STATISTICAL ANALYSIS PLAN

STUDY TITLE:	**A Phase 2, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of FG-3019 in Patients with Idiopathic Pulmonary Fibrosis**
PROTOCOL NUMBER:	**Protocol FGCL-3019-067, Amendment 5.0**
SPONSOR:	FibroGen, Inc. 409 Illinois Street San Francisco, California 94158 USA
Version	FINAL
Release Date:	July 26, 2017

CONFIDENTIALITY STATEMENT

STATISTICAL ANALYSIS PLAN SIGNATURE PAGE

Approvals

I have reviewed and accepted the information in this document to be a true and accurate representation of the Statistical Analysis Plan for Study FGCL-3019-067 (Amendment 5.0).

Initiator:

Signature: _____   **Date:** _____

        █████████
        Project Statistician, ████████████████
        FibroGen, Inc.

Reviewed by:

Signature: _____   **Date:** _____

        ████████
        █████████████████

Signature: _____   **Date:** _____

        ████████
        ████████████████████

Signature: _____   **Date:** _____

        █████
        Project Statistician
        Q-square Business Intelligence

Signature: _____   **Date:** _____

        ██████████████

Signature Significance

The following significance is lent to the signatures on the Approvals page of this document.

Signatory	Significance
Initiator	By signing, the author is attesting that the content of the document is complete and accurate.
Reviewer	By signing, the reviewer is attesting that the document's approach and contents are compliant with the study protocol, all appropriate, regulatory requirements, and other significant guidelines. This individual(s) has reviewed the document for accuracy and completeness.

CHANGE HISTORY

Version	Date	Description
0.1	July 6, 2015	Draft for the AE summaries for the 2015 IND Update
0.2	January 11, 2016	Safety data summaries
0.3	February 4, 2016	Add summaries of FVC data
0.4	March 1, 2016	First draft to study team
0.5	April 29, 2016	Add more details
0.6	October 31, 2016	Add more details
0.7	January 9, 2017	Add multiple imputation procedure to key efficacy endpoints.
0.8	January 31, 2017	Add more details
0.9	June 7, 2017	Revise imputation methods and efficacy analysis models, per discussion with ██████ ████ and ██████
1.0	July 17, 2017	Add references and justifications of the statistical methods; pre-final
1.1	July 21, 2017	Incorporated comments from the previous version
FINAL	July 26, 2017	Sign off

Table of Contents

List of Abbreviations

Δ	Change from baseline
ΔFVCpp48	Change from baseline to Week 48 in FVC % predicted (this abbreviation rule is applied in other parameters)
ADaM	Analysis Data Model
AE	Adverse event
ALT	Alanine aminotransferase
ANCOVA	Analysis of covariance
ANOVA	Analysis of variance
AST	Aspartate aminotransferase
ATC	Anatomical Therapeutic Chemical
AUC	Area under the curve
BL	Baseline
BP	Blood pressure
BPM	Beats per minute
CAD	Computer assisted diagnosis
CBC	Complete blood count
CM	Concomitant medication
C_{max}	maximal concentration
C_{min}	minimal concentration
CRF	Case report form
CS	Clinically significant
CTCAE	Common Terminology Criteria for Adverse Events
CTGF	Connective tissue growth factor
DLCO	Diffusing capacity of the lung for carbon monoxide
DLCOpp	DLCO % predicted
DLCOpp-hb	DLCO % predicted corrected for hemoglobin
EMA	European Medicines Agency

ECG	Electrocardiogram
EE	Efficacy evaluable
EOS	End of study
EOT	End of treatment
EX	Extension
FAS	Full analysis set
FEV$_1$	Forced expiratory volume in 1 second
FG-3019	FibroGen-3019 (recombinant human monoclonal antibody)
FIB	HRCT fibrosis score
FIB-WZ	Fibrosis at worst zone
FRC	Functional residual capacity
FVC	Forced vital capacity
FVC (L)	FVC in litre
FVCpp	FVC % predicted
GG	HRCT ground glass score
HAHA	Human anti-human antibody
Hb	Hemoglobin
HC	HRCT honeycombing score
HRCT	High resolution computed tomography
HRQoL	health-related quality of life
IPF	Idiopathic pulmonary fibrosis
ITT	Intent-to-treat
IV	Intravenous
kg	Kilogram
LOCF	Last observation carried forward
LS Means	Least squares means
MedDRA	Medical Dictionary for Regulatory Activities

mg	Milligram
MAR	Missing at random
MCMC	Markov chain Monte Carlo
MI	Multiple Imputation
MI Imputed	Average imputed value from multiple imputations
MMRM	Mixed effect Model Repeated Measures
MNAR	Missing not at random
NCI	National Cancer Institute
NCS	Not clinically significant
OC	Observed cases
PD	Progressive disease
PFT	Pulmonary Function Test
PINP	Procollagen type I N-peptide
PIIINP	Procollagen type III N-peptide
PK	Pharmacokinetic
PMM	Pattern-mixture model
PRO	Patient reported outcome
PT	Prothrombin time, Preferred Term
PTT	Partial prothrombin time
QILD	HRCT quantitative interstitial lung disease
QoL	Quality of life
RPM	Respirations per minute
SAE	Serious adverse event
SAP	Statistical analysis plan
S.D.	Standard deviation
S.E. or s.e.	Standard error
SGRQ	Saint Georges Respiratory Questionnaire

SOBQ	Shortness of Breath Questionnaire
SOC	System organ class
SDTM	Study Data Tabulation Model
TEAE	Treatment emergent adverse event
TLC	Total lung capacity
TLFs	Tables, listings and figures
UCSD-SOBQ	University of California, San Diego Shortness of Breath Questionnaire
ULN	Upper limit of normal
WHODrug	World Health Organization Drug Dictionary

1   INTRODUCTION

This Statistical Analysis Plan (SAP), Part 1, documents planned analyses for the blinded Randomized Treatment Period for Study FGCL-3019-067 based on the study protocol dated 12DEC2016 (Amendment 5.0): A Phase 2, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Efficacy of FG-3019 in Patients with Idiopathic Pulmonary Fibrosis.

This study allows for subjects who complete the Blinded Randomized Treatment Period to continue receiving open-label treatment in an Extended Treatment Period, if they qualify. After all subjects have completed the Randomized Treatment Period, the database will be locked. Data from this study period will be summarized as a stand-alone package as Part 1 of the study summaries. Tables, listings, and figures (TLFs) in this package will include all data from subjects who did not enter the extended treatment period, and all data up to the date prior to the first extended treatment dosing date for subjects who entered the extended treatment period. Data from subjects participating in the Open-Label Extended Treatment Period will be summarized and analyzed separately in a Part 2 report. Summaries of pharmacokinetics, CTGF (post-treatment), HAHA, and biomarker (post-treatment) data will be conducted after treatment unblinding.

After the completion of the Extended Treatment Period, all data for subjects treated in extension will be summarized in Part 2 of the study summaries. Part 2 summaries will also include the overall safety profile for all subjects in the entire study. Additional analysis methods and data handling rules for Part 2 summaries will be documented in a separate SAP (Part 2).

This SAP includes statistical analysis methods, models, definitions, and data handling rules. It supersedes the statistical sections in the protocol in case of differences.

2 STUDY OBJECTIVES

The overall objective of this study is to evaluate the safety, tolerability, and efficacy of FG 3019 in subjects with IPF.

2.1 Primary Safety Objective

To determine the safety and tolerability of FG-3019 administered at a dose of 30 mg/kg by intravenous (IV) infusion every 3 weeks.

2.2 Primary Efficacy Objective

To determine the effect of FG-3019 on forced vital capacity (FVC) (percent of predicted value) in the target population.

