STATISTICAL ANALYSIS PLAN

Efficacy and safety of SAR156597 in the treatment of Idiopathic Pulmonary Fibrosis (IPF): A randomized, double-blind, placebo-controlled, 52-week dose-ranging study

SAR156597 - DRI11772

STATISTICIAN: [redacted]

DATE OF ISSUE: 16-Jun-2017

NCT02345070

Total number of pages: 82

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

6MWT:	6-meter walking test
AE:	adverse event
AESI:	adverse event of special interest
ALP:	alkaline phosphatase
ALT:	alanine aminotransferase
AST:	aspartate aminotransferase
ATC:	anatomic or therapeutic categories
BMI:	body mass idex
CI:	confidence interval
CIR:	copy increment from reference
CK:	creatine kinase
CMQ:	Customized MedDRA Query
DLCO:	diffusing capacity of the lung for carbon monoxyde
ESR:	erythrocyte sedimentation rate
FEV1:	forced expiratory volume in 1 second
FVC:	forced vital capacity
GGO:	Ground Glass Opacification
hsCRP:	high-sensitivity C-reactive protein
ILD:	Identifiable Causes of Interstitial Lung Disease
IMP:	Investigational Medicinal Product
ITT:	intent-to-treat
IVRS:	interactive voice response system
IWRS:	interactive web response system
PK:	pharmacokinetics
PPI:	proton pump inhibitors
RV:	residual volume
SAE:	serious adverse event
SMQ:	Standardized MedDRA Query
SpO2:	peripheral oxygen saturation
TEAE:	treatment-emergent adverse event
TLC:	total lung capacity
UIP:	usual interstitial pneumonitis
WHO-DD:	World Health Organization Drug Dictionary

1 OVERVIEW AND INVESTIGATIONAL PLAN

1.1 STUDY DESIGN AND RANDOMIZATION

This is a phase 2, proof-of-concept, multicenter, multinational, randomized, placebo-controlled, double-blind, 3-parallel-arm (placebo qw, SAR156597 200mg qw and SAR156597 200mg q2w alternating with placebo every other week), stratified study design.

After a screening phase of up to 4 weeks, patients will be centrally randomized (using allocation from block randomization schedule with stratifying factors) via IRT in a 1:1:1 ratio to 1 of the 3 treatment groups and treated double-blind for approximately 1 year.

Randomization will be stratified according to background therapy in authorized countries at the moment of randomization (patient with background therapy with either pirfenidone or nintedanib versus patient without background therapy). According to protocol amendment No. 10, it is planned to include a maximum of 66% of patient with background therapy.

Only patients with adjudicated IPF status were allowed to be enrolled in the study.

Approximately 300 patients (100 patients per treatment arm) will be recruited and randomized from 109 initiated sites.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of this study is to evaluate, in comparison with placebo, the efficacy of 2 dose regimens of SAR156597 (200 mg qw every week and 200 mg q2w every two weeks alternating with placebo every other week) administered subcutaneously during 52 weeks on lung function of patients with IPF.

1.2.2 Secondary objectives

The secondary objectives of this study are:

- To evaluate the efficacy of 2 dose levels/regimens of SAR156597 compared to placebo on IPF disease progression
- To evaluate the safety of 2 dose levels/regimens of SAR156597 compared to placebo in patients with IPF
- To evaluate the pharmacokinetics (through plasma concentrations) of SAR156597 200 mg administered subcutaneously during 52 weeks
- To assess the potential immunogenicity of SAR156597

- To evaluate the effect of SAR156597 on circulating biomarkers
- To explore the effect of SAR156597 on quality of life

1.3 DETERMINATION OF SAMPLE SIZE

Approximately 300 patients will be randomized in this study, 100 per treatment arm. Based on the enrollment in prior studies of this population approximately half of the patients are expected to be older than 65.

The sample size calculations are based on the primary efficacy variable of % predicted FVC change from baseline to week 52, with the following assumptions:

- A common standard deviation of 12% is assumed based on the data from previous trials (26)
- A 5% mean difference between each dose and placebo in change from baseline in % predicted FVC
- A 2-sided 5% significance level with 80% power

Ninety two (92) patients per treatment arm will yield 80% power to detect a 5 percentage point difference between the treatment groups and placebo. Additional patients will be randomized in each regimen to allow for dropouts; therefore, approximately 100 patients will be randomized into each treatment arm, for a total sample size of approximately 300 patients for the study.

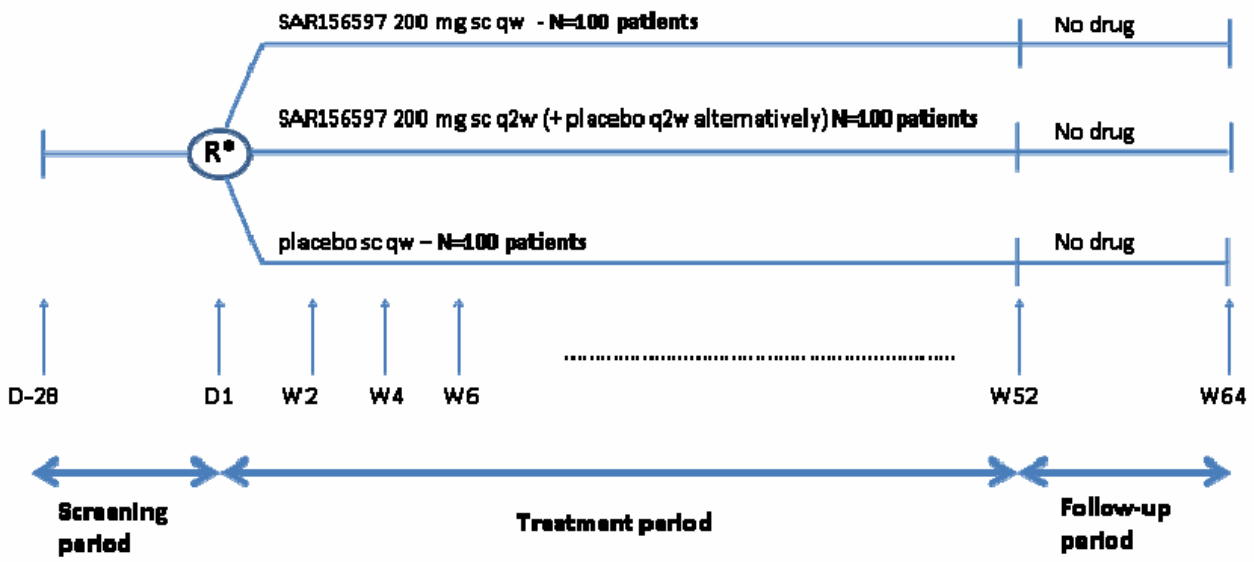
As a rank-based method will be used for analysis the actual power may differ from these calculations, however, calculations based on the ASCEND study (28), which used the same analysis method and endpoint confirm that the planned analysis should have at least 80% power. No adjustment is made for multiplicity since a hierarchical testing procedure is proposed.

Calculations were made using nQuery Advisor 7.0.

1.4 STUDY PLAN

The following figure presents the graphical study design:

Figure 1 - Graphical study design (screening period, double-blind period and follow-up period)



R*: randomization
Sc: subcutaneous
D: day
W: week

Approximately 300 patients will be randomized in a ratio 1:1:1 to the following three treatment arms:

- SAR156597 (qw) arm (n = 100): Patients will receive SAR156597 administered subcutaneously in 200 mg doses once every week (qw)
- SAR156597 (q2w) arm (n = 100): Patients will receive SAR156597 administered subcutaneously in 200 mg doses once every 2 weeks (q2w) alternating with placebo subcutaneously once every 2 weeks
- Placebo (qw) arm (n = 100): Patients will receive placebo subcutaneously once every week.

The study population will be stratified at baseline according to a background therapy with pirfenidone or nintedanib (standard of care for IPF in authorized countries).

The study will last up to 68 weeks as follows:

- Up to 4 weeks of screening
- 52 weeks of study treatment (from Day 1 dosing to Week 52 dosing)
- 12 weeks of follow-up

The total duration will be approximately 64 weeks (excluding screening period) per patient.

1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

This section summarizes major changes to the protocol statistical section with emphasis on changes after study start (after the first patient was enrolled). The protocol history table below gives the timing, rationale, and key details of major changes to the protocol statistical section.

The first patient was screened on 01-May-2015.

The first patient was randomized on 27-May-2015.

Table 1 - Protocol amendment statistical changes

Amendment Number	Date Approved	Rationale	Description of statistical changes
10	21-Jul-2015	An early analysis of the main efficacy endpoints and safety data will be performed after the last patient completed the End of treatment visit (Visit 9, week 52) to allow an evaluation of the benefit/risk for the treatment of IPF with SAR156597 and take the decision to move to Phase 3 of clinical development and prepare for future phase 3 studies.	Not applicable.
10	21-Jul-2015	In case of major recruitment issues, a futility analysis could be considered in order to make further decisions regarding trial continuation/discontinuation.	No major recruitment issues occurred during the enrollment therefore no interim analysis was performed.
10	21-Jul-2015	Due to the recent approval of pirfenidone (Esbriet®) and nintedanib (Ofev®) and thus the availability on the market in almost all countries participating in the study, the rate of patients with authorized background therapy is likely to be higher than that initially expected at time of study design. In consequence, in order to limit the impact of background therapy on efficacy and safety assessment of SAR156597 for the treatment of IPF, the rate of patients in the background therapy stratum will be capped to a maximum of 66% (i.e, 33% in the without background therapy stratum).	In protocol: stratification factor was already planned for primary analysis but was not described for secondary analyses. In SAP, description for secondary analyses and specification whether specific parameters will be described by With/Without Background Therapy status will be specified.
NA	NA	Duration of IMP exposure is defined as: last dose date – first dose date + 14 days, however for 200 mg SAR156597 q2w arm, there is an IMP injection every two weeks alternating with placebo the other week.	Duration of IMP exposure will be defined as: last dose date – first dose date + 7 days.
NA	NA	Change in mITT definition due a too low selection of patients (more than 5% of patients were discarded with the protocol definition).	The new definition is all randomized patients who take at least one dose or part of a dose of IMP. Patients will be analyzed using the treatment arm to which they were randomized. This results in considering all mITT patients

Amendment Number	Date Approved	Rationale	Description of statistical changes
			in efficacy analysis provided that the endpoint was measured or could be calculated (e.g. at least 1 baseline and 1 post-baseline values for imputing and calculating a change).
NA	NA	Baseline definition has been revised because the timeslot between a screening and a baseline value is large (up to 4 weeks) and there will not be several assessments at 52 weeks.	Last available value before first IMP administration.
NA	NA	%predicted DLCO analysis is of interest, it was missing in protocol.	Addition of the %predicted DLCO exploratory endpoint analysis.
NA	NA	Primary efficacy endpoint - change in week 52 missing data (due to death or lung-transplant) imputation approach: the time-to-death or lung transplant approach is more sensitive in ranking patients. For example, a patient who dies early will be more penalized than a patient who dies lately during the study. Primary efficacy endpoint - change in week 52 missing data (due to reasons other death or lung-transplant) imputation approach: • patients taking treatment at 52 weeks will be imputed using a linear regression run on their treatment arm completers. • patients not taking treatment at 52 weeks will be imputed using a linear regression run on the placebo completers.	Time-to-death or lung transplants will be used instead of allocating the lowest rank for all died and lung transplanted patients. For patients with a week 52 missing data (due to reasons other than death or lung transplant), depending on treatment intake at 52 weeks, a linear regression will be performed on a particular set of completers.
NA	NA	Disease progression endpoint - change in analysis cut-off date: Day 375 will be used instead of superior boundary of week 52 analysis window (Day 406) in order to prevent from non-random censures.	Patients not meeting the definition for disease progression will be considered censored at the time of their week 52 FVC or DLCO assessment provided that it was performed before Day 375. For patients with a missing week 52 assessment, the censoring time will be the time of the last measured FVC or DLCO assessment. For patients with week 52 assessment performed after Day 375, the censoring time will be Day 375.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

Not applicable.

2 STATISTICAL AND ANALYTICAL PROCEDURES

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value is defined as the last available value obtained before or equal to the date and time of the first double-blind Investigation Medicinal Product (IMP) administration. For patients randomized and not treated, the baseline value is defined as the last available value obtained before or equal to the date and time of randomization.

All baseline safety and efficacy parameters are presented along with the on-treatment summary statistics in the safety and efficacy sections (Section 2.4.4 and Section 2.4.5).

Demographic characteristics

Demographic variables at baseline are gender (Male, Female), race (Caucasian/White, Black, Asian/Oriental, Other), age in years (quantitative and qualitative variable: <40, ≥40 and >65, ≥65 and <75, ≥75 years), ethnicity (Hispanic/Latino, Not Hispanic/Latino).

Medical or surgical history

Medical (or surgical) history includes (but are not limited to) the following previous relevant medical and surgery history collected at baseline:

- All prior significant medical conditions, and/or hospitalization history: eg. pneumonia, asthma, Chronic Obstructive Pulmonary Disease (COPD), diabetes, hypertension, pulmonary arterial hypertension, congestive heart failure, esophageal reflux, depression, sleep apnea, myocardial infarction or arrhythmias, need for major surgery or surgical complications.

Habits

Habits includes:

- smoking habits (never smoke, quit smoking (have smoked regularly in their lifetimes but have stopped at least 8 days ago), currently smokes (smoked at least 1 cigarette, as a mean, per day during the past 7 days)), pack years (if currently smoking or quit smoking: One pack year is smoking 20 cigarettes a day for one year), mean smoking duration for quit and current smokers in years.
- alcohol habits (frequency of alcoholic drinks in the last 12 months (Never/At least monthly/At least weekly/At least daily/Occasionally), number of standard drinks (1 or 2/Greater than 2 where standard drink means 1 pint/bottle of beer, 1 glass of wine, 1 shot of hard liquor…) per day when drinking alcohol)

Disease characteristics at baseline

Specific disease history includes:

- IPF diagnosis algorithm: Identifiable Causes of Interstitial Lung Disease (ILD) (yes, no), Diagnosis confirmation by Central review (yes, no), IPF diagnosis based on HRCT alone OR HRCT and Surgical Lung biopsy
- HRCT UIP (usual interstitial pneumonitis) determination during adjudication (Definite UIP, Probable UIP, Possible UIP and Definitely not UIP)
- Lung Biopsy UIP determination during adjudication (Definite UIP, Probable UIP, Possible UIP and Definitely not UIP)
- Time since diagnosis of IPF (in years and by category: <1, 1-≤3 and >3 years)
- Time since first symptoms of IPF (years)
- Any acute exacerbation within the past 12 months before enrollment (yes, no) and if yes, time from most recent date of acute exacerbation (months and by category (<6 months and ≥6 months))
- Family IPF history (yes, no, unknown)
- Oxygen therapy need (yes, no) and if yes number of hours the patient is on supplemental oxygen therapy per day
- Currently taking pirfenidone or nintedanib for at least 12 weeks prior to enrollment (yes, no) and if yes, will continue with pirfenidone or nintedanib during the study (yes, no)

The following efficacy parameters will be also summarized at baseline (see definition in Section 2.1.3):

- FVC (forced vital capacity): L and % predicted (quantitative and by tercile (≤60, >60 and ≤75, >75)),
- FEV1 (forced expiratory volume in 1 second): L and % predicted,
- FEV1/FVC ratio,
- % predicted DLCO (diffusing capacity of the lung for carbon monoxide): quantitative and by tercile (≤40, >40 and ≤55, >55),
- residual volume (RV): L,
- total lung capacity (TLC): L,
- SpO2 (peripheral oxygen saturation): % under oxygenotherapy and under ambient air,
- SGRQ total score,
- EQ-5D-5L total score,
- 6MWT (6-meter walking test) distance (m),
- Total Fibrosing ILD score.

