NRS} \times 1) + (\text{PsAID12 (depression) NRS} \times 1)\end{aligned}$$

The total is divided by 20. Thus the range of the final PsAID value is 0-10 where higher figures indicate worse status.

If one of the 12 NRS values composing the PsAID is missing, the imputation is as follows: calculate the mean value of the 11 other (nonmissing) NRS (range, 0 – 10), impute this value for the missing NRS. Then, calculate the PsAID as explained above. If 2 or more of the NRS are missing, the PSAID is considered missing value (no imputation).

6.4.1.8. ACR70

ACR70 is similarly defined as ACR20 (see Section 6.2.1), except that the improvement threshold from Baseline is 70%.

6.4.1.9. Disease Activity Score for 28 joint count using C-reactive protein (DAS28 (CRP)), DAS28 (CRP) LDA, DAS28 (CRP) remission and Time to Achieve DAS28(CRP) LDA

The DAS28(CRP) score is calculated as follows:

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

where

Ln(CRP + 1) is the natural logarithm of (CRP value [mg/L] + 1);

PGADA is on 0 – 100 scale.

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

If any of the components is missing, the DAS28(CRP) will be set as missing.

The derived DAS28(CRP) binary parameters include:

- DAS28(CRP) remission: DAS28(CRP) < 2.6,
- DAS28(CRP) LDA: DAS28(CRP) ≤ 3.2.

The time to achieve DAS28(CRP) LDA is defined as:

Date of first [DAS28(CRP) ≤ 3.2] – First dose date of any study drug + 1

For each subject included in the analysis, all available measurement for DAS28(CRP) will be considered for the time to achieve DAS28(CRP) LDA based on the treatment policy estimand.

The time to achieve DAS28(CRP) LDA can be derived for the subject in each treatment group based the data during Main study phase. If the DAS28(CRP) LDA is not achieved during Main study phase, the time to achieve DAS28(CRP) LDA will be censored at the last nonmissing DAS28(CRP) assessment date during Main study phase. If the component scores of DAS28(CRP) are at different dates for a visit, the latest date will be used for the derivation of time to achieve DAS28(CRP) LDA.

6.4.1.10. Psoriatic Arthritis Response Criteria (PsARC) response

PsARC responder is defined as having an improvement in at least 2 of the 4 factors (with at least one factor being a joint count) and no worsening in the remaining factors:

- TJC68 (improvement defined as a decrease from baseline of at least 30%),
- SJC66 (improvement defined as a decrease from baseline of at least 30%),
- PGADA (0 – 100 VAS, improvement defined as a decrease from baseline of at least 20),
- PhGADA (0 – 100 VAS, improvement defined as a decrease from baseline of at least 20).

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

6.4.1.11. PASI50, PASI90, PASI100

PASI50, PASI90 and PASI100 response are defined similarly as PASI75, except that the improvement threshold from baseline are 50%, 90% and 100% in PASI score, respectively.

6.4.1.12. Spondyloarthritis Research Consortium of Canada (SPARCC) Enthesitis Index

The enthesitis examination is based on the 16 anatomical sites: the medial epicondyle (left and right), the lateral epicondyle (left and right), the supraspinatus insertion (left and right), the bilateral greater trochanter (left and right), the quadriceps tendon insertion into superior border of patella (left and right), the patellar ligament insertion into inferior pole of patella or tibial tuberosity (left and right), the achilles tendon insertion (left and right), and the plantar fascia insertion (left and right). Enthesitis at each site was scored on a dichotomous basis: 0 means Enthesitis Absent and 1 means Enthesitis Present.

SPARCC Enthesitis Index is derived as the sum of the enthesitis score over the 16 sites mentioned above.

If there are missing items, then available data are used to derive SPARCC Enthesitis Index.

The summaries and analyses of SPARCC Enthesitis Index will be conducted on the FAS subjects with enthesitis at Baseline.

6.4.2. Analysis Methods for Other Secondary Endpoints

The FAS will be used for all summaries and analyses of other secondary efficacy endpoints.

For the binary endpoints of the ACR20 response, the ACR50 response, the PASI75 response in subjects with psoriasis covering ≥3% of the BSA at Baseline and the MDA response, for subjects with observed outcome, the similar analysis as the key secondary endpoint analyses described in Section 6.3.2 will be applied. The proportions of subjects achieving above endpoints (observed case) will be plotted over time during Main study by treatment and by visit.