2.3 Secondary objectives

- To evaluate the effect of FG-3019 on the extent of pulmonary fibrosis as measured by high resolution computed tomography (HRCT) scans of the chest

- To evaluate the relationship between changes in quantified scores of pulmonary fibrosis and clinical outcomes

- To evaluate the effect of FG-3019 on progression of IPF and the frequency of respiratory-related hospitalizations and respiratory-related mortality

- To evaluate the effect of FG-3019 on health-related quality of life (HRQoL)

2.4 Exploratory Objectives

- To evaluate the effect of FG-3019 on FVC (liters)

- To evaluate the effect of FG-3019 on a panel of serum and plasma biomarkers

- To evaluate the relationship between known chromosomal loci associated with IPF and the response to FG-3019

2.5 Objectives of the Substudy (Amendment 4.0)

To assess the safety, pharmacokinetics (PK), and efficacy of FG-3019 when given with standard of care (pirfenidone or nintedanib) IPF background medication.

3 STUDY DESIGN

3.1 Overview

This is a Phase 2, randomized, double-blind, placebo-controlled, multicenter study to evaluate the safety, tolerability, and efficacy of FG-3019 30 mg/kg comparing with placebo. The study includes two components: the main study which includes subjects not receiving other IPF treatment and the Substudy which includes subjects receiving either pirfenidone or nintedanib. Below is a brief description of the study components.

Table 1. Overview of the Main and Substudy

	Main Study	**Substudy**
Planned sample size	136 Total 110 first-line treatment 26 prior treatment failure	30 on pirfenidone 30 on nintedanib
Randomization ratio (FG-3019 : placebo)	1:1	2:1
Dosing	Day 1 -- Week 45; Every 3 weeks; 16 infusions	Day 1 – Week 21; Every 3 weeks; 8 infusions
Dose level	30 mg/kg	15 mg/kg at first 2 infusions; 30 mg/kg the next 6 infusions
Last FVC and HRCT assessment	Week 48	Week 24
PK	None	PK of FG-3019 and PK of background medication
Extended treatment	Subjects meet protocol defined criteria	Not offered

3.2 Study Population

Subjects 40 to 80 years old who have IPF diagnosed in accordance with criteria published in the 2011 international consensus guidelines.

Subjects who have failed or are intolerant of products approved for treating IPF are eligible under amendment 1. No subjects are enrolled in this category.

Subjects who are on pirfenidone or nintedanib are eligible for substudy under amendment 4.

3.3 Study Assessments

This section includes a brief summary of the study assessments in the Randomized Treatment Period. More detailed descriptions can be found in the protocol.

- Spirometry (FVC and FEV1) is performed at Screening, Day 1, and every 12 weeks.

- HRCT scans of the chest are performed at Screening and every 24 weeks.

- Saint Georges Respiratory Questionnaire (SGRQ) is administered on Day 1 and every 12 weeks.

- Serum samples for central lab tests (CBC with differentials and serum chemistry) are collected at Screening, Day 1, and every 12 weeks. Spot urine samples for urinalysis are collected with the same schedule.

- Serum and plasma samples for biomarker tests (including CTGF) are collected on Day 1 and every 24 weeks.

- Blood sample for DNA testing is collected on Day 1.

- Plasma samples for assessing HAHA are collected on Day 1 and at the End of Study (EOS) safety follow-up.

- Adverse events, concomitant medications, procedures, non-drug therapies, oxygen use are collected from signing ICF through End of Study (EOS).

- Vital signs are measured at every clinical visit. On dosing days, vital signs are measured at pre-, intra-, and post-infusion.

- Electrocardiograms (ECG) are performed at Screening and every 24 weeks.

- A complete physical examination is performed at Screening and at Week 48. Focused physical examinations are performed every 6 weeks from Day 1 to Weeks 42.

- For subjects enrolled in the substudy, blood samples for assessing FG-3019 PK and PK for pirfenidone or nintedanib are collected on Day 1, Week 9, and Week 21.

- DLCO is measured at Screening to assess entry eligibility for all subjects. For subjects who enrolled prior to Amendement 2, DLCO is measured on Day 1 prior to first dose and every 12 weeks during the randomized treatment period.

- For subjects enrolled under Amendment 2 and subsequent amendments, post-treatment DLCO, lung volumes (TLC and FRC), UCSD questionnaire are not evaluated.

Assessments and sample collections on Day 1 are all performed prior to first infusion. For subjects who discontinue the study early, evaluations scheduled for Week 48 are performed at the visit of early termination.

3.4 Sample Size Determination

This study is planned to enroll approximately 136 subjects with no background therapy and 60 subjects with background therapy.

Among those who are not receiving background therapy, it is planned to enroll 110 subjects who are seeking first-line therapy and 26 subjects who failed pirfenidone and/or nintedanib treatment. First-line therapy is defined as no prior therapy with pirfenidone/nintedanib or <10% relative decrease in FVC % predicted while receiving either pirfenidone or nintedanib. Failure of pirfenidone and or nintedanib treatment is defined as having ≥10% relative decrease in FVC % predicted during prior pirfenidone/nintedanib treatment. The effect of first-line treatment is the primary interest of the study, and is the basis for sample size determination. Assuming a 10% dropout rate and a common standard deviation of 7.3, a sample size of 55 subjects per arm in the primary analysis population provides approximately 80% power to detect a treatment difference of 4.2 in change from baseline to Week 48 in percent predicted FVC, using a two sample t-test at the two sided significance level of 0.05. The expected treatment difference is estimated using data from study FG-3019-049 and historical placebo data from published studies. In Study FG-3019-049, a mean decline of 2.8 in % predicted FVC from baseline to Week 48 was observed. The expected decline of 7.0 % predicted FVC in the placebo arm is estimated based on the weighted average of published studies (Azuma, 2005, Daniels, 2010, Demedts, 2005, King, Jr., 2008, King, Jr., 2009, King, Jr., 2011, Meier-Kriesche, 2004, Noble, 2011, Noth, 2012, Raghu, 2004, Raghu, 2008, Raghu, 2012, Richeldi, 2011, Taniguchi, 2010, Zisman, 2010).

No subject is enrolled in the category of pirfenidone/nintedanib failure.

3.5 Randomization

Subjects are randomized in a 1:1 ratio to receive FG-3019 or placebo under the original protocol. (main study)

Subjects are randomized in a 1:1 ratio, stratifying by prior therapy of pirfenidone/nintedanib (yes/no), under protocol amendment 1. (No subjects enrolled in this category)

Subjects in substudy under amendment 4 are randomized in a 2:1 ratio to FG-3019 + background therapy or background therapy alone, stratifying by background therapy (pirfenidone or nintedanib).

Subject randomization and drug dispense are performed by an interactive responsive technology (IRT) system developed by Endpoint Clinical Inc. (San Francisco, CA). The randomization specifications are documented in different versions of Randomization Specification Forms developed for the original protocol and subsequent amendments.

3.6 Treatment Blinding

After a subject completes the Randomized Treatment Period and assessments at Week 48, his/her treatment assignment will be unblinded only to designated personnel not associated with the study conduct to determine eligibility to enter the Extended Treatment Period. The treatment assignment is only used (by the mentioned designated personnel) with the purposes to assess eligibility into the Extension Period. Study subjects, investigators and site staff, as well as the operation team members at Fibrogen and contract organizations remain blinded to treatment assignment throughout the study. The study treatment unblinding for final analyses will occur after all subjects have completed the Randomized Treatment Period and the 4-week Follow-up, and the database has been cleaned and locked. Data modules that are not included in the database lock prior to unblinding are: PK, CTGF (post-treatment), HAHA, biomarkers (post-treatment), and HRCT visual reading data.