Other baseline characteristics

Other baseline characteristics include stratification factor (With/Without Background Therapy) as per IVRS/IWRS and e-CRF (patients taking at least a dose or part of a dose of Nintedanib or Pirfenidone at randomization (Visit 2/Day 1) will be considered in the With Background Therapy stratum).

Any technical details related to computation, dates and imputation for missing dates are described in Section 2.5.

2.1.2 Prior or concomitant medications

All medications taken before randomization and until the end of the study are to be reported in one of the specific case report form pages:

- Pirfenidone or Nintedanib (within 12 weeks prior to randomization (baseline) and during the course of the study) on the dedicated page using ATC code,
- Systemic Corticosteroids category will gather the routes described in Appendix C (within 4 weeks prior to randomization (baseline) and during the course of the study).
- Other prior and concomitant specific medications (within 30 days before randomization (baseline) and during the course of the study) on the dedicated page:
  - In particular, N-acetylcysteine (administered at the dosage recommended by Czech Republic guidelines) will be selected using a WHO-DD codelist.
  - Anti-acids will be selected using a WHO-DD codelist.
    - in particular, H2 receptor blocker will be selected using a WHO-DD codelist
    - proton pump inhibitors (PPI) will be selected using a WHO-DD codelist.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version currently in effect at Sanofi at the time of database lock. All WHO-DD codelists used in this study are detailed in Appendix F.

- Prior medications are those the patient used from 12 weeks prior screening up to first investigational medicinal product (IMP) intake. Prior medications can be discontinued before first administration or can be ongoing during treatment phase.
- Concomitant medications are any treatments received by the patient concomitantly to the IMPs, from first IMP administration to the last IMP administration + 84 days. A given medication can be classified both as a prior medication and as a concomitant medication. Concomitant medications do not include medications started during the post-treatment period (as defined in Section 2.1.4).
- Post-treatment medications are those the patient took in the period running from the day after the last IMP administration + 84 days period (ie, 85 days after the day of last dose of double-blind IMP injection).

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.5.

2.1.3 Efficacy endpoints

Unless otherwise specified, all efficacy parameters values (scheduled or unscheduled) may be used to provide a value for the primary and secondary efficacy endpoints.

For FVC, FEV1 and FEV1/FVC parameters, a value is considered valid if this is an ATSBEST value with a grade “A” or “B” and a status “CLEAN” or “MANUAL”.

All measurements (scheduled or unscheduled) will be assigned to analysis windows defined in Table 6, in order to provide an assessment of the efficacy parameter at each time point planned to be collected as per protocol. For all post-baseline time points, the value used for the analyses at a given time point (e.g at week 52) is the valid value obtained within the corresponding analysis window. If multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the best value will be considered.

The baseline value is the last available and valid measurement obtained before or equal to the date and time of the first double-blind IMP injection. For patients randomized and not treated, the baseline value is defined as the last available and valid value obtained before or equal to the date and time of randomization.

2.1.3.1 Primary efficacy endpoint(s)

The pulmonary function test % predicted FVC is planned to be assessed at baseline, Week 12 (Visit 6), Week 24 (Visit 7), Week 36 (Visit 8), Week 52 (Visit 9/End of treatment) and Week 64 (Visit 10/End of study).

The primary efficacy endpoint of the study is the absolute change from baseline in % predicted FVC at 52 weeks in the mITT population.

2.1.3.2 Secondary efficacy endpoint(s)

2.1.3.2.1 Key secondary efficacy endpoints

Two key secondary endpoints will evaluate other aspects of the efficacy of SAR156597 in IPF:

- Patients experiencing Disease progression at 52 weeks in the mITT population
- Deaths (all-cause) at 52 weeks in the mITT population

2.1.3.2.2 Disease progression endpoint

Disease progression is defined as time from randomization to the first occurrence of any of the following events:

- decrease in absolute % predicted FVC≥10% (as in Section 2.1.3.1)
- decrease in absolute % predicted DLCO≥15% (as in Section 2.4.4.2.3)
- lung transplant (recorded on the dedicated e-CRF page during the study)
- death (recorded on the dedicated e-CRF page during the study)

Disease progression will be analyzed as a time-to-event variable: event date – randomization date + 1 day where the event date is described in Table 2.

Table 2 - Time-to-event variable definition for disease progression

Scenarii	Event status	Event date
First event up to Day 375 is decrease in absolute % predicted FVC≥10% or decrease in absolute % predicted DLCO≥15%	event	Date of the 1st decrease observed
First event up to Day 375 is death or lung transplant	event	Date of death or date or lung transplant
No event prior to last available and valid assessment of %predicted FVC or last available assessment of %predicted DLCO	censored	Min(date of last available %predicted FVC, date of last available %predicted DLCO)
No event up to Day 375 (last available and valid assessment of %predicted FVC and last available assessment of %predicted DLCO performed after Day 375)	censored	Date of Day 375

2.1.3.2.3 All-cause mortality endpoint

Deaths are reported during the study on the Death e-CRF page.

All-cause mortality endpoint will be analyzed as a time-to-event variable: event date - randomization date + 1 day where the event date is described in Table 3.

Table 3 - Time-to-event definition for all-cause mortality

Scenarii	Event status	Event date
Death up to Day 375	event	date of death
Lung transplant	event	Date of transplant
Study discontinuation (according to EOS e-CRF page) before Day 375	censored	Min{max(date of last available information, SAE recovery date), date of Day 375}
No death up to Day 375	censored	Date of Day 375

2.1.3.2.4 Exploratory efficacy endpoints

The other secondary efficacy endpoints are considered exploratory and include:

- Rate of acute IPF exacerbations at 52 weeks in the mITT population
- Rate of respiratory hospitalizations and rate of non-elective hospitalizations at 52 weeks in the mITT population
- Change from baseline in 6-MWT distance at 52 weeks in the mITT population
- Change from baseline in HRCT lung interstitial score at 52 weeks in the mITT population
- Change from baseline in the total score of the SGRQ at 52 weeks in the mITT population
- Change from baseline in the total score of the EQ-5D-5L at 52 weeks in the mITT population

2.1.3.2.5 Acute IPF exacerbations endpoint

All acute IPF exacerbations events are recorded during the study on the AE e-CRF page. To help the investigator in diagnosing an acute IPF exacerbation, an "IPF Exacerbation Complementary" e-CRF form has to be completed.

For efficacy analysis, only acute IPF exacerbations with a confirmed diagnosis on the "IPF Exacerbation Complementary" e-CRF form and occurring from randomization to Day 375 will be considered.

In addition, the following information is recorded on this complementary form:

- Diagnosis of AE,
- Clinical evaluations
  - unexplained worsening or development of dyspnea within 30 days,
    - if no what is the symptom or finding of the patient related to this AE,
  - worsening of hypoxemia from a known baseline arterial blood gas,
  - pulmonary infection ruled out by endotracheal aspiration or broncho-alveolar lavage (BAL),
    - if no, was pulmonary infection ruled out by routine clinical practice and/or microbiological studies
  - have alternative causes (including left heart failure, P.E., acute lung injury, tec.) been excluded
- Radiology assessment
  - HRCT performed
    - If yes, did it show new ground-glass abnormality and/or consolidation consistent with a UIP pattern

- Final diagnosis
  - Based on response provided, do you believe that the patient experienced an acute IPF exacerbation
- Management
  - Was corrective treatment/therapy used

2.1.3.2.6 Respiratory and non-elective hospitalizations endpoints

All hospitalizations are recorded during the study on the "Hospitalization" e-CRF page.

The respiratory hospitalizations occurring from randomization to Day 375 will be considered will be selected from the "Hospitalization" e-CRF page using reason = "Respiratory".

The non-elective hospitalizations occurring from randomization to Day 375 will be considered will be selected using the Yes response to the "nonelective hospitalization" item on the "Hospitalization" e-CRF page.

2.1.3.2.7 Change from baseline in 6-MWT distance endpoint

The total distance covered by the patient over a 6 minute walk is planned to be measured at Visit 2 (Day 1) predose, Visit 6 (Week 12), Visit 7 (Week 24), Visit 8 (Week 36), Visit 9/EOT (Week 52), Visit 10/EOS (Week 64), and reported in meters on the "6 Minute Walk Test - 6MWT" e-CRF page.

Additional information to assess quality will be reportedusing the "6 Minute Walk Test - 6MWT" e-CRF page:

- the completion of the 6 minute walk test (If No, what was the time patient walked? (in seconds)),
- the need of any pauses or breaks (if Yes, duration of pause (in seconds) and reason for pause were recorded).

Of note: If during visits 6, 7, 8, 9 and 10, oxygen saturation is less than 88% under ambient air at rest, the patient should perform 6-MWT under 4L oxygen.

2.1.3.2.8 Change from baseline in HRCT lung interstitial score endpoint

The HRCT lung interstitial score is adjudicated in a blinded manner, and planned to be assessed at screening (unless already done within 1 year prior to screening) and Visit 9/EOT (Week 52) (in Germany, second assessement at 52 weeks based on Investigator's judgment).

The Total Fibrosing ILD, Reticular, Ground Glass Opacification (GGO), Honeycomb and Emphysema scores are recorded in six lobar regions (namely: RUL, LUL, RML, LML, RLL and LLL).

For each lobar region:

- The sum of the 3 patterns (Reticular, GGO and Honeycomb) should be 100%
- The Total Fibrosing ILD and the Emphysema scores ranges from 0 to 100%

The score of interest will be the Total Fibrosing ILD score averaged on the 6 lobar regions.

Each pattern score and Emphysema total score are the mean of values observed in the 6 lobar regions.

In case of image quality issue, a value NE will be recorded. For any other issue, a value NA will be recorded. A score will be calculated only if at least 4 of the 6 lobar regions are filled in (other than NE or NA).

2.1.3.2.9 Change from baseline in total score of St George Respiratory Questionnaire (SGRQ) endpoint

Impact of IPF on quality of life health status will be assessed using SGRQ planned to be measured at Visit 2 (Day 1) predose, Visit 6 (Week 12), Visit 7 (Week 24), Visit 8 (Week 36), Visit 9/EOT (Week 52) and Visit 10/EOS (Week 64).

Three component scores are calculated for the SGRQ:

- **Symptoms:** effect of respiratory symptoms, their frequency and severity
- **Activity:** activities that cause or are limited by breathlessness
- **Impacts:** range of aspects concerned with social functioning and psychological disturbances resulting from airways disease

A **Total** score is also calculated which summarizes the impact of the disease on overall health status.

Scores are expressed as a percentage of overall impairment where 100 represents worst possible health status and 0 indicates best possible health status (1).

Data handling conventions and detailed definitions are described in Table 5.

2.1.3.2.10 Change from baseline in total score of EuroQol Questionnaire, 5 level system (EQ-5D-5L) endpoint

EQ-5D is a generic measure of health status developed by the EuroQol Group in order to provide a simple, measure of health for clinical and economic appraisal. It was designed for self-completion by patients.

EQ-5D-5L is planned to be measured at Visit 2 (Day 1) predose, Visit 8 (Week 36), Visit 9/EOT (Week 52) and Visit 10/EOS (Week 64).

EQ-5D-5L is the 5-level system for the 5-dimensional format of the EQ-5D and covers the measurement of mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. Scores on the 5 levels are used to create an index score derived from population weights. The United Kingdom population-based scoring will be used to derive an index score from -0.594 to 1.0 renamed the EQ-5D-5L total score.

In addition, the EQ-5D has a single visual analogue scale (VAS) (0-100) to evaluate perceived current health.

Of note: There should be only one response for each dimension. Missing values can be coded as '9' and ambiguous values (e.g. 2 boxes are ticked for a single dimension) should be treated as missing values (2).

Further explanations are available in Appendix D.

2.1.4 Safety endpoints

The safety analysis will be based on:

- the reported adverse events (AEs) including serious adverse events (SAEs) and adverse events of special interest (AESIs)
- and other safety information, such as:
  - clinical laboratory data,
  - vital signs,
  - electrocardiogram ECG.

Observation period

The observation period will be divided as follows:

- The **pre-treatment** period is defined as the time from the signed informed consent date up to the first administration of IMP.
- The **treatment-emergent adverse event** (TEAE) period is defined as the time from the first administration of the IMP to the last administration of the IMP + 84 days. It will include both **treatment** and **residual treatment** periods:
  - The **treatment** period is defined as the time from the first administration of the IMP to the last administration of the IMP + 7 days
  - The **residual treatment** period is defined as the time from the last administration of the IMP + 8 days to the last administration of the IMP + 84 days.
- The **post-treatment** period is defined as the period of time starting the day after the end of the TEAE period (ie, 85 days after the day of last dose of double-blind IMP injection).

The on-study observation period is defined as the time from the first administration of IMP until the end of the study (defined as last protocol planned visit (end of study visit) or the resolution/stabilization of all SAEs and adverse events of special interest (AESIs)).

2.1.4.1 Adverse events variables

Adverse event observation period

- Pre-treatment adverse events are AEs that developed or worsened or became serious during the pre-treatment period
- Treatment-emergent adverse events are AEs that developed or worsened or became serious during the TEAE period
- Post-treatment adverse events are AEs that developed or worsened or became serious during the post-treatment period

All adverse events (including SAEs and AESIs) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

Record the occurrence of AEs (including SAEs and AESIs) from the time of signed informed consent until the end of the study.

Adverse events of special interest (AESIs) include the following terms (their complete descriptions are provided in the protocol):

- Pregnancy of a female subject entered in a study as well as pregnancy occurring in a female partner of a male subject entered in a study with IMP selected using the e-CRF "Pregnancy" tick box on the AE page.
- Symptomatic overdose (serious or non-serious) with IMP, selected using the e-CRF "Symptomatic overdose of the IP" tick box on the AE.
- Increase in ALT ≥3·ULN (if baseline ALT <ULN) or ≥2·baseline value (if baseline ALT ≥ULN), selected using laboratory data.
- [redacted]
- Hypersensitivity:
  - Anaphylactic reactions or acute allergic reactions that require immediate treatment, selected using a CMQ coding list.
  - Severe injection site reactions, selected using HLT="Injection site reaction" and an intensity equal to severe.
- Tuberculosis, selected using a CMQ coding list.
- Initiation of medications for suspected tuberculosis, selected using a WHODD CDG00737 "initiation of medications for suspected tuberculosis".

- Acute renal failure, selected using a CMQ coding list.

The summary of CMQ coding lists to be used in determining these AESIs is displayed in Appendix E.

2.1.4.2 Deaths

The deaths observation period are per the observation periods defined above.

- Death on-study: deaths occurring during the on-study observation period
- Death on-treatment: deaths occurring during the TEAE period
- Death post-study: deaths occurring after the end of the study visit

2.1.4.3 Laboratory safety variables

Clinical laboratory data consists of blood analysis, including hematology, clinical chemistry, and urinalysis. Clinical laboratory values after conversion will be analyzed into standard international units and international units will be used in all listings and tables.