For all the other binary endpoints, only above mentioned observed response rate, 2-sided 95% CI and non-stratified response rate difference along with its 95% CI will be provided by visit.

For all the continuous endpoints, absolute value and change from Baseline will be summarized using descriptive statistics (sample size, mean, SD, median, Q1, Q3, minimum, and maximum) by treatment and by visit.

For the continuous endpoints of the change from baseline in HAQ-DI, the change from Baseline in SF-36v2 PCS, the change from Baseline in LEI for subjects with enthesitis at Baseline, the change from Baseline in FACIT-Fatigue and the change from Baseline in LDI for subjects with dactylitis at Baseline, plots of mean ± SD for change from baseline by visit will be presented.

For time to event analyses, Kaplan-Meier plot of each treatment group during Main study will be provided. Number of subjects having events and censored, KM estimates of median (95% CI), Q1, Q3, minimum and maximum of the time to events will be summarized by treatment group.

The by-subject listing of each endpoint and corresponding components, if available, including data during the Main Study and LTE phases will be provided.

6.5. Changes From Protocol-Specified Efficacy Analyses

Due to study termination and insufficient number of subjects enrolled, graphical approach of multiple test procedures will not be implemented, and only the nominal p values will be presented.

Due to small number of subjects available, for key secondary binary endpoints, logistic regression analyses may not be applicable. In such cases, no p-value will be provided. For key secondary continuous endpoints, MMRM analyses may not be applicable. In such cases, only descriptive summary statistics will be provided.

No sensitivity analyses will be performed.

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

7.1.2. Adverse Event Severity

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

7.1.3. Relationship of Adverse Events to Study Drug

Related AEs are those for which the investigator selected "Related" on the AE 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.

7.1.4. Serious Adverse Events

Serious adverse events (SAEs) will be identified and captured as SAEs if the AEs met the definitions of SAEs that were specified in the study protocol. SAEs captured and stored in the clinical database will be reconciled with the SAE captured in Gilead safety database 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 30 days after permanent discontinuation of study drug
- Any AEs leading to premature discontinuation of study drug.

For subjects who entered LTE, TEAEs with an onset date prior to first dose date of any study drug for LTE will be allocated to the Main Study phase. TEAEs with an onset date on or after the first dose date of any study drug for LTE will be allocated to the LTE phase.

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 30 days after the date of the last dose of study drug

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

For subjects who entered LTE, an AE with incomplete onset date that is prior to the month and year (or year) of the first dose date of any study drug for LTE will be allocated to the Main Study phase. An AE with incomplete onset date that is after the month and year (or year) of the first dose date of any study drug for LTE will be allocated to the LTE phase.

In addition, an AE with completely missing onset date or incomplete onset date that is same as the month and year (or year) of the first dose date of any study drug for LTE, and stop date that is prior to the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose date of any study drug for LTE, will be allocated to the Main Study phase. An AE with completely missing onset date or incomplete onset date that is same as the month and year (or year) of the first dose date of any study drug for LTE, and stop date that is missing, or the same as, or after the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose of any study drug for LTE will be allocated to the LTE phase.

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

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 the treatment group:

- TEAE (by maximum severity)
- TEAEs with Grade 3 or higher (by maximum severity)

- TE treatment-related AEs (by maximum severity)
- TE treatment-related AEs with Grade 3 or higher (by maximum severity)
- TE SAEs
- TE treatment-related SAEs
- TEAEs leading to premature discontinuation of any study drug
- TEAEs leading to premature discontinuation of study
- TE AEs leading to death (i.e., outcome of death)
- TEAEs leading to temporary interruption of any study drug

A brief, high-level summary of the number and percentage of subjects who experienced at least 1 TEAE in the categories described above will be provided by treatment group. All deaths observed in the Main Study will also be 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 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 SAEs
- All Deaths
- All SAEs leading to death (i.e., outcome of death)
- All AEs with severity of Grade 3 or higher
- All AEs leading to premature discontinuation of any study drug
- All AEs leading to premature discontinuation of study
- All 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 MedDRA search terms (MSTs). However, should additional cases not detected by the predefined search term listings identified during the clinical review process, these cases will also be reported by respective category.

7.1.7.1. Adjudication Committee for Major Adverse Cardiovascular Events (MACE) and Thromboembolic Events

An independent adjudication committee governed by a charter will be set up to perform adjudication of potential MACE as well as thromboembolic events reported during the study.