3.7 Protocol Amendments

Below is a brief description of changes from the original protocol that affect data analyses.

- In Amendment 1 (09 December 2014), the study sample size was increased to enroll a total of 136 subjects., of which up to 26 subjects with ≥10% relative decrease in forced vital capacity (FVC) % predicted during pirfenidone/nintedanib treatment will be enrolled to explore whether subjects who progress on pirfenidone and/or nintedanib may respond to FG-3019.

  - Randomization is stratified by subjects who received prior IPF treatment or not. The eligibility criterion for subjects in the FG-3019 arm to receive extended treatment is changed from <3% decrease from baseline in FVC % predicted value to no decrease.

  - DLCO is only assessed at Screening to evaluate entry eligibility; post-treatment assessments are not performed for subjects enrolled under Amendment 1 and later amendments.

  - UCSD questionnaire and Lung volume parameters (TLC and FRC) are not assessed in subjects enrolled under Amendment 1 and subsequent amendments.

  - Eligibility of receiving extended treatment for subjects who are randomized in FG-3019 is changed from no more than 3% decline from baseline to no decline from baseline.

- Amendment 2 (09 March 2015) includes only clarifications of the previous amendment.

- Protocol amendment 3 (28 May 2015), safety follow-up is changed from 7 to 10 weeks after the last dose of study medication.

- In protocol amendment 4 (04 January 2016, US and Canada sites only): a substudy is added to assess the safety, pharmacokinetics (PK), and efficacy of FG-3019 when given with standard of care IPF background medication (pirfenidone or nintedanib). A total of 60 subjects (30 with background therapy of pirfenidone and 30 with nintedanib) is planned to be enrolled into the substudy with randomization ratio of 2:1 to FG-3019 + background therapy or background therapy alone, stratifying by the type of background therapy.

- In Amendments 5 and 5.1 (12 December 2016): safety follow-up is changed from 10 weeks to 4 weeks after the last dose.

3.8 Changes from the Protocol

- The study is planned to enroll approximately 136 subjects, who will include a target of 110 subjects who are seeking first-line therapy and approximately 26 subjects who failed pirfenidone and/or nintedanib treatment. Due to lack of availability of subjects who failed pirfenidone and/or nintedanib treatment during the enrollment period, it was decided not to further pursue enrolling this category of subjects, as documented in Note-to-File dated 05 FEB 2016.

4 STUDY ENDPOINTS AND DEFINITIONS

4.1 Study Endpoints

4.1.1 Efficacy Endpoints

4.1.1.1 Primary Efficacy Endpoint

- Change from baseline in FVC (percent of predicted value) at Week 48

4.1.1.2 Other Efficacy Endpoints

- Change in pulmonary fibrosis score by quantitative HRCT at Week 24, Week 48, and later time points. (Note that efficacy endpoints after Week 48 are described in Part 2 of SAP.)

- Change from baseline in HRQoL at Week 24, Week 48, and later time points.

- Progression of IPF, defined as time from Day 1 to any one of the following:
  - Death from any cause
  - Absolute decline in FVC % of predicted value of ≥10% not due to intercurrent illness, confirmed by repeat spirometry
  - Clinical diagnosis of IPF progression

- Proportion of subjects with at least one respiratory-related hospitalization

- Proportion of subjects with respiratory-related death, censored at Week 52

- Categorical assessment of absolute change from baseline in FVC percent of predicted value at Week 48

4.1.2 PK Parameters and Biomarker Endpoints

- Change from baseline to Week 24 and from baseline to Week 48 in FVC (liters).

- Change from Day 1 to Week 48 in a panel of serum and plasma biomarkers that will be determined by FibroGen.

- Relationship between known chromosomal loci associated with IPF and the response to FG-3019.

- Population pharmacokinetics in all subjects to assess trough and peak drug concentrations (Substudy)
  - FG-3019 Cmax of 1st, 4th 8th dose
  - Pirfenidone blood concentration before the first morning dose and one hour after study drug dosing completion on Day 1, Week 9 and Week 21.
  - Nintedanib blood concentration before the first morning dose and two hours after study drug dosing completion on Day 1, Week 9 and Week 21.

4.1.3 Safety Endpoints

- AEs and SAEs

- Laboratory test abnormalities (graded by NCI Common Terminology Criteria for Adverse Events [NCI CTCAE] v4.0)

- Vital signs

- 12-lead electrocardiograms (ECGs)

- Physical examinations

- Human anti-human antibody (HAHA)

4.2   Detailed Definitions of the Study Endpoints

4.2.1   Change from baseline in FVC (% predicted) at Week 48

FVC in litres is measured at Screening, Day 1, and every 12 weeks during the randomized treatment period. Unscheduled assessments may be performed if subjects are experiencing respiratory conditions that are un-related to IPF and that affect PFT performance. The percent predicted FVC value (FVC % predicted) is calculated using the algorithm provided by Hankinson et al. (1999) for the corresponding gender-age group (Appendix I).

This primary efficacy endpoint, change from baseline to Week 48 in FVC % predicted (ΔFVCpp48), is defined only for subjects in the Full Analysis Set population (see definition in Section 5.1.3) in the main study. Baseline is defined as the mean of the last Screening visit and the Day 1 visit. Week 48 window is target day (Day 337) $\pm$ 6 weeks (see Section 5.7). ΔFVCpp48 will be characterized in mean, median, proportion of subjects with no decline (change >=0), and proportion of subjects with absolute decline $\geq$ 10% (change $\leq$-10%). The categorical parameters are described in Section 4.2.7.

The primary analysis of ΔFVCpp48 will be using a random coefficients linear regression model, assuming missing data to be missing at random (see Section 6.6.1.1). Only the observed cases will be included in the model. No imputation is needed for the primary analysis.

ΔFVCpp48 will be characterized in mean, median, categorical responses such as no decline (change >=0) and decline $\geq$ 10% (change $\leq$-10%), as well as empirical distribution. For these summaries, missing data will be imputed using the algorithm described in Section 5.4.1. Other imputation methods will be used for sensitivity analyses.

4.2.2   Change in Fibrosis Score by Quantitative HRCT at Weeks 24 and 48

HRCT scan is performed at Screening, Weeks 24 and 48 weeks during the randomized treatment period. HRCT images are sent to MedQIA, LLC, the central reading center, for quantification of disease severity.

Each lung is divided into five lobes (right upper, right middle, right lower, left upper, left lower). A computer algorithm takes samples of pixels; classifies each pixel in the sample into fibrotic reticulation or not; and calculates percent of fibrotic reticulation in each lung lobe. The average of the 5 lung lobes is defined to be fibrosis score of the whole lung.

Change from baseline in HRCT fibrosis score at Weeks 24 and 48 (ΔFIB24 and ΔFIB48) are only defined for subjects in the FAS Population. ΔFIB48 is defined for the main study only. The fibrosis score at Screening is defined as the baseline. Visit windows are defined as target day (day 169 and 337 for Weeks 24 and 48, respectively) $\pm$ 12 weeks (see Section 5.7). Missing data

are imputed using the MI method for subjects in the FAS Population. Detail imputation algorithm is described in Section 5.4.2. This endpoint will be characterized in mean, median, proportion of responders which are defined as change from baseline (1) <= 0 (not worsened) and (2) < 0 (improved). Responder definition is based on observed + imputed data defined in Section 5.4.2.