The laboratory parameters will be classified as follows:

- Hematology
  - **Red blood cells and platelets and coagulation**: hemoglobin, hematocrit, red blood cell count (erythrocytes), platelet count, erythrocyte sedimentation rate (ESR)
  - **White blood cells:** white blood cell count with differential (leukocytes, neutrophils, lymphocytes, monocytes, basophils, eosinophils)
  - **Tuberculosis screen:** a Quantiferon®-TB gold evaluation.
- Chemistry
  - **Metabolism:** glucose, total protein, albumin, creatine kinase (CK), high-sensitivity C-reactive protein (hsCRP), Hemoglobin A1C, Cholesterol, Triglycerides
  - **Electrolytes**: sodium, potassium, chloride, calcium, phosphorous, bicarbonate
  - **Renal function**: creatinine, creatinine clearance, blood urea nitrogen, urate
  - **Liver function**: alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase (ALP), bilirubin
  - **Pregnancy test**: (For women of childbearing potential) β-HCG blood test, urine pregnancy tests.
  - [redacted]
- Serology
  - **Hepatitis screen:** hepatitis B surface antigen, hepatitis B surface antibody, hepatitis B core antibody and hepatitis C antibodies

- [redacted]

- Cytometry
  - [redacted]

- Urine samples will be collected as follows:
  - **Urinalysis** (by dipstick): pH, specific gravity, protein, glucose, ketones, occult blood, nitrite, leukocyte esterase, urobilinogen and bilirubin. If any parameter on the dipstick is abnormal, a urine sample should be sent to the central laboratory for testing. If positive for proteins, microscopic analysis is performed by central laboratory.

Technical formulas are described in Section 2.5.1.

2.1.4.4 Vital signs variables

Vital signs include: heart rate (beats per minute), diastolic and systolic blood pressure (mmHg) after 5 minutes of resting in sitting position and body weight (kg) and height (cm). body mass index (BMI) is calculated automatically at each visit on e-CRF page.

In addition to quantitative variables, the following qualitative variables will be displayed at baseline for weight in kilograms (<50, ≥50 and <100, ≥100), body mass index (BMI) in kg/m² (<30, ≥30).

Moreover, the values and changes from baseline in SPO2 over time will be displayed from the "Pulse Oximetry - Oxygenotherapy" e-CRF page and "Pulse Oximetry – Ambient air" e-CRF page separately.

2.1.4.5 Electrocardiogram variables

A standard 12-lead ECG will be performed and evaluate locally.

ECG parameters include heart rate (beats per minute), PR interval (msec), QRS duration (msec), QT interval (msec), ST deviation (mm), T-wave and U-wave morphologies as normal or abnormal. In addition, ECG assessments will be described as normal or abnormal (clinically significant or not) according to the Investigator's interpretation on the ECG eCRF page. The corrected QTc (according to Bazett and Fridericia) will be derived afterwards.

2.1.5 Pharmacokinetic variables

All measurements (scheduled or unscheduled) will be assigned to analysis windows defined in Table 7, in order to provide an assessment of the PK parameter at each time point planned to be collected as per protocol.

Samples for PK analysis are planned to be collected at pre-dose (within 2 hours before each dose administration) during Visits 2 (Day 1), 4 (Week 4), 6 (Week 12), 7 (Week 24), 8 (Week 36) and 9/EOT (Week 52). Pharmacokinetics samples during Visit 10 will be collected in the morning. An additional sample should be taken in a subgroup (at least 30 patients by cohort) between 5 and 10 days after last dose (i.e. after Visit 9). The possibility to perform this additional sample will be discussed between Investigator and patient on a case by case basis.

Pharmacokinetic variable is the SAR156597 concentration at each time point using the time windows defined in Table 8. Depending on the timing of the sample versus the previous injection, C_{trough}, $C_{trough,av,}$ C_{max} and $C_{Follow-Up}$ will be defined as follows:

- C_{max}: SAR156597 concentration sample taken 5 to 10 days after last injection (week 52);
- C_{trough}: SAR156597 concentration sample within 2h before each administration (day 1, week 4, week 12, week 24, week 36 and week 52);
  - $C_{trough,av,ss}$: SAR156597 average trough concentrations ("$C_{trough,av,ss}$") will be calculated for each patient as the mean of "$C_{troughs}$" considered at steady state (ss). The occurrence of steady state will be assessed graphically, by plotting "C_{trough}" throughout the study visits over all patients.
- $C_{Follow-up}$: SAR156597 concentration sample taken in week 64, 14 weeks after last injection.

2.1.6 Immunogenicity variables

All measurements (scheduled or unscheduled) will be assigned to analysis windows defined in Table 7, in order to provide an assessment of the ADA parameter at each time point planned to be collected as per protocol.

Anti-SAR156597 antibodies (ADA) samples are planned to be collected at approximately the same times as the PK samples, namely: Visit 2 (Day 1), Visit 4 (Week 4), Visit 6 (Week 12), Visit 7 (Week 24), Visit 8 (Week 36), Visit 9 (EOT visit at Week 52) and Visit 10 (EOS visit at Week 64).

ADA positive patients are patients with at least 1 treatment-induced or treatment-boosted ADA positive sample during the TEAE period, where

- Treatment induced ADAs: ADAs developed de novo (seroconversion) following administration of the biotherapeutic (i.e., formation of ADAs any time after the initial drug administration in a subject without pre-existing ADAs). If the baseline ADA sample is missing or non-reportable and at least one reportable ADA sample is available during the treatment (including follow-up period) the baseline sample will be considered as "negative" for data analysis. This is considered being a conservative approach for ADA assessment.
- Treatment boosted ADA: Pre-existing ADAs that were boosted to a higher level following administration of biotherapeutic (i.e., any time after the initial drug administration) the ADA titer is significantly higher than the baseline titer. A low serial dilution schema (2-fold or 3 fold) should be applied during titration. A difference in titer values of two titer

steps between an on treatment or follow-up sample and its baseline sample is considered significant. For examples, at least a 4-fold increase in titers for 2-fold serial dilution schema (or 9-fold increase in titers for 3- fold serial dilution schema). If no titer could be determined for a positive sample, the titer will be reported as the MRD of the assay.

The rest will be classified as ADA negative or inconclusive patients.

The following variables will be described:

- ADA response (Positive, Negative or inconclusive). For ADA positive:
  - Titer levels
  - Neutralizing status (Positive or inconclusive)
- Pre-existing positive ADA defined as patients with positive ADA response at baseline with less than a specific increase (depending on titer calculation, e.g. 4-fold increase in titer for 2-fold serial dilution schema, 9–fold increase in titer for 3-fold serial dilution schema) in the post-baseline period
- Treatment-emergent positive ADA response defined as 1) Patients with no ADA positive response at baseline but with any positive response in the post-baseline period (up to follow-up visit) or 2) Patients with a positive ADA response at baseline and at least a specific increase (depending on titer calculation, e.g. 4-fold increase in titer for 2-fold serial dilution schema, 9–fold increase in titer for 3-fold serial dilution schema) in titer in the post-baseline period (up to follow-up visit). For treatment-emergent positive ADA, the following categories for ADA duration will be applied:
  - A persistent positive response is a treatment-emergent ADA positive response detected in at least 2 or more post-baseline samples separated by at least a 16-week period (irrespective of any negative sample in between)
  - An indeterminate duration positive response is defined as ADA present only at the last sampling time point (and all previous samples negative) or with two last samples are positives but separated by a period less than 16 weeks.
- A transient positive response is defined as
  - Treatment induced ADA detected only at one sampling time during treatment or follow-up observation period (excluding last sampling time)
  - Treatment induced ADA detected at two or more sampling time during treatment where the first and last ADA positive sample are separated by a period less than 16 weeks and last sampling time is negative.
- Duration of ADA defined as longevity of treatment induced ADA (in days): date of last treatment induced ADA sample minus date of first treatment induced or treatment boosted ADA sample + 1. This will be calculated only for subjects with at least 2 positive ADA samples.
- Time to onset of treatment-emergent ADA positive response (in days): date of first ADA positive – date of first IMP administration.

2.1.7 Biomarker endpoints

All measurements (scheduled or unscheduled) will be assigned to efficacy analysis windows defined in Table 6, in order to describe all BM assessments by time points even in case of premature treatment discontinuation.

Peripheral blood are planned to be collected for measurement of protein biomarkers in serum or plasma at the study visits 2 (Day 1), 7 (Week 24) and 9/EOT (Week 52).

The protein biomarkers include, but are not limited to:

- Inter-cellular adhesion molecule 1 (ICAM1),
- TARC,
- Periostin,
- Human Epipidymis protein 4 (HE4).
- IL-6

2.1.8 Health economic endpoints

Health economic endpoints other than quality of life will not be assessed.

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient study status and the patient analysis populations.

Screened patients are defined as any patients who signed the informed consent.

Randomized patients consist of all patients with a signed informed consent form who have had a treatment kit number allocated and recorded in the IVRS/IWRS database, regardless of whether the treatment kit was used.

For patient study status, the total number of patients in each of the following categories will be presented in the clinical study report using a flowchart diagram or summary table by treatment arm and overall:

- Screened patients
- Screen failure patients and reasons for screen failure
- Non-randomized but treated patients
- Randomized patients
- Randomized but not treated patients and reason for not being treated
- Randomized and treated patients

- Patients who complete and did not complete the study treatment period as per protocol
- Patients who discontinued study treatment by main reason for permanent treatment discontinuation
- Patients who complete and did not complete the study follow-up period as per protocol
- Patients who discontinued study by main reason for study discontinuation
- Status at last study contact and patient ongoing at the moment of the early analysis

This table will be duplicated for background therapy status.

For all categories of patients (except for the screened and non-randomized categories) percentages will be calculated using the number of randomized patients as the denominator. Reasons for treatment discontinuation will be supplied in tables giving numbers and percentages by treatment arm. This summary will be provided by treatment arm.

The incidence of premature treatment discontinuation (irrespective of the reason) and premature treatment discontinuation due to AEs will be presented graphically by treatment arm, background therapy status on randomized and treated patients (as randomized), using Kaplan-Meier method. The time-to-event variable is defined as: event date – randomization date + 7 days, where an event could be:

- a permanent treatment discontinuation recorded on the “End of Treatment” eCRF page with a No answer to “Did the subject complete treatment period per protocol?”.
- a death before week 52 (without previous EOT) recorded on the “Death” eCRF page.

In case of a complete treatment period (with a Yes answer to “Did the subject complete treatment period per protocol?” recorded on the End of Treatment” eCRF page), the patient will be censored at the completion date.

The incidence of premature study discontinuation will be presented graphically by treatment arm, background therapy status on randomized patients, using Kaplan-Meier method. The time-to-event variable is defined as: event date – randomization date + 1 day, where an event could be:

- a study discontinuation without a 12-week follow-up.
- a study discontinuation with a 12-week follow-up but occurring before week 52.
- a death occurring before week 52 (without previous EOS) recorded on the “Death” eCRF page.

In case of a complete study period (i.e. with at least 12 weeks of follow-up ending after week 52), the patient will be censored at the completion date.

A patient is considered lost to follow-up at the end of the study if he/she is not assessed at the last protocol planned visit and if the time from the last successful contact (from the vital patient status case report form) to the last protocol planned visit is greater than 2 days, unless died before.

Patient with insufficient follow-up will be described.

A patient is considered with insufficient follow-up or without follow-up at the end of the study in the following cases:

- If the patient is not assessed at the scheduled follow-up visit, or any post-treatment visit, unless patient died before;
- If the follow-up visit is less than 12 weeks after the last double-blind IMP injection.

All critical or major quantitative deviations potentially impacting efficacy, randomization and drug-dispensing irregularities, and other important deviations will be summarized in tables giving numbers and percentages of deviations by treatment arm. These deviations are listed in the data review and surveillance plan. All critical or major qualitative deviations will be described.

Additionally, the analysis populations for safety, efficacy, pharmacokinetics, immunology and biomarkers will be summarized in a table by number of patients on the randomized population.

- Efficacy population: modified intent-to-treat (mITT) population
- Safety population
- Pharmacokinetics population
- Anti-Drug Antibody population
- Pharmacodynamics biomarker populations: Biomarker TARC population, Biomarker ICAM1 population, Biomarker HE4 population, Biomarker Periostin population, Biomarker IL-6 population.

Definitions of the study populations are provided in Section 2.3.

2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as
   a) an ineligible patient is randomized,
   b) a patient is randomized based on an incorrect stratum,
   c) a patient is randomized twice.

   OR
2. A patient is dispensed an IMP kit not allocated by the protocol-defined randomization, such as
   a) a patient at any time in the study is dispensed a different treatment kit than as randomized (which may or may not contain the correct-as-randomized IMP), or
   b) a non-randomized patient is treated with IMP reserved for randomized patients.

Randomization and drug-dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All randomization and drug-dispensing irregularities will be documented in the clinical study report. If the number of irregularities is large enough to make a tabular summary useful, the irregularities will be categorized and summarized among randomized patients (number and percentages). Non-randomized, treated patients will be described separately.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

Randomization and drug allocation irregularities
Kit dispensation without IRT transaction
Erroneous kit dispensation
Kit not available
Randomization by error
Patient randomized twice
Stratification error
Patient switched to another site

2.3 ANALYSIS POPULATIONS

Patients treated without or before being randomized will not be considered randomized and will not be included in any analysis population. The safety experience of these patients will be reported separately.

The randomized population includes any patient who has been allocated to a randomized treatment regardless of whether the treatment kit was used.

For any patient randomized more than once, only the data associated with the first randomization will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

2.3.1 Efficacy populations

2.3.1.1 Modified intent-to-treat population

The modified intent-to-treat (mITT) population is defined as all randomized patients who received at least 1 dose or part of a dose of the IMP. Patients in the mITT population will be analyzed according to the treatment arm to which they are randomized.

2.3.2 Safety population

The safety population is defined as randomized population who actually received at least 1 dose or part of a dose of the IMP, analyzed according to the treatment actually received.

In addition:

- Non-randomized but treated patients will not be part of the safety population; however, their safety data will be presented separately
- Randomized patients for whom it is unclear whether they took the IMP will be included in the safety population as randomized
- For patients receiving more than 1 IMP during the trial, the treatment arm allocation for as-treated analysis will be the one received at the majority of injections during the first 52 weeks of treatment in terms of kits.

2.3.3 Pharmacokinetics population

The PK analysis will be performed on all randomized and treated patients (safety population) with at least one available PK sample post first double-blind IMP injection.

2.3.4 Immunogenicity population

The anti-SAR156597 antibody analysis will be performed on all randomized and treated patients (safety population) with at least one available ADA sample post first double-blind IMP injection.

2.3.5 Biomarker population

Baseline biomarker population includes the randomized and treated patients with a baseline sample successfully analyzed.

Pharmacodynamics biomarker population includes the randomized and treated patients with a baseline sample and at least one post first double-blind IMP injection sample successfully analyzed.

2.4 STATISTICAL METHODS

Continuous data will be summarized using the number of available data, mean, standard deviation (SD), median, minimum and maximum for each treatment arm. Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment arm.

2.4.1 Demographics and baseline characteristics

Parameters will be summarized on the randomized population analyzed in the treatment arm to which they were randomized and according to background therapy status.

Parameters described in Section 2.1.1 will be summarized by treatment arm and overall treatment groups using descriptive statistics.

Medical/surgical history will be summarized in each treatment arm by primary SOC, HLT and PT. Events will be sorted by SOC internationally agreed order and decreasing frequency of PT based on the incidence in the overall treatment arm.

P-values on demographic and baseline characteristic data will not be calculated.

A specific quantitative description of the efficacy parameters will be provided at baseline by treatment arm and by background therapy status. In particular, Q1 and Q3 will be described for PFTs. In addition, the terciles categories at baseline will be displayed for % predicted FVC and % predicted DLCO parameters.

2.4.2 Prior or concomitant medications

Specific prior medications (Nintedanib, Pirfenidone, Systemic Corticosteroids, N-Acetylcysteine, Anti-acid (including PPI and H2 receptor blocker)) and other prior medications will be presented separately for the randomized population.

Specific concomitant medications (Nintedanib, Pirfenidone, Systemic Corticosteroids, N-Acetylcysteine, Anti-acid (including PPI and H2 receptor blocker)) and other concomitant medications will be presented separately for the randomized population.