7.1.7.1.1. Major Adverse Cardiovascular Events (MACE)

MACE will be comprised of cardiovascular (CV) death, non-fatal myocardial infarction (MI) and non-fatal stroke. To identify the MACE, the following potential cases identified will be adjudicated:

- 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

7.1.7.1.2. Thromboembolic Events

Thromboembolism events will be comprised of arterial systemic thromboembolism events (ASTE) and venous thromboembolism events (VTE) associated with deep vein thrombosis (DVT) and/or pulmonary embolism (PE). To identify the thromboembolism events, the potential cases identified using the Embolic And Thrombotic Events SMQ search will be adjudicated.

The adjudication committee will review those potential MACE and thromboembolic events and related clinical data to determine whether a MACE or thromboembolism event has developed.

The number and percentage of subjects with positively adjudicated MACE or thromboembolic events will be summarized by treatment group using PT.

By-subject listings for all randomized subjects who have potential MACE or thromboembolic events will be provided respectively.

7.1.7.2. Other Adverse Events of Special Interest

In addition to general safety parameters, MACE and Thromboembolic Events, safety information on other adverse events of special interest (AESIs) will also be analyzed. AESIs will be identified by laboratory results, SMQs, 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
- Malignancy (including lymphoma; excluding nonmelanoma skin cancer)
- Nonmelanoma skin cancer
- Gastrointestinal (GI) perforations

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

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

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 30 days after the last dose date.

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 3.6.

A baseline laboratory value will be defined as the last nonmissing measurement obtained on or prior to the date/time of first dose of any study drug.

For subjects who were re-randomized, the baseline laboratory value for the LTE phase will be the last available measurement obtained on or prior to the first dose date of filgotinib. The lab abnormalities after the first dose date of filgotinib will be assessed from the re-randomized baseline.

For subjects who entered LTE phase, the laboratory collected on or prior to the first dose date of any study drug for LTE will be allocated to Main Study phase. The laboratory collected after the first dose date of any study drug for LTE will be allocated to LTE phase.

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 Baseline values, values at each postbaseline visit and change from Baseline at each postbaseline visit will be provided by treatment group for the following laboratory tests:

- Hematology
  - Hematocrit
  - Hemoglobin
  - Platelet count
  - Red blood cell (RBC) count
  - White blood cell (WBC) count
  - Mean corpuscular volume
  - Lymphocytes
  - Monocytes
  - Neutrophils
  - Eosinophils
  - Basophils

- Chemistry
  - Alkaline phosphatase (ALP)
  - Aspartate aminotransferase (AST)
  - Alanine aminotransferase (ALT)
  - Total bilirubin
  - Serum creatinine
  - Creatinine clearance by Cockcroft-Gault formula
  - Glucose
- Lipid
  - Triglycerides
  - Total cholesterol
  - HDL
  - LDL
  - LDL/HDL ratio

Change from baseline to a postbaseline visit will be defined as the visit value minus the baseline value. The mean, median, Q1, Q3, minimum, and maximum values will be displayed to the reported number of digits; SD values will be displayed to the reported number of digits plus 1.

Median (Q1, Q3) of the observed values for these laboratory tests will be plotted using 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 3.7.3.

7.2.2. Graded Laboratory Values

The CTCAE Version 5.0 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 (i.e., 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 baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days.

For subjects who entered LTE, treatment-emergent laboratory abnormalities during Main study phase are defined as values increase at least 1 toxicity grade from the baseline at any postbaseline time point, up to and including the first dose date of study drug for LTE. Treatment-emergent laboratory abnormalities during LTE phase are defined as values that increase at least 1 toxicity grade from the baseline (re-randomized baseline for re-randomized subjects) at any postbaseline time point after first dose date of study drug for LTE, up to and including the date of last dose of any study drug plus 30 days.

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

7.2.2.2. Treatment-Emergent Marked Laboratory Abnormalities

Treatment-emergent marked laboratory abnormalities are defined as values that increase at least 3 toxicity grade from baseline at any postbaseline time point, up to and including the date of last dose of any study drug plus 30 days.

For subjects who entered LTE, treatment-emergent marked laboratory abnormalities during Main study phase are defined as values increase at least 3 toxicity grade from the baseline at any postbaseline time point, up to and including the first dose date of study drug for LTE. Treatment-emergent marked laboratory abnormalities during LTE phase are defined as values that increase at least 3 toxicity grade from the baseline (re-randomized baseline for re-randomized subjects) at any postbaseline time point after first dose date of study drug for LTE, up to and including the date of last dose of any study drug plus 30 days.