4.2.3   Change from baseline in HRQoL at Weeks 24 and 48

Subject's health-related quality of life (HRQoL) is assessed by the Saint George's Respiratory questionnaire (SGRQ). This instrument was designed to measure health impairment in patients with asthma and Chronic Obstructive Pulmonary Disease (COPD). Its application in IPF patient population was evaluated by Swigris et al. (2010). It was used in nintedanib pivotal trials to evaluate treatment benefit in IPF patients (Richeldi et al., 2014).

The questionnaire includes 17 questions covering 3 domains: Symptoms (SYM); Activity (ACT); and Impacts (IMP). The domain and the total (TOT) scores range from 0 to 100, where 0 indicates the best and 100 the worst possible health status. The scoring algorithm is provided by the SGRQ manual which is extracted in Appendix II. The manual includes rules on how to handle missing responses to individual questions. Subjects self-administer the questionnaire on Day 1 and at Weeks12, 24, 36, and 48.

Changes from baseline in the 3 domain scores and the total score (abbreviated as ΔSYM, ΔACT, ΔIMP, and ΔTOT) are defined only for subjects in the FAS Population. Day 1 assessment is defined as the baseline. Change from baseline to Week 48 is defined for the main study only. Visit windows are defined as target day ± 6 weeks (see Section 5.7). These endpoints will be estimated based on LS means obtained from a mixed effect repeated measures model (MMRM) described in Section 6.6.3. Missing data will not be imputed explicitly for these endpoints.

Detail description of the questionnaire is presented in St. George's Respiratory Questionnaire Manual, Version 2.3 (June 2009), by Jones, Paul, St. George's, University of London. A summary of the scoring algorithm is presented in Appendix II for reference.

4.2.4   Time to Progression of IPF

Time to progression of IPF is defined as the time from Day 1 to the first occurrence of the following events:

- Death from any cause

- Absolute decline in FVC % of predicted value of ≥10% not due to intercurrent illness, confirmed by repeat spirometry

- Clinical diagnosis of IPF progression

If a subject discontinued treatment due to disease progression as recorded on the Disposition CRF, it is considered clinical diagnosis of IPF progression. To determine whether a subject's FVC (% predicted) declined ≥10%, it should be either confirmed by a subsequent test or by the fact that the subject has discontinued the study. If confirmed, the earlier time is used to define this endpoint. Below are some examples.

1. If a subject's FVC declined below the threshold at Week 36 but the assessment increased above the threshold at the next test, this subject is not considered to have progressed at Week 36.

2. If a subject's FVC declined below the threshold at Week 36 and the assessment was confirmed at the next test, this subject is defined to have progressed at Week 36.

3. If a subject's FVC declined below the threshold at Week 36 and dropped out subsequently due to any reason, this subject is defined to have progressed at Week 36.

If the last assessment is at week 48, no confirmation will be required. Only observed FVC values (no imputed values) are used to assess progression. All events collected during year 1 (including Day 365) are included. If no event is observed, PFS is censored on the date of last PFT in the Randomized Treatment Period. Mortality events reported after EOS but before or on Day 365 are included in this endpoint.

Table 2.   Examples of Year 1 Time to Progression and Censoring Rules

Subject	Day of Last PFT	PFT result	Day of death or acute IPF	Event or Censor	Days	Comment
1	300	normal	320	event	320	
2	300	normal	370	censor	300	Event beyond day 365. Technically the subject's status between days 300-370 is unknown. It should not be marked as censor at 365.
3	300	normal	none	censor	300	

4.2.5   Proportion of Subjects with Respiratory-Related Hospitalization

Respiratory-related hospitalization is recorded on the AE CRF.

4.2.6   Proportion of Subjects with Respiratory-Related Death

Respiratory-related and IPF related death is recorded on the Death CRF.

4.2.7   Categorical Assessment of ΔFVCpp48

Two categorical endpoints will be summarized:

- Proportion of subjects with no decline (change ≥0)
- Proportion of subjects with absolute decline ≥ 10% (change ≤ -10%).

The observed and imputed value of ΔFVCpp48 (defined in Section 5.4.1) will be used to classify these categories. Dropouts due to deaths and disease progression are classified in the failure category by the imputation algorithm.

4.2.8 Prior and Concomitant Medication

Medications are coded using The World Health Organization Drug Dictionary (WHODD: March 1, 2015). The medications are reported by therapeutic class and generic name.

Prior medications are defined as those that are stopped prior to first infusion of FG-3019. These medications are summarized separately from the concomitant medications. If the stop date is incomplete or missing, the imputation rules are described in Section 5.4.10.

4.2.9 Study Drug Exposure and Treatment Compliance

Duration of study drug exposure = date of last non-zero dose – date of first dose + 1. This calculation does not account for FG-3019's half-life and the resulting drug exposure after the last dose.

Number of infusions is defined as the number of non-zero dose infusions.

The calculated dose amount of FG-3019 = nominal dose * body weight.

The latest measurement of body weight is used in calculating the dose for each infusion. The calculated dose amount is used to verify the actual amount administered.

Treatment compliance is defined as:

Compliance = (# of actual doses administered)/ (# of expected doses) * 100, where # of expected doses = (last dosing week)/3 + 1.

4.2.10 FG-3019 Pharmacokinetic (PK) Parameters

For subjects in the substudy, six FG-3019 plasma samples are collected: pre-any study medication and 0-1 hour from end of infusion on Day 1 and at Weeks 9 and 21. Post-dose samples are collected within 1 hour after the end of infusion. PK parameters include peak and trough concentrations, approximated by pre- and post-dose samples. Population PK (/PD) analysis may be performed separately with details described in the CSR (if applicable).

4.2.11 Pirfenidone and Nintedanib Pharmacokinetic (PK) Parameters

For subjects in the substudy, six plasma samples are collected for each subject to assess the pirfenidone and nintedanib concentration level: pre-dose and post-dose on Day 1 and at Weeks 9 and 21. Post-dose samples are collected at 3 hours after dosing for pirfenidone, and at 4 hours after dosing for nintedanib. Trough and peak concentration level, estimated by pre- and post-dosing samples, are evaluated. Metabolite-to-parent ratio will also be calculated in the post-dose samples for both pirfenidone and nintedanib.

4.2.12 AEs and SAEs

AEs and SAEs are collected on the AE CRFs. SAEs are special cases of AEs. In this document the more general terms such as AE and TEAE are used to cover both cases of AEs and SAEs.

The definitions of AEs, SAEs, severity, and relationship to study medication are described in Section 11of the protocol. Adverse events are coded using the Medical Dictionary for Regulatory Activities (MedDRA: 18.0) for system organ class (SOC) and preferred term for summary.

Treatment Emergent Adverse Event in the Study (TEAE) – Randomized Treatment Period

TEAEs during the randomized treatment period are defined as <u>new</u> or <u>worsening</u> AEs that occurred in the window of first dose of infusion of the blinded study drug (FG-3019 or placebo) (Day 1) and within 28 days of the last infusion of FG-3019/placebo or before the first FG-3019 infusion in the extended treatment period, whichever occurs first.

Table 3   Definition of TEAE and Study Period When AE Occurred

AE Onset Date Relative to Dosing*	AEPRIOR	TEAE	Study Period (EPOCH)
AE onset date < Day 1	Any	No	screen
AE onset date = Day 1	Yes**	No	screen
	No**	Yes	treatment
Day 1 < AE onset date ≤ date of last dose in randomized treatment period	Any	Yes	treatment
Date of last year 1 dose < AE onset date ≤ earlier date of {last year 1 dose + 28, Day 1-EX}	Any	Yes	Within 28 days from last dose follow-up
AE onset date > date of last year 1 dose + 28 and AE onset date < Day 1-EX for subjects entered EX	Any	No	Additional Follow-up beyond 70 days
*Dosing refers to FG-3019 or placebo infusion			
** If AE onset time is known, TEAE will be determined by AE onset date/time; AEPRIOR will not be used.			