Specific post-treatment medications (Nintedanib, Pirfenidone, Systemic Corticosteroids, N-Acetylcysteine, Anti-acid (including PPI and H2 receptor blocker)) and other concomitant medications will be presented separately for the randomized population.

Specific prior, concomitant, post-treatment medications will be summarized by treatment arm according to the WHO-DD dictionary, considering the first digit of the anatomic category (ATC) class (anatomic category) and standardized medication names.

Other prior, concomitant, post-treatment medications will be summarized by treatment arm according to the WHO-DD dictionary, considering the first digit of the anatomic category (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and patients will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore patients may be counted several times for the same medication.

The table for prior specific medications will be sorted by decreasing frequency of ATC followed by the standardized medication names based on the overall incidence across treatment groups. The table for other prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across treatment groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories or standardized medication names), alphabetical order will be used.

The tables for specific concomitant and post-treatment medications will be sorted by decreasing frequency of ATC followed by standardized medication names based on the incidence on the column all. The tables for other concomitant and post-treatment medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence on

the column all. In case of equal frequency regarding ATCs (anatomic or therapeutic categories or standardized medication names), alphabetical order will be used.

In addition, the following specific medications will be summarized on the randomized population:

- The prohibited medications will be presented in the deviations table.
- For the specific concomitant medications (Pirfenidone, Nintedanib, Systemic Corsticosteroids, NAC and Anti-acid therapy (combined anti-acid, PPI and H2 receptor blocker)), the number of patients taking at least one dose at randomization will be summarized. For Pirfenidone, Nintedanib and Systemic Corsticosteroids only, the mean, median, standard deviation (SD), minimum and maximum dose will be summarized (use of the standard prednisone equivalent dose for Systemic Corticosteroids). If several doses for a given medication were taken by patient at this timepoint, the sum of the doses will be used.
- The change in dose for Nintedanib and Pirfenidone medications will be described by visit and treatment arm according to the "Status of IPF Background Therapy" eCRF page: Has the dose of the background therapy changed since the previous visit? (Yes, No).

A list of ATC codes to be used in determining these specific medications is displayed in Appendix F.

2.4.3 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure and compliance will be assessed and summarized by actual treatment within the safety population (Section 2.3.2).

2.4.3.1 Extent of investigational medicinal product exposure

The extent of IMP exposure will be assessed by the duration of IMP exposure.

Duration of IMP exposure is defined as last dose date – first dose date + 7 days, regardless of unplanned intermittent discontinuations (see Section 2.5.3 for calculation in case of missing or incomplete data).

Duration of IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum) in weeks. In addition, duration of treatment exposure will also be summarized categorically by numbers and percentages for each of the following categories and cumulatively according to these categories: ≥1 day and <2 weeks, ≥2 weeks and <4 weeks, ≥4 weeks and <6 weeks, ≥6 weeks and <12 weeks, ≥12 weeks and <24 weeks, ≥24 weeks and <36 weeks, ≥36 weeks and <44 weeks, ≥44 weeks and <53 weeks and ≥53 weeks. Non-integer values will be rounded to 1 decimal place.

Additionally, the cumulative duration of treatment exposure will be provided, defined as the sum of the duration of treatment exposure for all patients, and will be expressed in patient years.

The graphical representation of exposure to investigational product and time since randomization will also be displayed by treatment arm.

2.4.3.2 Compliance

A given administration will be considered noncompliant if the patient did not take the planned dose of treatment as required by the protocol. No imputation will be made for patients with missing or incomplete data.

Overall percentage of compliance for a patient will be defined as 100-(% day with below-planned dosing + % days with above-planned dosing), considering that injection should be performed every week +/- 2 days as per protocol.

Above-planned dosing percentage for a patient will be defined as the number of days with more than 1 injection administration within the 5 days before divided by the duration of IMP injection exposure in days.

Below-planned dosing percentage for a patient will be defined as the number of days with no injection administration within the previous 9 days divided by the duration of IMP injection exposure in days.

These parameters will be summarized descriptively as quantitative variables (number, mean, SD, median, minimum, and maximum).

The overall percentage of patients whose compliance is ≥80% will be summarized. In addition, numbers and percentages of patients with >0 and ≤5, >5 and ≤10, >10 and ≤20 and >20% days with above-planned dose administrations will also be provided, as well as numbers and percentages of patients with >0 and ≤5, >5 and ≤10, >10 and ≤20 and >20% days with below-planned dose administrations.

In addition, the following parameters will be summarized by treatment arm:

- The number of IMP injections (mean (SD), median, Min:Max description).
- The number of advanced injections (less than 5 days between two consecutive injections) and the number of delayed injections (more than 9 days between two consecutive injections).
- The number of patients with at least one advanced injection and the number of patients with at least one delayed injection. A patient could be counted in several categories.
- The frequency of IMP injections (mean (SD), median, Min:Max description) will be defined for each patient as the average number of days between two consecutive injections: (last injection date – first injection date)/(number of injections -1) for patients with at least 2 injections.

Cases of overdose are reported in the AE e-CRF pages as AESI if symptomatic or AE if asymptomatic. The reported cases of overdose will be described in the AE analysis (see Section 2.1.4 and Section 2.4.5.1).

2.4.4 Analyses of efficacy endpoints

All visits will be re-allocated as given in Table 6, and re-allocated visits measurements will be included in the by-time point summaries unless otherwise specified.

2.4.4.1 Analysis of primary efficacy endpoint(s)

The primary efficacy endpoint will be analyzed on the mITT population. In the primary analysis, assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment.

Due to the non-normality of the endpoint distribution, the absolute change from baseline in % predicted FVC at week 52 as defined in Section 2.1.3.1 will be analyzed in the mITT population using a rank ANCOVA model adjusted on the stratification factor (with/without background therapy). This model will use the ranks of absolute change from baseline to week 52 in % predicted FVC as the outcome and the ranks of baseline % predicted FVC as a covariate. In case of ties, the mean of the corresponding ranks will be assigned.

To summarize, the following steps will be performed:

A) On the pool of arms SAR156597 200mg qw and placebo:

1. Imputation of missing data at 52 weeks

a) Week 52 missing values due to reasons other than death or lung transplant

i. Patients in active arm and taking treatment at 52 weeks: Missing values at 52 weeks will be imputed using a model estimated from the subset of patients in the active arm who completed the 52-week treatment period and with a week 52 value. The imputation model will be based on a Mixed Model with Repeated Measures (MMRM) with fixed categorical effect of stratification factor (with/without Background Therapy) and repeated categorical effect for time point (baseline, week 12, week 24, week 36, week 52). Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation. Missing values at 52 weeks will be imputed using the empirical best linear unbiased prediction (EBLUP).

ii. Other patients (patients in active arm and not taking treatment at 52 weeks; or patients in placebo arm): Missing values at 52 weeks will be imputed using a model estimated from the subset of patients in the placebo arm who completed the 52-week treatment period and with a week 52 value. The imputation model will be based on a Mixed Model with Repeated Measures (MMRM) with fixed categorical effect of stratification factor (with/without Background Therapy) and repeated categorical effect for time point (baseline, week 12, week 24, week 36, week 52). Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using

Satterthwaite's approximation. Missing values at 52 weeks will be imputed using the empirical best linear unbiased prediction (EBLUP).

iii. When running the imputation model, if the predicted absolute change from baseline calculated from the predicted week 52 value is greater than zero (increases above baseline), it will be set to zero.

iv. The non-monotone missing patterns with a week 52 value available will not be imputed.

b) Week 52 missing values due to lung transplant:

i. Any % predicted FVC value assessed after a lung transplant will be considered as missing.

ii. No imputation of a week 52 value for % predicted FVC. In step 2, patients will be ranked worse than missing data due to reasons other than death or lung transplant.

c) Week 52 missing values due to death :

i. No imputation of a week 52 value for % predicted FVC. In step 2, patients will be ranked worse than missing data due to lung transplant.

2. Rank ANCOVA

a) Ranking within each stratum of the stratification factor (with/without Background Therapy)

i. Death: Missing data due to death will be ranked based on a time-to-death, with shortest time until death as the worst (lowest) rank (from 1 to $n_{stratum,death}$).

ii. Lung transplant: Missing data due to lung transplant will be ranked based on a time-to-lung transplant, with shortest time until lung transplant as the worst (lowest) rank (from $n_{stratum,death}+1$ to $n_{stratum,death}+n_{stratum,lungtransplant}$). So, missing data due to lung transplant will less penalized than deaths.

iii. Other: Absolute change from baseline at 52 weeks in % predicted FVC values will be calculated from observed and imputed week 52 % predicted FVC values and then ranked (from $n_{stratum,death}+n_{stratum,lungtransplant}+1$ to $n_{stratum}$).

iv. Standardization of ranks using the NPLUS1 option of PROC RANK on SAS with the by statement for the stratification factor (with/without Background Therapy).

b) Adjustment on baseline % predicted FVC covariate

i. The residuals from the linear regression of response ranks on baseline ranks will be calculated within each stratum of the stratification factor (with/without Background Therapy) using the following PROC GLM on SAS:

```
proc GLM data=xx;
by stratification_factor;
model chg_Rank=base_Rank;
output out=residual r=resid;
run;
```

c) Test of treatment effect using the CMH2 option of PROC FREQ on SAS and the residuals as scores
   i. The qw arm will be compared to placebo using a Mantel-Haenszel mean score chi-square test at a two-sided type-I error level of 5%. This test will be stratified on the stratification factor.

B) On the pool of arms SAR156597 200mg q2w and placebo:
Same steps as A except for 2c: the q2w arm will be compared to placebo using a Mantel-Haenszel mean score chi-square test at a two-sided type-I error level of 5% only if the qw arm versus placebo comparison was significant. This test will be stratified on the stratification factor.

Descriptive statistics for values and changes from baseline will be presented by treatment arm and background therapy status for each time point. The mean (SD) % predicted FVC values and the mean (SD) absolute change from baseline in % predicted FVC over time will be graphically represented by treatment arm and by background therapy status without any imputation.

A specific table will describe the following items by treatment arm and background therapy:

- The number (%) of patients with an observed week 52 value for % predicted FVC
  - The mean (SD) and median (Q1:Q3) of baseline and week 52 values of % predicted FVC for those patients
  - The mean (SD) and median (Q1:Q3) of absolute change from baseline at 52 weeks in % predicted FVC for those patients
- The number (%) of patients having a missing data due to death at week 52,
- The number (%) of patients having a missing data due to lung transplant at week 52,
- The number (%) of patients having a missing data due to reasons other than death or lung transplant,
- The mean (SD) of baseline % predicted FVC value for all patients included in the analysis
- The p-values versus placebo of rank ANCOVA (only by treatment arm).

The magnitude of treatment effect will be assessed as follows:

- For each quartile of ranks per pool (A then B) of treatment arms, the following items will be displayed:
  - The number (%) of patients with death
  - The number (%) of patients with lung transplant
  - The number (%) of patients with an observed or imputed week 52 value for % predicted FVC
    - The median (Q1:Q3) absolute change from baseline at 52 weeks in % predicted FVC for those patients

This will allow to compare across treatment arms the distribution of deaths, lung transplants and absolute change from baseline at 52 weeks within each quartile (first quartile corresponding to the worst ranks). As ranks will be performed by background therapy status, one table will be provided on patients with background therapy status and another one will be provided on patients without background therapy status.

- The worsening of the change from baseline at 24 weeks and 52 weeks (imputed and observed values) will be described thanks to the number (%) of patients by treatment arm and by background therapy status in following categories:
  - Stability or improvement (decrease ≤0%)
  - Decrease >0 % to ≤5%
  - Decrease >5% to ≤10%
  - Decrease >10% or death or lung transplant

Supportive analysis

The cumulative distribution of decline from baseline at 24 weeks and at 52 weeks in % predicted FVC will be graphically represented by treatment arm and by background therapy status without any imputation.

2.4.4.1.1 Sensitivity analyses

The description of all sensitivity analyses is available on the below Table 4:

Table 4 - primary efficacy outcome algorithm in case of missing data

Type of missing data	Sensitivity 1	Sensitivity 2
Missing data at week 52 due to reasons other than death or lung transplant	Multiple imputation of week 52 value using copy increment from reference (CIR) approach (assuming normality)	No imputation
Missing data at week 52 due to death	Multiple imputation of week 52 value using the distribution of the last available % Predicted FVC assessment of dead and lung transplanted patients	No imputation
Missing data at week 52 due to lung transplant	Multiple imputation of week 52 value using the distribution of the last available % Predicted FVC assessment of dead and lung transplanted patients	No imputation

Sensitivity analysis 1: Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment.

The absolute change from baseline to week 52 in %predicted FVC will be analyzed on the mITT population using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as well as the continuous fixed covariate of % predicted FVC baseline. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation.

As for primary analysis, the non-monotone missing patterns with a week 52 value available will not be imputed.

Week 52 missing values (due to reasons other than death or lung transplant) will be imputed using the CIR multiple imputation approach (3). The CIR approach is a multiple imputation methodology based on a Gaussian repeated-measures model:

- First, a large number of datasets will be imputed using the posterior distribution of a linear model containing:
  - baseline value,
  - week 12, week 24 and week 36 time points values (if occurred before the current time point to be imputed),
  - stratification factor,
  - treatment arm* time point interaction,
  - treatment arm.

  This requires a monotone missingness, which is why all time points will be imputed sequentially to determine week 52 value.
- Then, the analysis will be run independently on those imputed datasets to estimate parameters.
- Finally, all the results will be combined thanks to Rubin's rule.
- The particularity of this approach is that it will takes into account the disease progression:
  - Before withdrawal, all patterns have the same profile. A patient will be assumed to have the same mean response as his/her treatment arm.
  - After withdrawal (which results in missing values for all subsequent visits):
    - a patient in placebo arm will be imputed as in MAR framework.
    - a patient in active arm will be imputed assuming that the advantage due to past treatment is incremented with placebo slope assessed from withdrawal visit to imputed visit.

- In practice:
  - When running the imputation model, if the absolute change from baseline calculated from the predicted week 52 value is greater than zero (increases above baseline), it will be set to zero.
  - The MMRM will be performed on each imputed dataset.
  - The Rubin's rule will be performed using Roger's macro or MIANALYZE SAS procedure as appropriate
    - The number of datasets (1000 if appropriate) will be chosen to achieve p-value stability.

Week 52 missing data due to death or lung transplant will be imputed using the distribution of their last available % Predicted FVC assessment. The mean (SD) will be calculated and then, for each dead or lung transplanted patient, a value will be drawn several times (same number as for CIR methodology) between this mean and 0 (translation of this distribution with mean as maximal value achievable and right truncated at 0 if required). If the value drawn is higher than the last available patient's value, the imputed week 52 value will be set to the last available patient's value.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 5% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 5% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE), 90% and 95% CIs will be assessed for each active arm versus placebo.

A specific table will describe the following items by treatment arm:

- The number (%) of patients with an observed week 52 value for % predicted FVC
  - The mean (SD) and median (Q1:Q3) of baseline and week 52 values of % predicted FVC for those patients
  - The mean (SD) and median (Q1:Q3) of absolute change from baseline at 52 weeks in % predicted FVC for those patients
- The number (%) of patients having a missing data due to death at week 52,
- The number (%) of patients having a missing data due to lung transplant at week 52,
- The number (%) of patients having a missing data due to reasons other than death or lung transplant,
- The mean (SD) of baseline % predicted FVC value for all patients included in the analysis,
- The LSMEANS (SE), the LSMEAN Differences versus placebo (SE, corresponding 90% and 95% CIs) and the p-values versus placebo of MMRM.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status.