If the relevant 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 in the study with the given response at 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.

By-subject listings 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.

7.2.3. 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:

- Aspartate aminotransferase (AST): (a) > 3 times of the upper limit of reference range (ULN); (b) > 5 x ULN; (c) > 10 x ULN
- Alanine aminotransferase (ALT): (a) > 3 x ULN; (b) > 5 x ULN; (c) > 10 x ULN
- AST or ALT: (a) > 3 x ULN; (b) > 5 x ULN; (c) > 10 x ULN
- Total bilirubin: > 2 x ULN
- Alkaline phosphatase (ALP) > 1.5 x ULN
- AST or ALT > 3 x ULN and total bilirubin: (a) > 1.5 x ULN; (b) > 2 x ULN

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

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

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

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

- Platelet count: (a) any postbaseline worsening CTCAE grade from baseline; (b) baseline value of less than Grade 3 and increase to Grade 3 or higher at worst postbaseline; (c) 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 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 pressures (mmHg), pulse (beats/min)) as follows:

- Baseline value

- Values at each postbaseline visit

- Change from baseline at each postbaseline visit

A baseline value will be defined as the last available value collected on or prior to the date/time of first dose of study drug.

For subjects who were re-randomized, the baseline value for LTE phase will be defined as the last available value collected on or prior to the date/time of the first dose of study drug for LTE.

For subjects who entered LTE phase, the body weight, BMI and vital signs collected on or prior to the first dose date of any study drug for LTE will be allocated to Main Study phase. The body weight, BMI and vital signs collected after the first dose date of any study drug for LTE will be allocated to LTE phase.

Change from baseline to a postbaseline visit will be defined as the postbaseline value minus the baseline value. Body weight and vital signs measured at unscheduled visits will be included for the baseline value selection.

In the case of multiple values in an analysis window, data will be selected for analysis as described in Section 3.7.3. No formal statistical testing is planned.

A by-subject listing of vital signs will be provided by subject ID number and 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

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

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

7.4.1. Prior Medications

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

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

For the purposes of analysis, any medications with a start date prior to or on the first dosing date of study drug and continued to be taken after the first dosing date, or started after the first dosing date but prior to or on the last dosing date of study drug will be considered concomitant medications. Medications started and stopped on the same day as the first dosing date or the last dosing date of study drug will also be considered concomitant. Medications that 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 study drug or a start date after the last dosing date of study drug will be excluded from the concomitant medication summary. If a partial stop date is

entered, any medication with the month and year (if day is missing) or year (if day and month are missing) prior to the date of first study drug administration will be excluded from the concomitant medication summary. If a partial start date is entered, any medication with the month and year (if day is missing) or year (if day and month are missing) after the study drug stop date will be excluded from the concomitant medication summary. Medications with completely missing start and stop dates will be included in the concomitant medication summary, unless otherwise specified. Summaries will be based on the Safety Analysis Set. No formal statistical testing is planned.

For subjects who entered LTE, concomitant medications with a start date prior to the first dose date of any study drug for LTE will be allocated to the Main Study phase. Concomitant medications with a start date on or after first dose date of any study drug for LTE will be allocated to the LTE phase.

For subjects who entered LTE, a concomitant medication with incomplete start date that is prior to the month and year (or year) of the first dose date of any study drug for LTE will be allocated to the Main Study phase. A concomitant medication with incomplete start date that is after the month and year (or year) of the first dose date of any study drug for LTE will be allocated to the LTE phase.

In addition, a concomitant medication with completely missing start date or incomplete start date that is same as the month and year (or year) of the first dose date of study drug for LTE, and stop date that is prior to the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose of study drug for LTE will be allocated to the Main Study phase. A concomitant medication with completely missing start date or incomplete start date that is same as the month and year (or year) of the first dose date of study drug for LTE, and stop date that is missing, or the same as, or after the date (or month and year if day is not recorded; or year alone if month is not recorded) of the first dose of study drug for LTE will be allocated to the LTE phase.

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. Other Safety Measures

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

7.6. Changes From Protocol-Specified Safety and Other Analyses

Safety analyses based on treatment policy estimand will not be performed.

CCI [redacted]

8. REFERENCES

Healy P, Helliwell P. Measuring dactylitis in clinical trials: which is the best instrument to use? Journal of Rheumatology 2007;34:1302-6.