AEs may be recorded in multiple records due to changes of characteristics such as seriousness, severity, frequency, treatment, and other aspects. These records will be linked together for summary purposes. The most severe grade, seriousness, relatedness to study drug, and AE frequency will be used in AE summary tables. Below are the rules:

(1) Multiple AE records are linked as one AE if they have the same preferred term and onset date of a later record is the same as stop date or stop date + 1 of the previous record.

(2) If the linked AE starts in the TEAE window as shown in Table 3, it is a TEAE.

(3) If the linked AE starts in screening, then the worst seriousness, severity grade, and AE frequency of the screening records will be used as the reference. A post-dose AE is

considered TEAE if it is worsened comparing to the reference. 'Worsened' is defined as increased grade in seriousness, severity, and frequency.

(4) Once an AE record is determined to be a TEAE, all <u>subsequent records</u> of the linked AE, including the records past 28 days after last dose of the randomization treatment period, are considered TEAE. The highest grade of severity and seriousness of the <u>linked TEAE will be included in the AE summary tables.</u>

The hypothetic example below illustrates the algorithm. Of note, AE #1 occurred in Screening; AE #2 and #3 were not worse than AE #1; these 3 AEs are not TEAEs. AE#4 worsened during the TEAE window; the subsequent AE #5 is also a TEAE by the algorithm.

Table 4   Example of Definition of TEAE for Multiple Records of the Same AE

AE #	Onset	AE	Severity	Seriousness	Related to study drug?	TEAE
1	Screening	Headache	2	N	N	N
2	Day 2	Headache	1	N	N	N
3	Last dose	Headache	2	N	Y	N
4	10 days post last dose	Headache	3	Y	N	Y
5	45 days post last dose	Headache	5	Y	N	Y
Consolidated for summary		Headache	5	Y	Y	Y

AE Severity Rating

AEs are rated by the investigators as "Grade 1 - mild", "Grade 2 - moderate", "Grade 3 - severe", "Grade 4 - life threatening", or "Grade 5 - fatal" based on the CTCAE v4.0 grading system as described in Section 11.3.3 of the protocol. If a subject reports multiple occurrences of an AE within one system organ class or preferred term, the most severe occurrence will be presented in the summary by severity rating. Missing severity rating will not be imputed and it is ranked the lowest in summary by severity rating.

AEs Related to FG-3019

Investigators determine the relationship of AEs with the study medication. For summary purposes, 'Possibly related' AEs are grouped with the 'Related' AEs. If related and unrelated occurrences of an AE within one system organ class or preferred term are reported in a subject, the related occurrence will be presented in the summary by relationship. Missing relationship evaluation will not be imputed and it is ranked the lowest in summary by relationship.

Infusion-associated AE

TEAEs are considered possible infusion reactions if they occurred during a FG-3019 infusion or within 12 hours from the end of a FG-3019 infusion regardless of causality assessment. In addition, TEAEs reported by investigators as possible infusion reactions and TEAEs with verbatim containing the word 'infusion', regardless of timing, are classified as possible infusion reactions.

4.2.13 Laboratory Abnormalities Central Laboratory Evaluations

Safety lab tests are evaluated at Screening to assess entry eligibility and on Day 1 prior to first dose. The last assessment prior to the first dose is defined as baseline. Post-treatment assessments are performed every 12 weeks to monitor subject safety. Lab test abnormality grading defined by CTCAE v4.0 is provided by ICON Central Laboratories, Inc. (see reference).

Complete Blood Count with Differentials

Hemoglobin, hematocrit, red blood cell (RBC) count, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), white blood cell (WBC) count, neutrophils, lymphocytes, monocytes, eosinophils, basophils, platelet count and calculated absolute neutrophil count (ANC).

Serum Chemistry

Sodium, potassium, chloride, bicarbonate, creatinine, blood urea nitrogen, glucose, calcium, phosphorus, uric acid, albumin, aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, and total bilirubin.

Coagulation

Prothrombin time (PT) and activated partial thromboplastin time (aPTT).

Urinalysis

Specific gravity, pH, protein, glucose, ketones, bilirubin, leukocyte esterase, nitrite, urobilinogen, blood, clarity, color, and microscopic examinations if any of the preceding urinalysis tests (other than glucose or ketones) are abnormal.

Serum Pregnancy Test

Human chorionic gonadotropin (HCG) for females of child-bearing potential only.

The CTCAE grading of lab test results is provided in Appendix I. Potentially clinically significant lab test results are those that are CTCAE grade 3 or higher.

In addition to using CTCAE grading, normal ranges for each numeric laboratory variable are provided by the central lab. Numeric variable values are compared to the normal range and are flagged as low, normal, or high.

Baseline for lab tests is defined as the last assessment prior to the first dose.

4.2.14 Vital Signs

Vital signs parameters include systolic/diastolic blood pressures, pulse rate, respiratory rate, and body temperature. Baseline for vital signs is defined as the average of the last Screening value and the Day 1 pre-infusion measurement. .

Potentially clinically significant vital sign changes are those which meet both criteria in Table 5a.

Table 5a.   Potentially Clinically Significant Changes in Vital Signs

Parameter	Observed Values	Change from Baseline
Systolic BP (mm Hg)	<90 or>140 mm Hg	>20 mm Hg
Diastolic BP (mm Hg)	<50 or>90 mm Hg	>15 mm Hg
Pulse Rate (bpm)	<50 or>100 bpm	>20 bpm

BP=blood pressure; bpm=beats per minute;

4.2.15 Electrocardiogram

ECG is evaluated qualitatively by investigator as:

Normal,

Abnormal, clinically significant, or

Abnormal, not clinically significant.

Abnormal findings are described.

Table 5b.   Criteria for Potentially Clinically Significant ECG

Parameter	Criteria
PR interval (msec)	<120 or >200
QT interval (msec)*	≤300 or > 450
RR interval (msec)	<600 or >1200

*(Uncorrected for heart rate. For QTc use >450 for males and >470 for females)

4.2.16 Physical Examination

Body systems are evaluated as normal or abnormal based on investigator assessment. For abnormal cases, descriptions are provided.

4.2.17 DNA Analysis

Fingerlin et al. (2013) performed a genome-wide association study in subjects with fibrotic idiopathic interstitial pneumonias (N = 1616) and controls (N = 4683) and determined an association of disease with TERT and MUC5B on chromosomes 5p15 and 11p15, respectively, the chromosome 3q26 region near TERC, and identified 7 novel chromosome loci. The novel loci include FAM13A (4q22), DSP (6p24), OBFC1 (10q24), ATP11A (13q34), DPP9 (19p13),

and chromosomal regions 7q22 and 15q14-15. Their findings indicate that genes involved in host defense, cell-cell adhesion, and DNA repair contribute to the risk of fibrotic pulmonary disease. Patient DNA is isolated from whole blood samples. We will evaluate the relationship between clinical response to FG-3019 or placebo, in terms of changes in pulmonary function and fibrosis, and the specific identified variants in these chromosomal loci. A categorical analysis for each response [FVC (% predicted) change ≤ -10% or change >0%] will be provided for each gene variant of interest in terms of relative expression in active and placebo patients, similar to Figure 4A in Fingerlin.

4.2.18 Connective Tissue Growth Factor (CTGF)

Plasma samples are collected to measure the levels of whole CTGF and whole + N-terminal fragment CTGF. Concentration level at each time point will be summarized descriptively.