Sensitivity analysis 2: Only on-treatment data will be analyzed which means that only data during the treatment period (from randomization to last IMP administration date + 7 days) will be considered. Post-treatment data will be considered missing. Only patients having a baseline and a post-baseline (during the treatment period) % predicted FVC values in at least one time point analysis window will be considered in the analysis.

The absolute change from baseline to week 52 in %predicted FVC will be analyzed on the mITT population using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as well as the continuous fixed covariate of % predicted FVC baseline. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation.

No imputation of week 52 missing data due to death, lung transplant or reasons other than death or lung transplant will be performed.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 5% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 5% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE), 90% and 95% CIs will be assessed for each active arm versus placebo.

A specific table will describe the following items by treatment arm:

- The number (%) of patients with an observed week 52 value for % predicted FVC
  - The mean (SD) and median (Q1:Q3) of baseline and week 52 values of % predicted FVC for those patients
  - The mean (SD) and median (Q1:Q3) of absolute change from baseline at 52 weeks in % predicted FVC for those patients
- The number (%) of patients having a missing data due to death at week 52,
- The number (%) of patients having a missing data due to lung transplant at week 52,
- The number (%) of patients having a missing data due to reasons other than death or lung transplant,
- The mean (SD) of baseline % predicted FVC value for all patients included in the analysis,
- The LSMEANS (SE), the LSMEAN differences versus placebo (SEs, corresponding 90% and 95% CIs) and the p-values versus placebo of MMRM.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status.

2.4.4.1.2 Subgroups analyses

The following subgroups will be explored:

- Smoking status: Smoke / Never Smoke
- Gender: Male / Female
- Age: Age ≤65 / Age >65
- Stratification factor: With / Without background therapy
- % predicted FVC categories at baseline (using terciles: (≤60, >60 and ≤75, >75))
- ANA at baseline: Positive / Negative where a Positive ANA titer is defined as titer ≥1:40

For each subgroup, the change from baseline at 52 weeks in % Predicted FVC will be analyzed using a rank ANCOVA model. For each pool of active arm vs placebo (A and B), the subgroup-by-treatment interaction will be tested as follows:

A) For Smoking status, Gender, Age subgroups: rank ANCOVA model with covariate baseline and terms of treatment, stratification factor, subgroup, treatment-by-subgroup

i. First, compute ranks of baseline and response as in the primary analysis

```
proc rank data=xxx nplus1 ties=mean out=rankdata;
by stratum;
var base ranks;
ranks base_Rank ranks_Rank;
run;
```

ii. Then, calculate the residuals as in the primary analysis

```
proc glm data=xx;
by stratum;
model ranks_Rank=base_Rank;
output out=residual r=resid;
run;
```

iii. Finally, the subgroup-by-treatment interaction will be tested using residuals with PROC GLM

```
proc glm data=residual;
class armn subgroup stratum;
model resid=stratum armn subgroup subgroup*armn;
run;
```

B) For Stratification factor subgroup: rank ANCOVA model with covariate baseline and terms of treatment, stratification factor, treatment-by-stratification factor

i. First, compute ranks of baseline and response without by stratum statement

```
proc rank data=xxx nplus1 ties=mean out=rankdata;
var base ranks;
ranks base_Rank ranks_Rank;
run;
```

ii. Then, calculate the residuals without by stratum statement

```
proc glm data=xx;
model ranks_Rank=base_Rank;
output out=residual r=resid;
run;
```

iii. Finally, the subgroup-by-treatment interaction will be tested using residuals with PROC GLM

```
proc glm data=residual;
class armn stratum;
model resid=stratum armn stratum*armn;
run;
```

C) For % predicted FVC baseline categories: rank ANCOVA model with terms of treatment, stratification factor, subgroup, treatment-by-subgroup

i. First, compute ranks of response as in the primary analysis

```
proc rank data=xxx nplus1 ties=mean out=rankdata;
by stratum;
var ranks;
ranks ranks_Rank;
run;
```

ii. Second, the subgroup-by-treatment interaction will be tested using response ranks with PROC GLM

```
proc glm data=residual;
class armn subgroup stratum;
model ranks_Rank=stratum armn subgroup subgroup*armn;
run;
```

Descriptive statistics of baseline values, week 52 values and change from baseline at 52 weeks in %predicted FVC including number (%) of patients, mean, SD, median, Q1, Q3, Min, Max by subgroup will be reported by treatment arm without any imputation.

Supportive analysis

The MMRM of Sensitivity analysis 1 will also be performed in order to provide the LSMEANS (SE) by treatment arm in each subgroup and test the subgroup-by-treatment-by-time point interaction in each pool (A and B) of active arm versus placebo. The MMRM will contain in addition the subgroup variable and the subgroup-by-treatment-by-time point interaction to be tested. When the stratification factor or the % predicted FVC baseline terciles are the subgroups to be tested, the model will not be adjusted on those variables.

In addition, a forest plot will resume the adjusted differences of means (LSMEAN differences) and corresponding 95% CIs of each active arm versus placebo by subgroup.

2.4.4.2 Analyses of secondary efficacy endpoints

Secondary efficacy endpoints are described in Section 2.1.3.2 and will be analyzed using the mITT population.

2.4.4.2.1 Disease progression endpoint

The disease progression will be analyzed on the mITT population. Assessments at week 52 (performed before Day 375) will be used regardless of whether the patient has previously discontinued treatment. Only patients having a baseline and a post-baseline values in at least one time points analysis window for both the % predicted FVC and the % predicted DLCO parameters will be considered in the analysis. No week 52 missing data will be imputed for the disease progression.

The cumulative incidence of disease progression at 52 weeks will be presented graphically by treatment arm and background therapy status, using Kaplan-Meier method. The Kaplan Meier survival estimates (number of patients at risk, number of patients censored, number of disease progression events and probability of surviving without event (95% CI)) will be summarized by time points and overall (number of patients censored, number of events, median time to event).

The censoring reasons will be described by treatment arm and background therapy status.

The time to disease progression at 52 weeks will be tested comparing SAR156597 200 mg qw dose versus placebo using a log-rank test stratified on stratification factor (with/without Background Therapy) at a two-sided type I error level of 5%; and only if this comparison is significant will the SAR156597 200 mg q2w dose be tested versus placebo using a log-rank test stratified on stratification factor (with/without Background Therapy) at a two-sided type I error level of 5%.

The event rates for each of the 3 treatment arms, the hazard ratio and its 90 and 95% confidence interval (CI) compared to placebo for each active arm versus Placebo will be estimated from a

Cox model containing the treatment and the stratification factor variables. The underlying assumption of proportional hazards for Cox model will be checked by visual inspection.

In order to assess the significance, the magnitude and the repartition within components of the treatment effect of disease progression at 52 weeks, the following items will be described by treatment arm and by background therapy status:

- Summary of the disease progression at 52 weeks: number of events, event rates, p-values (only by treatment arm), estimated hazard ratios and their 90 and 95% CIs
- Description of each component as part of the disease progression: their number of events, event rates and their estimated hazard ratios
- Description of each component as individual (all events will be considered meaning that a patient could be represented several times per component): their number of events, event rates, their estimated hazard ratios and their 90 and 95% CIs

In addition, the maximal absolute decrease in % predicted FVC and the maximal absolute decrease in %predicted DLCO will be described by categories (≥0 and <5%, ≥5% and <10%, ≥10% and <15%, ≥15% and <20%, ≥20%) at 52 weeks.Decomposition of events and maximal events will be explored at 24 weeks as well.

Subgroup analyses

The following subgroups will be explored:

- Smoking status: Smoke / Never Smoke
- Gender: Male / Female
- Age: Age ≤65 / Age >65
- Stratification factor: With / Without background therapy
- % predicted FVC categories at baseline (using terciles (≤60, >60 and ≤75, >75))
- ANA at baseline: Positive / Negative where a Positive ANA titer is defined as titer ≥1:40

For each pool of active arm vs placebo (A and B), the subgroup-by-treatment interaction will be tested using a Cox model containing in addition the subgroup variable and the subgroup-by-treatment interaction to be tested. When the stratification factor is the subgroup to be tested, the model should not be stratified by this stratification factor.

Descriptive statistics including number (%) of patients, number of events, event rates, HR, 95% CI by subgroup will be reported by treatment arm.

In addition, a forest plot will resume the HR and corresponding 95% CIs of each active arm versus placebo by subgroup.

2.4.4.2.2 All-cause mortality endpoint

The all-cause mortality will be analyzed on the mITT population. Deaths from baseline to week 52 (performed before Day 375) will be used regardless of whether the patient has previously discontinued treatment.

The cumulative incidence of all-cause mortality will be presented graphically by treatment arm and background therapy status, using Kaplan-Meier method. The Kaplan Meier survival estimates (number of patients at risk, number of patients censored, number of death events and probability of surviving without event (95% CI)) will be summarized by time points and overall (number of patients censored, number of events, median time to event).

The censoring reasons will be described by treatment arm and background therapy status.

The event rates for each of the 3 treatment arms, the hazard ratio and its 90% and 95% CIs compared to placebo for each active arm versus Placebo will be estimated from a Cox model containing the treatment and the stratification factor variables. The underlying assumption of proportional hazards for Cox model will be checked by visual inspection.

No hypothesis testing will be performed as per protocol.

The same time-to-event analysis with IPF-related deaths only will be explored on the mITT population.

2.4.4.2.3 Exploratory endpoints

2.4.4.2.3.1 Acute IPF exacerbations endpoint

The number of acute IPF exacerbations and the number of patients having one, two or more than two acute IPF exacerbations will be described in the mITT population by treatment arm and background therapy status.

The annual rate in patient-year by treatment arm, the rate ratio and its 90% CI for each active arm versus placebo will be determined using a negative binomial model adjusted on the stratification factor variable and containing the stratification factor-by-treatment interaction. The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10%.

In addition, a description of the following items will be performed by treatment arm and by background therapy status for events occurring between randomization and Day 375:

- the number (%) of patients with any IPF event (complementary form filled whatever the final diagnosis question answer),
- the number (%) of patients with acute IPF exacerbation (final diagnosis question is equal to “yes”),

- the other diagnoses (final diagnosis question is equal to “no”, with associated PT term and verbatim),
- the clinical evaluation,
- the radiology assessment
- and any corrective treatment/therapy used.

2.4.4.2.3.2 Respiratory and non-elective hospitalizations endpoints

The numbers of respiratory and non-elective hospitalizations and the numbers of patients having at least one respiratory or at least one non-elective hospitalization will be described in the mITT population by treatment arm and background therapy status separately.

The annual rates for respiratory and non-elective hospitalizations in patient-year by treatment arm, the rate ratio and its 90% CI for each active arm versus placebo will be determined separately using a negative binomial model adjusted on the stratification factor variable and containing the stratification factor-by-treatment interaction. The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10%.

The number (%) of patients with any hospitalization, the type of hospitalization (elective, non-elective) and the reason (respiratory, other reason) will be provided by treatment arm and background therapy status for hospitalization occurring between randomization and Day 375.

Furthermore, a listing with details of reasons (respiratory and other reason) will be displayed.

2.4.4.2.3.3 Change from baseline in 6-MWT distance endpoint

The absolute change from baseline in 6-MWT distance endpoint will be analyzed on the mITT population. Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment. No imputation will be performed for the week 52 missing values of 6-MWT distance. Only patients having a baseline and a post-baseline 6-MWT distance values in at least one time points analysis window will be considered in the analysis.

The absolute change from baseline in 6-MWT distance at 52 weeks will be analysed using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, supplemental oxygen use (Yes/No answer to the “Does the subject need oxygen therapy” question on the “Idiopathic Pulmonary Fibrosis History“ e-CRF page) recorded at baseline, as well as the continuous fixed covariate of 6-MWT distance at baseline will be used. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite’s approximation.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE) and 90% CIs will be assessed for each active arm versus placebo.

Descriptive statistics for values and changes from baseline in 6-MWT distance in the mITT population will be presented by treatment arm and background therapy status for each time point. The mean (SD) 6-MWT distance values and the mean (SD) absolute change from baseline in 6MWT distance over time will be graphically represented by treatment arm and by background therapy status.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status.

In addition, to assess the quality of data:

- the number (%) of patients having a completed 6 minute walking test,
- the distribution of the time walked (in seconds) for those who did not complete 6-MWT,
- the number (%) of patients with any pause or break
- the distribution of the duration of pause (in seconds)

will be resumed by time point.

2.4.4.2.3.4 Change from baseline in HRCT lung interstitial score endpoint

The absolute change from baseline in HRCT lung interstitial score endpoint will be analyzed on the mITT population. Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment. No imputation will be performed for the week 52 missing values of HRCT lung interstitial scores. Only patients having both the baseline and week 52 Total Fibrosing ILD score assessments available will be considered in the analysis.

The change from baseline at 52 weeks in Total Fibrosing ILD score will be analyzed using an ANCOVA model with fixed effect for treatment arm, adjusted on baseline Total Fibrosing ILD score, Emphysema total score at baseline, and the stratification factor (with/without Background Therapy) covariates, with a stratification factor-by-treatment interaction will be performed. The SAR156597 qw arm will be tested versus Placebo at a two-sided type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10% using the appropriate contrast. The adjusted Least Square Means (LSMeans) and their 90% CIs will be assessed for each treatment arm.

Descriptive statistics for values and changes from baseline in Total Fibrosing ILD score will be described over time by treatment arm and by background therapy status. The mean (SD) Total Fibrosing ILD score values and the mean (SD) absolute change from baseline in Total Fibrosing

ILD score over time will be graphically represented by treatment arm and by background therapy status.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status for the Total Fibrosing ILD score.

In addition, descriptive statistics for values and changes from baseline in pattern score and the Emphysema total score will be described over time by treatment arm and by background therapy status. The mean (SD) scores values and the mean (SD) absolute change from baseline in scores over time will be graphically represented by treatment arm and by background therapy status.

2.4.4.2.3.5 Change from baseline in total score of St George Respiratory Questionnaire (SGRQ) endpoint

The absolute change from baseline in SGRQ total score endpoint will be analyzed on the mITT population. Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment. No imputation will be performed for the week 52 missing values of SGRQ total score. Only patients having a baseline and a post-baseline total SGRQ score values in at least one time points analysis window will be considered in the analysis.

The absolute change from baseline in SGRQ total score at 52 weeks will be analysed using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as well as the continuous fixed covariate of SGRQ total score at baseline will be used. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE) and 90% CIs will be assessed for each active arm versus placebo.

For each component score (Symptoms, Activity, Impacts) and the SGRQ total score, descriptive statistics for values and changes from baseline will be described over time by treatment arm and by background therapy status. The mean (SD) of each score values and the mean (SD) absolute change from baseline in each score over time will be graphically represented by treatment arm and by background therapy status.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status for the SGRQ total score.

2.4.4.2.3.6 Change from baseline in total score of EuroQol Questionnaire, 5 level system (EQ-5D-5L) endpoint

The absolute change from baseline in EQ-5D-5L total score endpoint will be analyzed on the mITT population. Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment. No imputation will be performed for the week 52 missing values of EQ-5D-5L total score. Only patients having a baseline and a post-baseline EQ-5D-5L total score values in at least one time points analysis window will be considered in the analysis.

The absolute change from baseline in EQ-5D-5L total score at 52 weeks will be analysed using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as well as the continuous fixed covariate of EQ-5D-5L total score at baseline will be used. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE) and 90% CIs will be assessed for each active arm versus placebo.

For EQ-5D-5L VAS and the EQ-5D-5L total score, descriptive statistics for values and changes from baseline will be described over time by treatment arm and by background therapy status. The mean (SD) of each score values and the mean (SD) absolute change from baseline in each score over time will be graphically represented by treatment arm and by background therapy status.

The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status for the EQ-5D-5L total score.