Liu JT, Tzeng CS, Tsou HH. Establishing Non-Inferiority of a New Treatment in a Three-Arm Trial: Apply a Step-Down Hierarchical Model in a Papulopustular Acne Study and an Oral Prophylactic Antibiotics Study. International Journal of Statistics in Med Research 2014;3 (1):11-20.

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

9. SOFTWARE

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

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

10. SAP REVISION

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

11. APPENDICES

Appendix 1. Heath 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 subject. 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 nonmissing. 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 nonmissing)

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 19. 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 nonmissing 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 nonmissing, 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 2. FACIT-Fatigue Scale (Version 4) and Scoring Guidelines

FACIT-Fatigue Scale (Version 4)

Below is a list of statements that other people with your illness have said are important.
Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

		Not at all	A little bit	Some-what	Quite a bit	Very much
HI7	I feel fatigued	0	1	2	3	4
HI12	I feel weak all over	0	1	2	3	4
An1	I feel listless ("washed out")	0	1	2	3	4
An2	I feel tired	0	1	2	3	4
An3	I have trouble starting things because I am tired	0	1	2	3	4
An4	I have trouble finishing things because I am tired	0	1	2	3	4
An5	I have energy	0	1	2	3	4
An7	I am able to do my usual activities	0	1	2	3	4
An8	I need to sleep during the day	0	1	2	3	4
An12	I am too tired to eat	0	1	2	3	4
An14	I need help doing my usual activities	0	1	2	3	4
An15	I am frustrated by being too tired to do the things I want to do	0	1	2	3	4
An16	I have to limit my social activity because I am tired	0	1	2	3	4

FACIT-Fatigue Subscale Scoring Guidelines (Version 4)

Instructions:

5) Record answers in "item response" column. If missing, mark with an X.

6) Perform reversals as indicated, and sum individual items to obtain a score.

7) Multiply the sum of the item scores by the number of items in the subscale, then divide by the number of items answered. This produces the subscale score.

8) The higher the score, the better the QOL.

Subscale	Item Code	Reverse Item?	Item Response	Item Score
Fatigue Subscale	HI7	4 -	_____	=_____
	HI12	4 -	_____	=_____
Score range: 0 – 52	An1	4 -	_____	=_____
	An2	4 -	_____	=_____
	An3	4 -	_____	=_____
	An4	4 -	_____	=_____
	An5	0 +	_____	=_____
	An7	0 +	_____	=_____
	An8	4 -	_____	=_____
	An12	4 -	_____	=_____
	An14	4 -	_____	=_____
	An15	4 -	_____	=_____
	An16	4 -	_____	=_____
			Sum individual item scores:	=_____
			Multiply by 13:	=_____
			Divide by number of items answered:	=_____
				= Fatigue Subscale score

Appendix 3. Oral Corticosteroids Dose Conversion Chart

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
Deflazacort	7.5 mg
Methylprednisolone	4 mg
Prednisolone	5 mg
Prednisone	5 mg
Triamcinolone Acetonide	4 mg

Appendix 4. Sample SAS Code for Analysis of Binary Endpoints

The following model statement will be used to construct the confidence interval for the binomial proportion difference described in Section 6.3.2:

```
proc freq data = test ;
   by avisitn;
   tables TRT01PN*aval_ / riskdiff (CORRECT);
   output out = f200vp_ci RISKDIFF;
   where TRT01PN = 1 | TRT01PN = 3;
run;
```

where the confidence interval for the difference of two independent binomial proportions is constructed based on the normal approximation (ie, the Wald method) with a continuity correction, to adjust for the difference between the normal approximation and the binomial distribution, which is a discrete distribution:

$$(\hat{p}_1 - \hat{p}_2) \pm \left(\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) + z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

The following model statement will be used to construct the confidence interval for the binomial proportion described in the Section 6.3.2:

```
proc freq data = test ;
   by avisitn TRT01PN;
   tables aval_ / BINOMIAL (CORRECT);
run;
```

the confidence interval for the binomial proportions is constructed based on the normal approximation to the binomial distribution with continuity correction:

$$\hat{p}_1 \pm \left(\frac{1}{2}\left(\frac{1}{n_1}\right) + z_{\alpha/2} \times \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}\right)$$

Appendix 5. Determining Missing and Virtual visits due to COVID-19

This appendix describes the site collection of COVID-19 data as pertains to missed/virtual visits and the data processing algorithm used to determine which visits were missing and which visits were virtual.