4.2.19 Human Anti-Human Antibodies to FG-3019 (HAHA)

Blood samples are collected at the end of study visit which is between 4 – 10 weeks from the last dose. HAHA level is evaluated. A sample is considered reactive and specific for antibodies to FG-3019 if the unspiked sample signal is greater than the cut-point.

4.3 Definitions of Additional Endpoints

4.3.1 Change from Baseline in FVC (L) at Week 48

Change from Baseline in FVC (L) at Week 48 (abbreviated as ΔFVC48) is the same measure of lung function as ΔFVCpp48 in the unit of volume. It is similarly defined as ΔFVCpp48. For subjects with missing assessments at Week 48, missing data will be imputed following the imputation rules are described in Section 5.4.2.

4.3.2 Other Endpoints Related to Change from Baseline in FVC (L) and FVC (% predicted)

- Proportion of subjects with change in FVC % predicted > 0 at all assessment time points

- Change from baseline in FVC (L) at all assessment time points

- Categorical assessment of change from baseline in FVC(% predicted) at all assessment time points

- Percent of relative change from baseline in FVC (L) at all assessment time points

  Percent of relative change from baseline is defined as [FVC (L) at a post-baseline visit – baselined FVC (L)] / baseline FVC (L) * 100. Observed and imputed values (see Section 5.4.1) will be used in defining % relative change at Week 48.

-

4.3.3 Other PFT Parameters

Other PFT parameters include FEV1 (% predicted) and FEV1/FVC ratio. DLCO, TLC, and FRC are not measured in subjects enrolled under protocol amendment 2.0 and subsequent amendments. Predicted value of FEV1is calculated using the formula presented in Appendix II. For summaries by visit, assessments are grouped based on windowing rules described in Section 5.6. Baseline is defined as the mean values of the last Screening visit and Day 1.

4.3.4 HRCT Visual Assessment

Each lung is divided into five lobes (right upper, right middle, right lower, left upper, left lower). For baseline, the extent of pulmonary abnormality in each lobe is scored using the following 6-point scale: none, 1 to 5%, 6 to 25%, 26 to 50%, 51-75% and more than 75%. The score of the worst lung lobe is defined as the highest score of the five lobes. Three types of abnormality are assessed: pure ground glass (GG, increased lung attenuation in the absence of reticular interstitial thickening or architectural distortion), fibrosis (FIB, reticular inter and intra-lobular interstitial thickening, traction bronchiectasis, and bronchiolectasis), and honeycomb cysts (HC, clustered air-filled lung cysts with contiguous walls). In addition, the airways are assessed for presence or absence of bronchiectasis per lobe. If ANY lung lobe shows presence or bronchiectasis, then whole lung will be considered having presence of bronchiectasis. For follow-up, the change in the extent of pulmonary abnormality in each lung lobe is scored using the following 3-point scale: better, same, worse, comparing to Baseline. To avoid bias in comparing follow-up with baseline images, readers are presented with two sets of HRCT scans (Baseline and Follow-up) in a randomized order and in a blinded fashion.

HRCT images are assessed by two independent thoracic radiologists. Any discrepancies are reviewed by a third independent reader for a consensus read. Only the final assessments will be used in analyses.

4.3.5 Other Quantitative HRCT Reading Scores

In addition to the fibrosis score described in Section 4.2.2, ground glass (GG) and honeycombing (HC) scores are included as part of the quantitative HRCT reading. An overall score (QILD) is defined as the sum of FIB, GG, and HC.

The worst lung zone is defined as the lung zone with the highest fibrosis score. If there are two or more lung zones with the same highest fibrosis score, these lung zones are combined and the average scores of these zones will be used in the follow-up visits.

4.3.6 Rate of Change in FVC and DLCO Prior to Enrollment

The rate of change in FVC and DLCO prior to enrollment is define as the linear regression slope based on the data collected up to day 1 prior to enrollment. This parameter is summarized as baseline characteristics. These two parameters may be evaluated whether they are prognostic factors to the clinical outcomes and change in FVC over time.

5 GENERAL STATISTICAL CONSIDERATIONS

5.1 Analysis Populations

Summaries of baseline characteristics, safety, study drug exposure, and biomarker are based on available data in the Safety population. Efficacy analyses are based on the FAS population. Pharmacokinetic parameters are summarized based on the PK population. All subjects in the Safety population are included in the data listings.

5.1.1 Randomized Population

The Randomized Population consists of all subjects who have received a randomized treatment assignment, regardless of whether they received study treatment.

5.1.2 Safety Population

The Safety Population consists of all randomized subjects who have received any amount of study medication. If actual treatment received differs from the randomized treatment arm, the actual treatment arm will be used in safety data analysis.

5.1.3 Full Analysis Set Population

The Full Analysis Set (FAS) Population consists of randomized subjects who have received any amount of study medication and have evaluable FVC assessments at baseline and at least one evaluable post baseline. Subjects who are mis-diagnosed as IPF will be excluded from FAS. The FAS population will be used in the primary analyses of the PFT endpoints and other efficacy endpoints. If actual treatment received differs from the randomized treatment arm, the actual treatment arm will be used in efficacy data analysis.

5.1.4 Substudy Population

The Safety Population consists of randomized subjects who have received any amount of study medication. If actual treatment received differs from the randomized treatment arm, the actual treatment arm will be used for safety data analysis.

5.1.5 Substudy PK Population

Substudy PK population includes subjects in substudy who have received at least one dose of study medication and have corresponding PK concentration data.

5.1.6 Extension Safety Population

Extension Safety Population (Safety-EX) includes subjects who have received any amount of study medication during the extension period.

5.2 Hypotheses and Decision Rules

The hypothesis of the study is whether treatment of FG-3019 is superior to placebo in attenuating the decline of FVC from baseline to Week 48 in patients with IPF. It will be tested based on the primary endpoint, change from baseline to Week 48 in FVC (% predicted), using the ANCOVA model described in Section 6.6.1, at a two-sided significance level of 0.05.

5.3 Adjustment for Multiple Comparisons

Using a gatekeeping procedure, if the primary endpoint achieves the statistical significance level, the secondary hypothesis of whether treatment of FG-3019 is superior to placebo in change of

lung fibrosis score at Week 48 will be tested at the significance level of 0.05 (two-sided). The model of testing this secondary endpoint is described in Section 6.6.2.

If both the primary and the first secondary endpoint achieve the statistical significance level, analyses of all other secondary endpoints will be performed at the significance level of α=0.05, using two-sided tests. No adjustments will be made for multiple tests on secondary endpoints except for change in fibrosis score at week 48.

5.4   Handling Dropouts and Missing Data

Patients with IPF have relatively short median survival time. Disease progression (PD) is the most frequent reason of early treatment discontinuation in the current study. Missing data are primarily caused by discontinuation due to PD. Other infrequent causes of missing data include discontinuation due to AE or withdrawal of consent, technical failures, and incompleteness in patient reported outcome. As recommended by regulators, appropriate methods based on plausible assumptions on the nature of the data and the mechanism of missing data should be used in analysis with missing data. This section presents the missing data handling methods for different efficacy parameters and the rationale of the associated assumptions. In addition, imputation of missing or incomplete start or stop date of AE and concomitant medication is also included in this section.

5.4.1   Imputation for Missing Data in ΔFVCpp48 and ΔFVC48

5.4.1.1   Assumptions and Rationales

The primary analysis of ΔFVCpp48 will be using a random coefficient linear regression model based on the missing at random (MAR) assumption (section 6.6.1.1). Only observed cases will be included in the model; no imputation will be performed. However, in order to characterize ΔFVCpp48 and ΔFVC48 in mean, median, categorical responses such as no decline (change >=0) and decline ≥ 10% (change ≤-10%), as well as empirical distribution, it is necessary to impute the missing data. Based on clinical experience, FVC rate of decline in IPF patients is gradual in the absence of IPF exacerbation but can be drastic after an IPF exacerbation or disease progression. Therefore, it seems realistic to model the missing data in subjects who discontinued due to PD/death utilizing the trend in a subgroup with similar characterics, and model other types of missing data utilizing the trend of the overall study population.