2.4.4.3 Multiplicity issues

The overall type I error rate for this study is 5%. For the primary endpoint, a hierarchical testing procedure will be used to control the type I error over the two doses. The SAR156597 qw arm will be compared to placebo first, and only if this comparison is significant will the SAR156597 q2w arm be compared to placebo.

The same procedure will be used for the disease progression endpoint if significance is achieved for both doses on the primary endpoint.

For other secondary endpoints considered exploratory, a hierarchical testing procedure will be used to control the type I error over the two doses. The SAR156597 qw arm will be compared to

placebo first at a two-sided type I error level of 10%, and only if this comparison is significant will the SAR156597 q2w arm be compared to placebo at a two-sided error level of 10%.

2.4.4.4 Additional efficacy analysis

2.4.4.4.1 Absolute change from baseline in % predicted DLCO

The absolute change from baseline in % predicted DLCO endpoint will be analyzed on the mITT population. Assessments at week 52 will be used regardless of whether the patient has previously discontinued treatment. No imputation will be performed for the week 52 missing values of %predicted DLCO. Only patients having a baseline and a post-baseline % predicted DLCO values in at least one time points analysis window will be considered in the analysis.

The absolute change from baseline in % predicted DLCO at 52 weeks will be analyzed using a Mixed Model with Repeated Measures (MMRM) with the fixed categorical effects of treatment arm, stratification factor (with/without Background Therapy), time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as well as the continuous fixed covariate of % predicted DLCO baseline. Parameters will be estimated with the Newton-Raphson algorithm, an unstructured covariance matrix will model the within-patient errors, and the denominator degrees of freedom will be estimated using Satterthwaite's approximation.

The SAR156597 qw arm will be tested versus Placebo at a type I error level of 10% first and if significant, will the SAR156597 q2w arm be tested versus Placebo at a two-sided type I error level of 10% using the appropriate contrast. The adjusted Least Square Means (LSMeans) estimates and standard errors (SE) will be provided for each of the 3 arms, adjusted differences of means (LSMEAN differences), corresponding standard errors (SE) and 90% CIs will be assessed for each active arm versus placebo.

Descriptive statistics for values and changes from baseline will be presented by treatment arm and background therapy status for each time point. The mean (SD) % predicted DLCO values and the mean (SD) absolute change from baseline in % predicted DLCO over time will be graphically represented by treatment arm and by background therapy status.
The representation of LSmeans (SE) over time will be displayed by treatment arm and background therapy status.

2.4.4.4.2 Absolute change from baseline in FEV1 (% Predicted and L), FEV1/FVC, FCV (% Predicted and L),

Descriptive statistics for values and changes from baseline in FEV1 (% Predicted and L), FEV1/FVC, FCV (% Predicted and L) will be presented by treatment arm and background therapy status for each time point.

2.4.5 Analyses of safety data

The summary of safety results will be presented by treatment arm on the basis of the safety population.

General common rules

All safety analyses will be performed on the safety population as defined in Section 2.3.2, unless otherwise specified, using the following common rules:

- Safety data in patients who do not belong to the safety population (eg, exposed but not randomized) will be listed separately. The baseline value is defined in Section 2.1.4.
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG (PCSA version dated March 2014 [Appendix A]).
- PCSA criteria will determine which patients had at least 1 PCSA during the TEAE period, taking into account all evaluations performed during the TEAE period, including unscheduled or repeated evaluations. The number of all such patients will be the numerator for the on-treatment PCSA percentage.
- The treatment-emergent PCSA denominator by arm for a given parameter will be based on the number of patients assessed for that given parameter in the TEAE period by treatment arm on the safety population.
- For quantitative safety parameters based on central laboratory/reading measurements, descriptive statistics will be used to summarize results and change from baseline values by time point and treatment arm. Summaries will include the endpoint baseline value, the last on-treatment value and change from baseline, the worst on-treatment value and change from baseline. The last on-treatment value is commonly defined as the last value collected during the treatment period (see Section 2.1.4). The worst on-treatment value is defined as the nadir and /or the peak post-baseline during the treatment period according to the direction (minimum or maximum) of the abnormality as defined in the PCSA list.
- All measurements (scheduled or unscheduled) will be re-allocated as given in Section 2.5.4 and re-allocated visits measurements will be included in the by-time point summary.
- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned. Relative risks versus placebo and their 95% CIs may be provided, if relevant.

2.4.5.1 Analyses of adverse events

Generalities

The primary focus of AE reporting will be on TEAE as defined in Section 2.3.2. Pre-treatment and post-treatment adverse events will be described separately.

If an AE date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the AE as pre-treatment, treatment-emergent, or post-treatment. The algorithm for imputing date/time of onset will be conservative and will classify an AE as treatment emergent unless there is definitive information to determine it is pre-

treatment or post-treatment. Details on classification of AEs with missing or partial onset dates are provided in Section 2.1.4.1.

Adverse event incidence tables will present by SOC, HLGT, HLT, and PT, for each treatment arm, the number (n) and percentage (%) of patients experiencing an AE sorting based on results of column all. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a treatment period. The denominator for computation of percentages is the safety population within each treatment arm.

Sorting within tables ensures the same presentation for the set of all AEs within the observation period (pre-treatment, treatment-emergent, and post-treatment). For that purpose, the table of all TEAEs presented by SOC and PT sorted by the internationally agreed SOC order and decreasing frequency of PTs within SOCs will define the presentation order for all other tables unless otherwise specified. Sorting will be based on results of column all.

Analysis of all treatment-emergent adverse events

The following TEAE summaries will be generated for the safety population.

- Overview of TEAEs by treatment arm and by background therapy status, summarizing number (%) of patients with any
  - Treatment-emergent adverse event
  - Serious treatment-emergent adverse event
  - Treatment-emergent adverse event leading to death
  - Treatment-emergent adverse event leading to permanent treatment discontinuation
- All TEAES by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least 1 TEAE sorted by the SOC internationally agreed order. The other levels (HLGT, HLT, and PT) will be presented in alphabetical order. This table will be presented by treatment arm.
- All TEAEs by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This sorting order will be applied to all other tables, unless otherwise specified. This table will be presented by treatment arm and by background therapy status.
- All TEAEs regardless of relationship and related by primary SOC, HLGT, HLT and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, and PT) will be presented in alphabetical order. This table will be presented by treatment arm and by background therapy status.
- All TEAEs by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This table will be presented by treatment arm and by background therapy status.

- All TEAEs by maximal severity, presented by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE by severity (ie, mild, moderate, or severe), sorted by the sorting order defined above. This table will be presented by treatment arm.

Analysis of all treatment emergent serious adverse event(s)

- All treatment-emergent serious adverse events by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 serious TEAE, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, and PT) will be presented in alphabetical order. This table will be presented by treatment arm.
- All treatment-emergent serious adverse events by primary SOC and PT, showing the number (%) of patients with at least 1 serious TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This table will be presented by treatment arm and by background therapy status.
- All treatment-emergent serious adverse events regardless of relationship and related to IMP, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients with at least 1 treatment-emergent serious adverse event, sorted by the internationally agreed SOC order. The other levels (HLGT, HLT, and PT) will be presented in alphabetical order. This table will be presented by treatment arm.
- All treatment-emergent serious adverse events regardless of relationship and related to IMP, by primary SOC and PT, showing the number (%) of patients with at least 1 serious TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This table will be presented by treatment arm and by background therapy status.
- Listings will be provided for all SAEs by treatment arm during on-treatment period.

Analysis of all treatment-emergent adverse event(s) leading to permanent treatment discontinuation

- All TEAEs leading to treatment discontinuation, by primary SOC, HLGT, HLT, and PT, showing the number (%) of patients sorted by the internationally agreed SOC order (column “All”). The other levels (HLGT, HLT, and PT) will be presented in alphabetical order. Listings will be provided for all TEAEs leading to permanent treatment discontinuation by treatment arm and patient. This table will be presented by treatment arm.
- All TEAEs leading to treatment discontinuation, by primary SOC and PT, showing the number (%) of patients with at least 1 TEAE, sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This table will be presented by treatment arm and by background therapy status.

Analysis of treatment-emergent AESIs

All treatment emergent AESIs, by AESI category and PT, showing the number (%) of patients by treatment arm and overall, sorted by decreasing incidence of PT within each AESI category. This

table will be presented by background therapy. The AESIs categories and details of the MedDRA coding are provided in Section 2.1.4.1.

- Listing of all Treatment emergent AESIs experienced per patient and by treatment arm: Patient ID, AESI category, PT, date of onset (day of onset), date of recovery (day of end), duration, intensity, corrective treatment, seriousness, relationship to SAR, relationship to BTs.
- Kaplan-Meier cumulative incidence curve for time to first treatment emergent local injection site reaction. The time to first treatment emergent local injection site reaction will be derived as follows: date of randomization minus date of last IMP injection. Patients are censored at cut-off date or date of last injection in case of study drug discontinuation.
- Kaplan-Meir cumulative incidence curve for number of injections to first treatment emergent local injection site reaction. The number of injections to first treatment emergent local injection site reaction will be derived as follows: sum of injections from baseline to event. Patients are censored at cut-off date or date of last injection in case of study drug discontinuation.
- Number (%) of patients with at least one TEAE in local injection site reaction category by primary SOC/PT.

In addition, the following variables will be tabulated for the local injection site reactions TEAEs:

- Number (%) of patients with at least one local injection site reaction TEAE
- Local injection site reaction per patient (1, >1)
- Highest intensity of the event (Mild, Moderate, Severe, Very Severe)
- Mean duration (in days)
- Number of events divided by the number of double-blind IMP injections received
- Time from first double-blind IMP injection to first injection site reaction (in days)
- Time from most recent IMP injection before first injection site reaction (in days)
- Description of the highest intensity (and size) of each symptom (using pre-specified eCRF-terms: Erythema, Swelling, Itch, Pain, Tenderness; or using MedDRA coding if Other) from “Reaction at Injection Site” eCRF complementary form. Number will correspond to the count of patients with treatment emergent local injection site reaction and the corresponding symptom. More than one symptom may be reported for each injection site reaction. In case of several occurrences of a symptom, the maximal intensity is used
- Number of double-blind IMP injections received up to the first event (1, 2, 3, 4, ≥5 to <10, ≥10 to <20, ≥20)

In addition, the following variables will be tabulated for the ALT increase TEAEs:

- Number (%) of patients with at least one ALT increase TEAE
- Number (%) of patients who met the following symptoms:

- Asthenia
- Nausea
- Fever
- Pruritus
- Jaundice
- Joint pain
- Abdominal pain
- Vomiting
- Skin eruption
- Purpura
- Hepatomegaly
- Splenomegaly
- Lymph nodes enlarged
- Ascites
- Asterixis
- Confusion

In addition, the following variables will be tabulated for the ALT increase – Trigger Factors TEAEs:

- Number (%) of patients with at least one ALT increase – Trigger factors TEAE
- Number (%) of patients with:
  - Any blood transfusion within the last 12 months
  - Exposure to potential hepatotoxic environmental or occupational agent within the last 12 months
    - Duration of exposure
  - Alcohol consumption within the last 7 days prior to increase
    - Number of standard drinks
  - Traveled in country at risk of viral hepatitis within the last 6 months
  - Used addictive drug within the last 12 months
    - Details of addictive drug
    - IV administration intake
    - Time to first administration (in weeks)

In addition, the following variables will be tabulated for the other event TEAEs:

- [redacted]
- Number (%) of patients with at least one tuberculosis event TEAE
- Number (%) of patients with at least one anaphylaxis event TEAE

Analysis of pre-treatment and post-treatment adverse events

An overview presenting number of patients with any pre-/post-treatment adverse events, serious pre-/post-treatment adverse events, pre-/post-treatment adverse events leading to death, pretreatment adverse events leading to permanent treatment discontinuation will be displayed by treatment arm. If only few patients (<10% of randomized patients) are concerned, a global listing will be provided.

2.4.5.2 Deaths

The following summaries of deaths will be generated for the safety population by treatment arm.

- Number (%) of patients who died by study period (on-study, on TEAE period, post-study) and reasons for death summarized on the safety population by treatment received and by background therapy.
- Deaths in nonrandomized patients or randomized but not treated patients
- Treatment-emergent adverse events leading to death (death as an outcome on the AE case report form page as reported by the Investigator) by primary SOC, HLGT, HLT, and PT showing number (%) of patients sorted by internationally agreed SOC order, with HLGT, HLT, and PT presented in alphabetical order within each SOC. This table will be presented by treatment received and by background therapy.
- Treatment-emergent adverse events leading to death (death as an outcome on the AE case report form page as reported by the Investigator) by primary SOC and PT showing number (%) of patients with at least 1 TEAE leading to death, sorted by internationally agreed SOC order and by decreasing incidence of PTs within each SOC. This table will be presented by treatment received and by background therapy.
- All pre-treatment adverse events leading to death by primary SOC and PT, showing the number (%) of patients, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC
- All post-treatment adverse events leading to death by primary SOC and PT, showing the number (%) of patients, sorted by the internationally agreed SOC order and decreasing incidence of PTs within each SOC
- Listings of deaths will be provided. For deaths in non-randomized patients or randomized and not treated patients, a distinct listing will be provided.

2.4.5.3 Analyses of laboratory variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all laboratory variables (central laboratory values and changes from baseline) will be calculated for each time point or study assessment (baseline, each post-baseline time point, endpoint) by treatment arm. This section will be organized by biological function as specified in Section 2.1.4.3 (except for pregnancy test).

For parameters glucose, monocytes, basophils, eosinophils and bicarbonate, the last on-treatment and worst on-treatment value will also be provided.

The pH parameter will be described by category: ≤4.6, >4.6 and <8, ≥8.

[redacted]

A specific description of the following parameters will be provided at baseline by treatment arm and overall treatment groups:

- Tuberculosis information using MedDRA CMQ and the specific "Tuberculosis Assessment" eCRF page (signs and symptoms at screening (Yes, No), treated for tuberculosis in the past (Yes, No), receiving Active Anti-TB Therapy at screening (Yes, No)) and Quantiferon®-TB Gold test result (Positive, Negative, Indeterminate).
- Serology tests: hepatitis B surface antigen (Positive, Negative, Borderline), hepatitis B surface antibody (Positive, Negative, Borderline), hepatitis B core antibody (Positive, Negative, Borderline) and hepatitis C antibodies (Reactive, Unreactive, Borderline).
- [redacted] connective tissue disorders using MedDRA CMQs.

The incidence of PCSAs (list provided in Appendix A) at any time during the TEAE period will be summarized by biological function and treatment arm whatever the baseline level and according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

For parameters for which no PCSA criteria are defined, similar table(s) using the normal range will be provided.

Listings will be provided with flags indicating the out of range values as well as PCSA values.

Drug-induced liver injury

The liver function tests, namely AST, ALT, alkaline phosphatase, and total bilirubin, are used to assess possible drug-induced liver toxicity. The proportion of patients with PCSA values at any post-baseline visit by baseline status will be displayed by treatment arm for each parameter. The

proportion of patients with PCSA values at any post-baseline visit will also be displayed by duration of exposure for each treatment arm.

A graph of distribution of peak values of ALT versus peak values of total bilirubin will also be presented. Note that the ALT and total bilirubin values are presented on a logarithmic scale. The graph will be divided into 4 quadrants with a vertical line corresponding to 3 x ULN for ALT and a horizontal line corresponding to 2 x ULN for total bilirubin.

Summarize the normalization by parameter (to ≤1 x ULN or return to baseline) of elevated liver function tests by categories of elevation (1-<3 x, 3 x, 5 x, 10 x, 20 x ULN for ALT and AST, 1.5 x ULN for alkaline phosphatase, and 1.5 x and 2 x ULN for total bilirubin), with the following categories of normalization: never normalized, normalized after IMP discontinuation. Note that a patient will be counted only under the maximum elevation category: 1-3, 3-5, 5-10, 10-20, >20.