Data collection

A COVID-19 supplement to the eCRF Completion Guidelines (CCG) was provided by data management to instruct clinical trial sites with respect to data entry expectations pertaining to scenarios related to the COVID-19 pandemic. If a visit was missed, sites should enter "Visit missed due to COVID-19." If a visit was conducted virtually, sites should enter "Virtual visit due to COVID-19."

Determination of Missed and Virtual visits

Natural Language Processing (NLP) was used to search the CRF comment fields to identify instances of "COVID-19" (or synonyms, see Table 11-1) and "Virtual" (or synonyms, see Table 11-1). The search terms are maintained in a global lookup and can be modified to tune the NLP model. For any comments with COVID-19 search terms, assign "Missed visit" or "Virtual visit" as follows

i) If COVID-19 terms are identified through NLP and the visit date is missing then result is 'Missed Visit'

ii) If COVID-19 and Virtual terms are identified through NLP for a visit, then result is 'Virtual Visit'. When there are multiple records for the same subject and the same visit, NLP will be based on multiple records to ensure 1 unique category per subject per visit

iii) Otherwise result is missing

Table 11-1. Examples of search terms for "COVID-19" and "Virtual" used to identify missed and virtual visits.

Search terms for "COVID-19"	Search terms for "Virtual"
COVID19	VIRTUAL
CORONA	TELEMED
CORONAVIRUS	TELEHEALTH
PANDEMIC	TELEPHONE
OUTBREAK	REMOTE
CRISIS	TELEMEDICINE
LOCKDOWN	TELECONSULTATION
QUARANTINE	TELEPHONICALLY
SHELTER	PHONE
	HOME VISIT
	ZOOM
	SKYPE

Appendix 6. Lists of PsA Medications

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

- ABATACEPT
- ADALIMUMAB
- ALEFACEPT
- BIMEKIZUMAB
- BRODALUMAB
- CERTOLIZUMAB PEGOL
- ETANERCEPT
- GOLIMUMAB
- GUSELKUMAB
- INFLIXIMAB
- IXEKIZUMAB
- SECUKINUMAB
- TOCILIZUMAB
- USTEKINUMAB

2) List of csDMARDs (WHO Preferred Terms)

- AZATHIOPRINE
- CICLOSPORIN
- HYDROXYCHLOROQUINE SULFATE
- IGURATIMOD
- LEFLUNOMIDE
- METHOTREXATE
- METHOTREXATE SODIUM
- SULFASALAZINE

3) List of Corticosteroids (WHO Preferred Terms)

- BETAMETHASONE ACETATE;BETAMETHASONE SODIUM PHOSPHATE
- BETAMETHASONE BUTYRATE PROPIONATE;MAXACALCITOL
- BETAMETHASONE DIPROPIONATE;BETAMETHASONE SODIUM PHOSPHATE
- BETAMETHASONE DIPROPIONATE;CALCIPOTRIOL
- CLOBETASOL
- CLOBETASOL PROPINOATE
- CORTISONE
- DEFLAZACORT
- METHYLPREDNISOLONE
- METHYLPREDNISOLONE ACETATE
- METHYLPREDNISOLONE SODIUM SUCCINATE
- PREDNISOLONE
- PREDNISOLONE VALEROACETATE
- PREDNISONE
- TRIAMCINOLONE
- TRIAMCINOLONE ACETONIDE

4) List of NSAIDs (WHO Preferred Terms)

- ACECLOFENAC
- CELECOXIB
- CODEINE PHOSPHATE;IBUPROFEN
- DEXKETOPROFEN TROMETAMOL
- DICLOFENAC
- DICLOFENAC POTASSIUM

- DICLOFENAC SODIUM
- ESOMEPRAZOLE MAGNESIUM;NAPROXEN
- ETODOLAC
- ETORICOXIB
- IBUPROFEN
- KETOPROFEN
- LOXOPROFEN SODIUM DIHYDRATE
- MELOXICAM
- NABUMETONE
- NAPROXEN
- NAPROXEN SODIUM
- NIMESULIDE
- SULINDAC
- TENOXICAM

GS-US-431-4566-SAP-v1.0

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy hh:mm:ss)
PPD	Clinical Research eSigned	16-Jul-2021 23:30:19
PPD	Biostatistics eSigned	20-Jul-2021 00:37:49