This imputation strategy is conformed to the recommendations of the EMA Guideline on Missing Data in Confirmatory Clinical Trials (2011)

- *"An attractive approach for imputing missing data may be to employ a different pre-specified imputation technique for each different reason for withdrawal, rather than the same technique for all patients. While this would represent a relatively novel approach, there is no objection to this in principle. The strategy has more flexibility in handling different reasons for and timings of withdrawal and consequently the possible relationship between missing data and the outcome of interest. If used appropriately, it may better address the question of primary regulatory interest. The method also offers an intuitive framework for conducting a range of sensitivity analyses."*

- *"A combined strategy incorporating several methods for handling missingness (e.g. assume dropouts due to lack of efficacy and adverse events are MNAR and lost to follow-up are MAR) may also be considered."*

5.4.1.2 Imputation Procedure

Missing data in ΔFVCpp48 and ΔFVC48 are classified into 2 categories:

> (A) missing due to discontinuation for PD or death;

> (B) missing due to other reasons.

Step 1: Missing data in Category B are imputed as the average of the multiple imputed values from the MI procedure. The model specifications are provided in the next two subsections. All subjects in FAS are included in the models.

Step 2a: Missing data in Category A are imputed using a linear slope random coefficient model. Subjects in FAS who are classified as having PD, including dropouts due to PD/death and completers whose FVCpp declined ≥10 at Week 48, are included in this model.

Step 2b: If the regression projected decline in Step 2a is higher than the maximum observed decline in the study population, truncate the linear projected value to the maximum observed decline in the main study.

Step 2c: If the projected decline in FVCpp48 in Step 2a is less than 10 (ΔFVCpp48>-10), then assign imputed ΔFVCpp48 a value of -10. For FVC (L), the cutoff point corresponding to ΔFVCpp48 = -10 is determined as (-10) * mean [predicted FVC (L) at baseline in FAS] /100.

In theory, the same MI model used to impute missing data in Category B can be applied to subjects who are classified as PD to impute missing data in Category A. However, due to limited available observed data in the PD subset, the MI model cannot be implemented. The linear slope random coefficient model requires only 2 or more data points per subjects.

The rationales of including steps 2b and 2c are as follows: a large decline of FVC may occur when a subject is experiencing an IPF exacerbation. Such decline usually subsides after the subject's condition is stabilized. Truncating the linear projection in step 2b is more likely to mimic the actual outcome than without truncation. For step 2c, if a subject discontinues treatment due to progressive disease, his or her FVC value is likely to have declined substantially, although only moderate changes are observed prior to the subject's discontinuation.

5.4.1.3 Statistical Models Used in Imputing Missing Data ΔFVCpp48

(a) For subjects in FAS, generate a monotone missing pattern by Monte Carlo Markov Chain (MCMC) procedure:

Model – FVCpp_MI1

PROC MI data=FVCpp_main   nimpute = 100 round = .01 minimum = 30 maximum=130 seed=303019067 out=FVCpp_MI1;

BY trt;

MCMC impute=monotone;

VAR age sex DLCOPP-hb0 FVCpp0 FVCpp12 FVCpp24 FVCpp36 FVCpp48;

run;

Input dataset FVCpp includes FAS subjects except those who were mis-diagnosed as IPF.

(b) Use a monotone MAR-based multiple imputation approach from PROC MI to impute missing data from monotone missing data pattern generated by Model – ΔFVCpp_MI1.

Model – FVCpp_MI2

PROC MI data=fvc_mi1 nimpute = 1 round = .01 minimum =30 maximum = 130 seed=313019067 out=FVCpp_MI2;

BY _imputation_ trt;

VAR age sex DLCOPP-hb0 FVCpp0 FVCpp12 FVCpp24 FVCpp36 FVCpp48;

MONOTONE REGRESSION (FVCpp48 = age sex DLCOPP-hb0 FVCpp0 FVCpp12 FVCpp24 FVCpp36);

run;

(c) For subjects who dropped out due to death or PD and subjects whose FVCpp declined ≥10 at Week 48, fit the following random coefficient model:

Model – Random_coeff

proc mixed data=PD/Deaths;

class subjid;

model chg = visitn/ noint outp=pred_data;

random visitn /subject=subjid type=UN;

run;

If the iterations do not converge, other variance structure will be used. Baseline records with chg = 0 are include in the model.

5.4.1.4 Statistical Models Used in Imputing Missing Data FVC (L)

(d) For subjects in FAS except those who are mis-diagnosed as IPF, generate a monotone missing pattern by Monte Carlo Markov Chain (MCMC) procedure:

Model – MI_FVC1

PROC MI data=fvc_main nimpute = 100 round=.001 minimum=1 maximum=5 seed=203019067 out=fvc_mi1;

BY trt;

MCMC impute=monotone;

VAR age sex DLCOPP-hb0 fvc0 fvc12 fvc24 fvc36 fvc48;

run;

(e) Use a monotone MAR-based multiple imputation approach from PROC MI to impute missing data from monotone missing data pattern generated by Model-FVC1.

Model – MI_FVC2

PROC MI data=fvc_mi1 nimpute = 1 round=.001 minimum=1 maximum=5 seed=213019067 out=fvc_mi2;

BY _imputation_ trt;

VAR age sex DLCOPP-hb0 fvc0 fvc12 fvc24 fvc36 fvc48;

MONOTONE REGRESSION (fvc48 = age sex DLCOPP-hb0 fvc0 fvc12 fvc24 fvc36);

run;

(f) The same random coefficient model (Model Random_coeff) in (c) above is used to obtain the projection coefficient for ΔFVC.

5.4.2 Imputation for Missing HRCT Fibrosis Scores

HRCT fibrosis score is a measure of physiological condition of lung tissue. Unlike functional measures such as FVC, physiological changes are gradual and are in general predictable from the pattern of prior measurements and the pattern of the study population. Based on this understanding, missing HRCT fibrosis score will be imputed by regression model that includes prior assessments as predictive effects. The multiple imputation (MI) procedure, which can impute missing data for subjects with no post-baseline assessment, is an appropriate procedure in this situation.

Since HRCT is a required procedure to evaluate subject eligibility, baseline values are in general available except a few cases of which the quality of the HRCT images is inadequate for quantitative scoring. Missing post-baseline data are mostly due to subjects discontinuing the study. Therefore, missing data at weeks 24 and 48 are in general in a monotone pattern with a few exceptions. A multi-step multiple imputation method, as shown below, will be used to perform missing data imputation as well as data analysis. Subjects with missing baseline value will not be included in the imputation and analysis models. Imputation and analysis procedure will be performed for main study and substudy separately.

5.4.2.1 Main Study

1) Use Monte Carlo Markov Chain (MCMC) methodology from PROC MI by treatment group to generate a monotone missing pattern;

> Model - FIB1
>
> PROC MI data=fib_obs_main nimpute = 100 round=.1 minimum=0 maximum=100 seed=103019067 out=fib_mi1main;
>
> > BY trt;
> >
> > MCMC impute=monotone;
>
> VAR fib0 fib24 fib48;
>
> run;

2) Use a monotone MAR-based multiple imputation approach from PROC MI to impute missing data from monotone missing data pattern generated by Model-FIB1.