Summarize the incidence of liver-related AEs by treatment arm. The selection of preferred terms will be based on the drug related hepatic disorder SMQ.

2.4.5.4 Analyses of vital sign variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all vital signs variables (central laboratory values and changes from baseline) will be calculated for each time point or study assessment (baseline, each post-baseline time point, endpoint) by treatment arm. For parameters heart rate, systolic and diastolic blood pressures, the last on-treatment and worst on-treatment value will also be provided.

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment arm irrespective of the baseline level and according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

Listings will be provided with flags indicating the out of range values as well as PCSA values.

2.4.5.5 Analyses of electrocardiogram variables

The summary statistics (including number, mean, median, standard deviation, minimum and maximum) of all ECG variables (central laboratory values and changes from baseline) will be calculated for each time point or study assessment (baseline, each post-baseline time point, endpoints) by treatment arm.

The incidence of PCSAs at any time during the TEAE period will be summarized by treatment arm irrespective of the baseline level and according to the following baseline status categories:

- Normal/missing
- Abnormal according to PCSA criterion or criteria

Listings will be provided with flags indicating the out of range values as well as PCSA values.

2.4.6 Analyses of pharmacokinetic variables

C_{trough} Concentrations of SAR156597 in plasma will be described on the PK population for each time point by treatment arm and by background therapy status using descriptive statistics (number of patients, arithmetic mean (SD), geometric mean, coefficient of variation, minimum and maximum).

In a separate table, $C_{trough,av,ss}$, C_{max} and $C_{Follow\text{-}up}$ will be described on the PK population by treatment arm and background therapy status using descriptive statistics (number of patients, arithmetic mean (SD), geometric mean, coefficient of variation, minimum and maximum).

The representation of time profiles for C_{trough} concentration (mean ± SD or Median, as appropriate) will be also provided by treatment arm and by background therapy status. Average steady state achievement will be performed by visual inspection of mean plots.

According to study results, additional plots will be prepared, as deemed necessary (eg, to explore the relationship with some safety or efficacy endpoints of interest).

The results of population PK modeling will be reported separately from the study report.

2.4.7 Analyses of immunogenicity data

The following summaries will be performed on the ADA population, taking into account all samples regardless of timing in relation to injections:

- ADA results (negative, positive or inconclusive) by time point and by treatment arm
- Neutralizing status (inconclusive or positive) for positive ADA by time point and by treatment arm
- ADA titers using descriptive statistics (median, Q1, Q3, minimum and maximum) for positive ADA by time point and by treatment arm
- Number (%) of patients with pre-existing ADA and number (%) of patients with treatment-emergent ADA positive response by treatment arm
- Number (%) of patients with persistent/transient/indeterminate treatment-emergent ADA positive response by treatment arm
- Time to onset of treatment-emergent ADA positive response (in days) using descriptive statistics by treatment arm and also according to peack titer.
- Duration of ADA (in days, only for subjects with at least 2 positive ADA samples) using descriptive statistics by treatment arm.

Correlations between ADA parameters (eg, titers, treatment-emergent ADA positive status, neutralizing status) and PK, safety and/or main efficacy endpoint(s) will be explored (eg, scatter plot) according to the study results.

2.4.8 Analyses of Biomarker data

Biomarkers will be summarized on the Pharmacodynamics population using descriptive statistics by treatment arm and time point.

Time profile of biomarkers will be explored and link between biomarkers and treatment arm will be investigated.

Correlation between each biomarker and the main efficacy endpoint(s) will be assessed taking into account the dose and the background therapy status.

According to study results, correlation between each biomarker at baseline and $C_{trough,av,ss}$ will be explored.

In addition, transcriptomic (mRNA) and genomic (DNA) properties will be explored.

A detailed description of the analyses will be provided apart of this SAP. The results of biomarker, transcriptomic and genomic analyses will be reported separately from the study report.

2.5 DATA HANDLING CONVENTIONS

2.5.1 General conventions

The following formulas will be used for computation of parameters.

Study day

Study day = assessment date - first IMP injection date + 1 (if assessment date ≥first IMP injection date), otherwise study day = assessment date - first IMP injection date, the day of first IMP injection being Day 1. For randomized but not treated patients, Day 1 is the day of randomization.

Demographic formulas

- Age (years) = (date of informed consent – date of birth)/365.25

Disease characteristics formulas

- Time since diagnosis of IPF (years) = (date of randomization - date of first diagnosis)/365.25
- Time since first symptoms of IPF (years) = (date of randomization - date of first IPF symptoms)/365.25
- Time from most recent date of acute exacerbation (months) = (date of randomization – date of most recent acute exacerbation)/30.4375
- Mean smoking duration:
- for quit smokers: smoking duration = end year – start year. In case of missing end year, no imputation with analysis year.
- for current smokers: smoking duration = analysis year - start year

ECG formulas

Bazett (QTcB (ms)): $\mathrm{QT\,(ms)} \Big/ \sqrt{60 \Big/ HR(bpm)}$

Fridericia (QTcF (ms)): $\mathrm{QT\,(ms)} \Big/ \sqrt[3]{60 \Big/ HR(bpm)}$

Renal function formulas

Creatinine clearance value will be derived using the equation of Cockroft and Gault:

For Male:

$\mathrm{CL_{CR}\ (mL/min)} = (140 - age(years)) * weight(kg) \Big/ 0.814 * serumcreatinine(\mu mol / L)$

For Female: result above multiplied by 0.85.

Date of last dose of investigational medicinal product

The date of the last dose of IMP is equal to the last date of administration reported on the IMP administration case report form page, or missing if the last administration date is unknown.

Lipids variables, laboratory safety variables, hs-CRP

For data below the lower limit of quantification (LLOQ)/limit of linearity, half of the lower limit value (i.e; LLOQ/2) will be used for quantitative analyses. For data above the upper limit of quantification (ULOQ) / limit of linearity, the upper limit value (i.e., ULOQ) will be used for quantitative analyses.

For ANCA parameter: values '<1:20' are considered as negative; other results are considered as positive.

The above rules won't be applied for the calculated LDL-C and non-HDL-C when HDL-C value is below the LLOQ. The value of LLOQ/2 for HDL-C will be used to obtain the non-HDL-C and calculated LDL-C used for quantitative analyses.

ADA

Missing baseline ADA will be considered as negative

2.5.2 Data handling conventions for secondary efficacy variables

2.5.2.1 St George Respiratory Questionnaire (SGRQ) endpoint

Table 5 - Data handling convention for the SGRQ variable from scales

Secondary Efficacy variable	Short Description	Data Collection	Derivation rules/Missing data handling
SGQRT: "Symptoms component" score	100 x (Sum of weights for the positive responses to questions 1 to 8) / Sum of weights for all items in Symptoms component	Self-administration questionnaire on site, entered in the database	In case of more than 2 missing items, the sum will not be calculated. The weight for the missed item is subtracted from the total possible weight for the Symptoms component (662.5) and from the Total weight (3989.4)
SGQRT: "Activity component" score	100 x ('Sum of weights for the positive responses to questions 11 and 15) / Sum of weights for all items in Activity component	Self-administration questionnaire on site, entered in the database	In case of more than 4 missing items, the sum will not be calculated. The weight for the missed item is subtracted from the total possible weight for the Symptoms component (1209.1) and from the Total weight (3989.4).
SGQRT: "Impacts component" score	100 x (Sum of weights for the positive responses to questions 9 to 10, 12 to 14 and 16 to 17)/ Sum of weights for all items in Impacts component	Self-administration questionnaire on site, entered in the database	In case of more than 6 missing items, the sum will not be calculated. The weight for the missed item is subtracted from the total possible weight for the Symptoms component (2117.8) and from the Total weight (3989.4).
SGQRT total score	100 x (Sum of weights for all positive responses in the questionnaire)/ Sum of total weights for the questionnaire	Self-administration questionnaire on site, entered in the database	If more than 24% of items are missing in the questionnaire, the total score will not be calculated.

It will be noted that the questionnaire requests a single response to questions 1-7, 9-10 and 17. If multiple responses are given to one of these questions then averaging the weights for the positive responses for that question are acceptable. We feel that this is a better approach than losing an entire data set and have used this technique in calculating the results used in our validation studies. (Clearly a better approach is to prevent such multiple responses occurring, but it is difficult to prevent occasional accidents).

2.5.3 Missing data

For categorical variables, patients with missing data are not included in calculations of percentages unless otherwise specified. When relevant, the number of patients with missing data is presented.

Handling of computation of time since diagnosis of IPF if the date of first diagnosis is incomplete

If the day is missing, the date is imputed to first day of the month. If the month is missing, the date is imputed to first month of the year

Handling of computation of time since first symptoms of IPF if the date of first symptoms of IPF is incomplete

If the day is missing, the date is imputed to first day of the month. If the month is missing, the date is imputed to first month of the year

Handling of computation of treatment duration if investigational medicinal product end of treatment date is missing

For the calculation of the treatment duration, the date of the last dose of IMP is equal to the date of last administration reported on the Investigational Product Administration case report form page.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of medication missing/partial dates

No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior or concomitantly, it will be considered a prior, concomitant, and post-treatment medication.

Handling of adverse events with missing or partial date/time of onset

Missing or partial AE onset dates and times will be imputed so that if the partial AE onset date/time information does not indicate that the AE started prior to treatment or after the TEAE period, the AE will be classified as treatment-emergent. No imputation of AE end dates/times will be performed. These data imputations are for categorization purpose only and will not be used in listings. No imputation is planned for date/time of AE resolution.

Handling of adverse events when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all AEs that occurred on or after the day of randomization should be considered as TEAE. The exposure duration should be kept as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of missing assessment of relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP has to be assumed and the AE considered as such in the frequency tables of possibly related AE, but no imputation should be done at the data level.

Handling of missing severity of adverse events

If the severity is missing for 1 of the treatment-emergent occurrences of an AE, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, a "missing" category will be added in the summary table.

Handling of potentially clinically significant abnormalities

If a patient has a missing baseline he will be grouped in the category "normal/missing at baseline."

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is >0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

2.5.4 Windows for time points

Data analyzed by time point (including efficacy, PK, BM, ADA, safety) will be summarized using the analysis windows given in Table 6, Table 7 and Table 8. The analysis windows will be exhaustive so that data recorded at any time point (including unscheduled visits) has the potential to be summarized. These analysis windows will be applicable for all analyses (except for disease progression, all-cause mortality, IPF-related cause mortality, acute IPF exacerbations, respiratory hospitalizations, non-elective hospitalizations), and they are defined to provide more homogeneous data for time point-specific analyses. For efficacy, PK, BM, ADA, if multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the maximum value will be displayed. For safety, if multiple valid values of a variable exist within an analysis window, the nearest from the targeted study day will be selected. In case of ties, the mean will be displayed.

Table 6 - Analysis windows definition for efficacy

Time point	Targeted study day	Efficacy analysis window (in study days)
Week 12	85	2 to 126
Week 24	169	127 to 210
Week 36	253	211 to 308
Week 52	365	309 to 406
Week 64	449	>406

Note: Safety and BM will follow the same analysis windows.

Table 7 - Analysis windows definition for PK and ADA

Time point	Targeted study day	Analysis window (in study days)
Week 4	29	27 to 31
Week 12	85	83 to 87
Week 24	169	167 to 171
Week 36	253	251 to 255
Week 52	365	363 to 367
Optional: Week 52+ 5 to 10 days	370	370 to 375
Week 64	449	447 to 451

PK concentrations will be analyzed following time windows as defined below in Table 8. If the date of the previous injection is unknown, the SAR156597 concentration will not be considered for the analysis.

Table 8 - Time windows for PK variables definition

PK variables	Time
Cmax	W52 + 5 to 10 days
Ctrough	-2h (day 1, Weeks 12,24,36,52)
Cfollow-up	Week 64

2.5.5 Unscheduled visits

For all analyses, unscheduled visit measurements may be used to provide a measurement for a time point, a baseline, if appropriate according to their definitions.

2.5.6 Pooling of centers for statistical analyses

Not applicable.

2.5.7 Statistical technical issues

Not applicable.

3 INTERIM ANALYSIS

As there was no major enrollment issue, no futility analysis was performed.

Early analysis:

The primary analysis of efficacy and safety will be performed on the data collected during the 52-week treatment period. The results of this analysis will not be used to change the conduct of the follow-up in any aspect. The 52-week data will not be reanalyzed at the end of the study. Individuals who have access to patients' source documents will remain blinded with regards to the treatment arm of individual patients throughout the study. Treatment arm will not be disclosed on Data review and Surveillance report.

4 DATABASE LOCK

There are two database locks planned for this study as follows:

- First database lock (for early analysis of the main efficacy and safety analyses evaluation at Week 52): 4 weeks after last patient last efficacy visit (Visit 9 Week 52).
- Final database lock (for final analysis of efficacy endpoints and final safety analysis at Week 64): 4 weeks after last patient last visit (Visit 10 Week 64).

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS version 9.4.

6 REFERENCES

1. Jones P. St George's Respiratory Questionnaire. University of London. (June 2009)

2. van Reenen M. and Janssen B. EQ-5D-5L User Guide. EQ-5D. (April 2015)

3. Roger, J. A tool to fit Bayesian Gaussian Repeated Measures models within SAS proc MCM using conjugate priors. DIA working group on Missing Data. (2015)

7 LIST OF APPENDICES

Appendix A Potentially clinically significant abnormalities criteria

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES – March 2014

for phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Clinical Chemistry		
ALT	By distribution analysis : >3 ULN >5 ULN >10 ULN >20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
AST	By distribution analysis : >3 ULN >5 ULN >10 ULN >20 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. Categories are cumulative. First row is mandatory. Rows following one mentioning zero can be deleted.
Alkaline Phosphatase	>1.5 ULN	Enzymes activities must be expressed in ULN, not in IU/L. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008.
Total Bilirubin	>1.5 ULN >2 ULN	Must be expressed in ULN, not in µmol/L or mg/L. Categories are cumulative. Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008.
Conjugated Bilirubin	>35% Total Bilirubin and TBILI>1.5 ULN	Conjugated bilirubin dosed on a case-by-case basis.
ALT and Total Bilirubin	ALT>3 ULN and TBILI>2 ULN	Concept paper on DILI – FDA draft Guidance Oct 2007. Internal DILI WG Oct 2008. To be counted within a same treatment phase, whatever the interval between measurement.

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES – March 2014

for phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
CPK	>3 ULN	FDA Feb 2005.
	>10 ULN	Am J Cardiol April 2006.
		Categories are cumulative.
		First row is mandatory. Rows following one mentioning zero can be deleted.
CLcr (mL/min)	<15 (end stage renal disease)	FDA draft Guidance 2010
(Estimated creatinine clearance based on the Cokcroft-Gault equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling
	≥30 - <60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥90 (normal GFR)	
eGFR (mL/min/1.73m2)	<15 (end stage renal disease)	FDA draft Guidance 2010
(Estimate of GFR based on an MDRD equation)	≥15 - <30 (severe decrease in GFR)	Pharmacokinetics in patients with impaired renal function-study design, data analysis, and impact on dosing and labeling
	≥30 - <60 (moderate decrease in GFR)	
	≥60 - <90 (mild decrease in GFR)	
	≥90 (normal GFR)	
Creatinine	≥150 µmol/L (Adults)	Benichou C., 1994.
	≥30% change from baseline	
	≥100% change from baseline	
Uric Acid		Harrison- Principles of internal Medicine 17th Ed., 2008.
Hyperuricemia	>408 µmol/L	
Hypouricemia	<120 µmol/L	
Blood Urea Nitrogen	≥17 mmol/L	
Chloride	<80 mmol/L	
	>115 mmol/L	
Sodium	≤129 mmol/L	
	≥160 mmol/L	
Potassium	<3 mmol/L	FDA Feb 2005.
	≥5.5 mmol/L	
Total Cholesterol	≥7.74 mmol/L	Threshold for therapeutic intervention.
Triglycerides	≥4.6 mmol/L	Threshold for therapeutic intervention.
Lipasemia	≥3 ULN	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES – March 2014

for phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Amylasemia	≥3 ULN	
Glucose		
Hypoglycaemia	≤3.9 mmol/L and <LLN	ADA May 2005.
Hyperglycaemia	≥11.1 mmol/L (unfasted); ≥7 mmol/L (fasted)	ADA Jan 2008.
HbA1c	>8%	
Albumin	≤25 g/L	
CRP	>2 ULN or >10 mg/L (if ULN not provided)	FDA Sept 2005.
Hematology		
WBC	<3.0 Giga/L (Non-Black); <2.0 Giga/L (Black)	Increase in WBC: not relevant.
	≥16.0 Giga/L	To be interpreted only if no differential count available.
Lymphocytes	>4.0 Giga/L	
Neutrophils	<1.5 Giga/L (Non-Black);<1.0 Giga/L (Black)	International Consensus meeting on drug-induced blood cytopenias, 1991. FDA criteria.
Monocytes	>0.7 Giga/L	
Basophils	>0.1 Giga/L	
Eosinophils	>0.5 Giga/L or >ULN (if ULN≥0.5 Giga/L)	Harrison- Principles of internal Medicine 17th Ed., 2008.
Hemoglobin	≤115 g/L (Male); ≤95 g/L (Female) ≥185 g/L (Male); ≥165 g/L (Female) Decrease from Baseline ≥20 g/L	Criteria based upon decrease from baseline are more relevant than based on absolute value. Other categories for decrease from baseline can be used (≥30 g/L, ≥40 g/L, ≥50 g/L).
Hematocrit	≤0.37 v/v (Male) ; ≤0.32 v/v (Female) ≥0.55 v/v (Male) ; ≥0.5 v/v (Female)	
RBC	≥6 Tera/L	Unless specifically required for particular drug development, the analysis is redundant with that of Hb. Otherwise, consider FDA criteria.
Platelets	<100 Giga/L ≥700 Giga/L	International Consensus meeting on drug-induced blood cytopenias, 1991.
Urinalysis		
pH	≤4.6 ≥8	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES – March 2014

for phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
Vital signs		
HR	≤50 bpm and decrease from baseline ≥20 bpm ≥120 bpm and increase from baseline≥20 bpm	To be applied for all positions (including missing) except STANDING.
SBP	≤95 mmHg and decrease from baseline ≥20mmHg ≥160 mmHg and increase from baseline ≥20 mmHg	To be applied for all positions (including missing) except STANDING.
DBP	≤45 mmHg and decrease from baseline ≥10 mmHg ≥110 mmHg and increase from baseline ≥10 mmHg	To be applied for all positions (including missing) except STANDING.
Orthostatic Hypotension		
Orthostatic SDB	≤-20 mmHg	
Orthostatic DBP	≤-10 mmHg	
Weight	≥5% increase from baseline ≥5% decrease from baseline	FDA Feb 2007.
ECG		Ref.: ICH E14 guidance (2005) and E14 Q&A (2012), and Cardiac Safety Research Consortium White Paper on PR and QRS (Nada et al. Am Heart J. 2013; 165(4) : 489-500)
HR	<50 bpm	Categories are cumulative
	<50 bpm and decrease from baseline ≥20 bpm	
	<40 bpm	
	<40 bpm and decrease from baseline ≥20 bpm	
	<30 bpm	
	<30 bpm and decrease from baseline ≥20 bpm	
	>90 bpm	Categories are cumulative
	>90 bpm and increase from baseline ≥20bpm	
	>100 bpm	
	>100 bpm and increase from baseline ≥20bpm	
	>120 bpm	
	>120 bpm and increase from baseline ≥20 bpm	

CRITERIA for POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES – March 2014

for phase 2/3 studies (oncology excepted)

Parameter	PCSA	Comments
PR	>200 ms	Categories are cumulative
	>200 ms and increase from baseline ≥25%	
	>220 ms	
	>220 ms and increase from baseline ≥25%	
	>240 ms	
	>240 ms and increase from baseline ≥25%	
QRS	>110 ms	Categories are cumulative
	>110 msec and increase from baseline ≥25%	
	>120 ms	
	>120 ms and increase from baseline ≥25%	
QT	>500 ms	
QTc	Absolute values (ms)	To be applied to any kind of QT correction formula. Absolute values categories are cumulative
	>450 ms	
	>480 ms	QTc >480 ms and ΔQTc>60 ms are the 2 PCSA categories to be identified in individual subjects/patients listings.
	>500 ms	
	Increase from baseline	
	Increase from baseline]30-60] ms	
	Increase from baseline >60 ms	

Appendix B Summary of statistical analyses

EFFICACY ANALYSIS

Endpoint	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
Primary endpoint					
%predicted FVC: Change from baseline at week 52	mITT	Rank ANCOVA with adjustment covariates: stratification factor (with/without Background Therapy), %predicted FVC baseline; and a treatment fixed effect. Imputation of W52 values using linear regression (according to treatment arm and treated status 52 weeks), time-to-death for ranking deaths and time-to-lung transplants for ranking lung transplants.	MMRM with treatment arm, stratification factor, % predicted FVC baseline, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction. Multiple imputation of W52 values using CIR (other) and translated distribution of their last available assessments (death or lung transplant).	Subgroups: Gender, age ≤65 or <65, Smoking status, with/without Background Therapy, Baseline %predicted FVC% categories, ANA at baseline (Positive/Negative) using both rank ANCOVA model and MMRM.	MMRM on-treatment data with treatment arm, stratification factor, % predicted FVC baseline, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction. No imputation. Cumulative distribution of decline from baseline at 24 weeks and at 52 weeks in % predicted FVC
Secondary endpoints					
Time-to-disease progression at 52 weeks: time-to-first event between a decrease in change from baseline in %predicted FVC≥10%, a decrease in change from baseline in %predicted DLCO≥15%, death and lung transplant.	mITT	Log-rank stratified on stratification factor (with/without Background Therapy); Kaplan-Meier representation and Hazard Ratio using a Cox model with treatment, stratification factor,. No imputation.	No	Subgroups: Gender, age≤65 or <65, Smoking status, with/without Background Therapy, Baseline %predicted FVC% categories, ANA at baseline (Positive/Negative).	Time-to-disease progression at 24 weeks: Hazard Ratio using a Cox model with treatment, stratification factor , stratification factor-by-treatment. No imputation.

Endpoint	Analysis population	Primary analysis	Supportive analysis	Subgroup analysis	Other analyses
Cumulative incidence of all-cause mortality	*mITT*	*Kaplan-Meier representation and Hazard Ratio using a Cox model with treatment, stratification factor ,. No imputation.*	*No*	*No*	*IPF-related deaths: Kaplan-Meier representation and Hazard Ratio using a Cox model with treatment, stratification factor , stratification factor-by-treatment. No imputation.*
Exploratory endpoints					
Change from baseline to week 52 in %predicted DLCO	*mITT*	*MMRM with treatment arm, stratification factor, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, % predicted DLCO baseline.No imputation.*	*No*	*No*	*No*
Annual rate of acute IPF exacerbations	*mITT*	*Annual rate from negative binomial model adjusted on stratification factor (with/without Background Therapy) and stratification factor-by-treatment interaction. No imputation.*	*No*	*No*	*No*
Annual rates of respiratory and non-elective hospitalizations	*mITT*	*Annual rates from negative binomial model adjusted on stratification factor (with/without Background Therapy) and stratification factor-by-treatment interaction. No imputation.*	*No*	*No*	*No*
Change from baseline to week 52 in 6-MWT distance	*mITT*	*MMRM with the fixed categorical effects of treatment arm, stratification factor, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, supplemental oxygen use at baseline, 6-MWT distance at baseline. No imputation.*	*No*	*No*	*No*

Endpoint	*Analysis population*	*Primary analysis*	*Supportive analysis*	*Subgroup analysis*	*Other analyses*
Change from baseline to week 52 in HRCT lung interstitial score (Total Fibrosing ILD score)	*mITT*	*Ancova with baseline value, emphysema score at baseline, stratification factor, treatment, stratification factor-by-treatment interaction. No imputation.*	*No*	*No*	*No*
Change from baseline to week 52 in SGRQ	*mITT*	*MMRM with treatment arm, stratification factor, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, SGRQ total score at baseline. No imputation.*	*No*	*No*	*No*
Change from baseline to week 52 in EQ-5D-5L	*mITT*	*MMRM with treatment arm, stratification factor, time point, treatment-by-time point interaction, stratification factor-by-treatment-by-time point interaction, as EQ-5D-5L total score at baseline . No imputation.*	*No*	*No*	*No*

Appendix C Systemic corticosteroids routes

Route
Buccal
Enteral
Administration Via Hemodialysis
Intraarterial
Intraduodenal
Intraesophageal
Intragastric
Intraileal
Intrajejunal
Intraluminal
Intralymphatic
Intramuscular
Intramuscular or Subcutaneous
Intravascular
Intravenous
Intravenous Bolus
Intravenous Drip
Intraventricular
Nasogastric
Oral
Oral Gavage
Oromucosal
Oropharyngeal
Parenteral
Percutaneous
Rectal
Subcutaneous
Sublingual
Submucosal
Transdermal
Mucosal

Appendix D EQ-5D-5L Scoring

```
* Creation of variable :qsstresn ;
data qs1;
      set adsd.QS_all;
      where QSSTAT ne 'NOT DONE';
      if index(qsorres,"NO")>0 or index (qsorres,"NOT")>0
      then qsstresn=1;
      else if index(qsorres,"SOME") or index(qsorres,"MODERATE") or index(qsorres,"MODERATELY")
            then qsstresn=2;
            else if index(qsorres,"CONFINED") or index(qsorres,"UNABLE") or index(qsorres, "EXTREME") or
index(qsorres, "EXTREMELY")
                  then qsstresn=3;
                  else qsstresn=.;
run;

proc sort data=qs1 out=qs2;
      by usubjid visit QSTESTCD;
run;

data qs3;
      set qs2;
      by usubjid visit QSTESTCD;
      retain qsstreni profil;
      **********Mobility*********;
      if QSTESTCD="QS1" and qsstresn=1 then qsstreni=1;
      if QSTESTCD="QS1" and qsstresn=2 then qsstreni=1-0.069;
      if QSTESTCD="QS1" and qsstresn=3 then qsstreni=1-0.314;
      if QSTESTCD="QS1" then if qsstresn=. then profil=0; else profil=qsstresn;
      ********Self-care**********;
      if QSTESTCD="QS2" and qsstresn=1 then qsstreni=qsstreni;
      if QSTESTCD="QS2" and qsstresn=2 then qsstreni=qsstreni-0.104;
```

```
      if QSTESTCD="QS2" and qsstresn=3 then qsstreni=qsstreni-0.214;
      if QSTESTCD="QS2" then if qsstresn=. then profil=compress(profil!!0); else
profil=compress(profil!!qsstresn);
      ******Usual activities*****;
      if QSTESTCD="QS3" and qsstresn=1 then qsstreni=qsstreni;
      if QSTESTCD="QS3" and qsstresn=2 then qsstreni=qsstreni-0.036;
      if QSTESTCD="QS3" and qsstresn=3 then qsstreni=qsstreni-0.094;
      if QSTESTCD="QS3" then if qsstresn=. then profil=compress(profil!!0); else
profil=compress(profil!!qsstresn);
      *****Pain/discomfort******;
      if QSTESTCD="QS4" and qsstresn=1 then qsstreni=qsstreni;
      if QSTESTCD="QS4" and qsstresn=2 then qsstreni=qsstreni-0.123;
      if QSTESTCD="QS4" and qsstresn=3 then qsstreni=qsstreni-0.386;
      if QSTESTCD="QS4" then if qsstresn=. then profil=compress(profil!!0); else
profil=compress(profil!!qsstresn);
      *****Anxiety/depression*****;
      if QSTESTCD="QS5" and qsstresn=1 then qsstreni=qsstreni;
      if QSTESTCD="QS5" and qsstresn=2 then qsstreni=qsstreni-0.071;
      if QSTESTCD="QS5" and qsstresn=3 then qsstreni=qsstreni-0.236;
      if QSTESTCD="QS5" then if qsstresn=. then profil=compress(profil!!0); else
profil=compress(profil!!qsstresn);
      output;
      if last.visit then do;
            qsstresn=profil;

            if index(profil,'3') ne 0 then qsstreni=qsstreni-0.269;
            if index(profil,'2') ne 0 or index(profil,'3') ne 0 then qsstreni=qsstreni-0.081;
                  else if profil=0 then qsstreni=.;
                        else qsstreni=qsstreni;
            QSTESTCD='QS6';
            QSTEST='Total score';
            QSORRES='';
            QSSTRESC='';
            output;
      end;
run;
```

Appendix E MedDRA codelists for AESIs and related hepatic disorder

requested list	type (cmq, smq, others)	name or code (cmqname, smqcd, others)			description (cmqdesc or SMQ_ShortName, others)	scope (smq) / prim_sec (soc)
reqlist	type	Codename			desc	scope
Acute renal failure	CMQ	GLB_ACUTE RENAL FAILURE NARROW			Acute renal failure Narrow	Narrow
Tuberculosis	CMQ	GZ402673_TUBERCULOSIS			Tuberculosis	
Anaphylactic reactions or acute allergic reactions that require immediate treatment	CMQ	GLB_ANAPHYLACTIC REACTION NARROW			Anaphylactic reaction Narrow	Narrow
[redacted]						
Hepatic disorder	CMQ	GLB_HEPATIC DISORDERS NARROW	Hepatic disorders Narrow	Narrow	Hepatic disorders Narrow	Narrow

Appendix F WHODD codelists for prior and concomitant medications

requested lists	cdgname	cdgdesc
Use of any cytotoxic/immunosuppressive agent including but not limited to azathioprine, cyclophosphamide, methotrexate, and cyclosporine	CDG00696	Cytotoxic_or_immunosuppressive_agent
Use of any cytokine modulators (alefacept, etanercept, adalimumab, efalizumab, infliximab, golimumab, certolizumab, rituximab)	CDG00697	Cytokine modulators
Prohibited concomitant biologic agents	CDGSn073	Biologicals NARROW
Prohibited concomitant Alefacept	CDG00736	Alefacept_mono_and_multi_ingredients
Initiation of medications for suspected tuberculosis	CDG00737	Medications for suspected tuberculosis
ORAL CORTICOSTEROIDS	CDGSn010	Corticosteroids NARROW
acetylcystein	CDG20006	ACETYLCYSTEIN_mono_and_multi_ingredients
n-acetylcystein	CDG20006	ACETYLCYSTEIN_mono_and_multi_ingredients
NAC	CDG20006	ACETYLCYSTEIN_mono_and_multi_ingredients
Anti-Acids	CDG00592	ANTACIDS
Proton pump inhibitors (PPI)	CDG00593	PROTON PUMP INHIBITORS
H2-receptor blockers	CDGSn292	H2 receptor antagonists Narrow

DRI11772 16.1.9 Statistical analysis plan

ELECTRONIC SIGNATURES

Signed by	Meaning of Signature	Server Date (dd-MMM-yyyy HH:mm)
[redacted]	Clinical Approval	16-Jun-2017 22:13 GMT+0200
[redacted]	Clinical Approval	19-Jun-2017 16:20 GMT+0200