> Model - FIB2
>
> PROC MI data=fib_mi1main nimpute = 1 round=.1 minimum=0 maximum=100 seed=113019067 out=fib_mi2main;
>
> BY _imputation_ trt;

VAR fib0 fib24 fib48;

MONOTONE REGRESSION (fib24= fib0);

MONOTONE REGRESSION (fib48= fib0 fib24);

run;

A regression method is preferred over the predictive mean matching method because those who drop out of the study are likely to be sicker than those who complete the study.

If the final missing data pattern is monotone, then Model – FIB1 will not be needed and NIMPUTE in Model – FIB2 will be changed to 100.

Change from baseline at Week 24 is: $\Delta FIB24 = FIB24 - FIB0$;

Change from baseline at Week 48 is: $\Delta FIB48 = FIB48 - FIB0$.

5.4.2.2 Substudy

Since the substudy is relatively small and there is only one post-baseline assessment, the two background therapies are combined in MI procedure in order for it to perform properly. The following MI model will be used for the substudy:

Model – FIB3:

PROC MI data= fib_obs_sub nimpute = 100 round=.1 minimum=0 maximum=100 seed=133019067 out= fib_mi2sub;

BY trt ;

VAR fib0 fib24;

MONOTONE REGRESSION (fib24= fib0);

run;

If at the final data analysis there are not enough observed data points to perform the imputation in Models FIB3, variable trt will be moved from the 'BY' statement to the 'MONOTONE REGRESSION' statement.

5.4.3 Rules for Handling Missing Data in Individual Questions in SGRQ Scoring

Rules of handling missing data in individual questions in SGRQ questionnaire are presented in Appendix III.2. Missing data in the SGRQ domain and total scores will not be imputed. Analysis of these endpoints will be based on observed data using a MMRM model.

5.4.4 Handling Missing/Incomplete AE Onset Date

If the AE onset date is incomplete or missing, the following rules will be applied to obtain imputed AE onset date before applying Table 3 in Section 4.2.12.

- If year and month are present, only day is missing,
  a) If AE onset Year/month = Day 1 Year/month, assign onset date = date of Day 1;
  b) If AE onset Year/month ≠ Day 1 Year/month, assign onset Day = 1;
- If year is present, month and day are missing,

a) If onset year = year of Day 1, assign onset date = date of Day 1;

b) If onset year ≠ year of Day 1, assign January 1st to onset month and day.

- If onset date is completely missing, assign onset date = date of Day 1.

If the stop date is complete and the imputed start date as above is after the stop date, the start date will be imputed by the stop date.

5.4.5 Handling Missing/Incomplete CM Start/Stop Dates

For the purpose of grouping medications in different study period, the following rules will be used to impute incomplete CM start and end date:

- Incomplete CM start date: assign 1 to missing Day, January to missing Month.

- Incomplete CM end date: assign 30 to missing Day, December to missing Month. Impute CM end date only if 'ONGOING' is not checked.

If the imputed stop date is before the start date (imputed or non-imputed start date), then the imputed stop date will be replaced with the start date.

No imputation will be performed for the following cases:

- CM end date will not be imputed if 'ONGOING' is checked.

- Year of CM start or end is missing. If CM end year is missing, the CM will be grouped in the 'Concomitant Medication' category.

5.4.6 Handling Missing/Incomplete Dates for PFT Prior to Enrollment

For the purpose of estimating the rate of change in PFT parameters prior to enrollment, the following rules will be used to impute incomplete PFT performance date:

- If year and month are present, only day is missing, assign day = 15;

- If year or month is missing, then the corresponding PFT assessment is not evaluable for evaluating rate of change.

5.5 Adjustment for Covariates

Comparisons between treatment arms will be adjusted for baseline value of the parameter being analyzed.

5.6 Definition of Baseline

Baseline is defined in Section 4.2 for each parameter separately.

5.7 Efficacy Analysis Visit Window

Efficacy parameters (PFT, HRCT, and PRO) will be summarized by analysis visit defined by the following assessment windows in Tables 6 and 7.

Table 6. Assessments with 12-week interval (PFT and SGRQ)

Analysis Visit	Window
Baseline	Prior to first dose on Day 1

Week X	X = 12, 24, 36, 48
	Target day = 7 * X + 1
	Window = [Target day – 42, Target day + 41]

Table 7.   Year 1 assessments with 24-week interval (HRCT)

Analysis Visit	Window
Baseline	Prior to first dose on Day 1
Week X	X = 24, 48
	Target day = 7 * X + 1
	Window = [Target day – 84, Target day + 83]

All scheduled and unscheduled assessments are included. If multiple FVC assessments are available in the same window, the highest value will be used in the analysis. If multiple HRCT assessments or SGRQ assessments are in the same window, the one closest to the scheduled visit will be used in the analysis. If two assessments with equal distance to the target day, the later assessment will be used in analysis. Assessments that do not fall in any windows are considered not evaluable and will not be included in analyses.

Safety data (lab tests, vital signs, ECG, physical exams) are summarized by nominal clinic visit. Data collected during the unscheduled visits are not included in the summary tables, but are included in data listings and in evaluations for clinically significant abnormal changes.

The upper bound of the Week 48 window is at the earlier time point of Target day +41 (or +83 for HRCT) days or Day 1-EX.

5.8   Interim Analyses and Data Monitoring

Unblinded safety data summaries are performed by an independent vendor and are reviewed an independent Data Monitoring Committee periodically. DMC recommendations are documented in the DMC meeting minutes.

5.9   Pooling Data of Study Sites

All study sites are pooled in all analyses due to the small number of subjects enrolled at each site. Effect of a region, or country, or site will be evaluated as part of exploratory analyses.

5.10 General Layout

All study parameters, including baseline characteristics, efficacy, safety, PK and biomarker, will be summarized descriptively. Descriptive statistics including the number of subjects (n), mean, standard deviation (SD) for baseline variables or standard error (SE) for efficacy parameters, median, minimum and maximum will be presented for continuous variables. For continuous PK parameters, coefficient (CV) and geometric mean may also be presented. Number (n) and percentage (%) of subjects in each category will be summarized for categorical variables.

Efficacy parameters will be summarized analytically. Analytical statistics include LS mean and standard error, 95% CI for the mean or median, p-value, treatment difference expressed in absolute difference, odds ratio, or hazard ratio depending on the nature of the parameter.

All summaries will be presented by treatment group as well as a combined group (FG-3019, Placebo, Overall), unless specified otherwise. Subjects who are enrolled in the substudy will be summarized separately from those who are enrolled in the main study. Only data collected during the Randomized Treatment Period are included in this set of analyses and summaries. Data collected during the extension period will be summarized in a separate package.

Data listings are sorted by patient sub-group: main study, substudy pirfenidone, and substudy nintedanib. Within each sub-group, subjects are sorted by treatment arm (FG-3019 first) and finally by subject ID.

Figures, such as line-chart, bar-chart, box-plot, scatter plot, forest plot, or waterfall plot, are in general included to facilitate comparison between treatment arms and evaluation of trend.

5.11 Data Errata and Hard-Coding

Data errors identified after database lock are documented in an errata log. Under special circumstances, hard-coding in SAS programs is necessary to correct data errors in order to avoid obscure study results. Only limited cases deemed necessary and are approved by study team will be hard-coded. The changed values as well as approval from the study team will be documented.

6 STATISTICAL ANALYSES

6.1 Subject Enrollment and Disposition

The number of subjects enrolled in each study site will be summarized.

The number of subjects in each study population (Randomized, Safety, FAS, Substudy, and Extension) will be summarized by treatment group for main and substudy separately. The number of subjects who completed or discontinued the study as well as the reasons for early discontinuation will be summarized. Treatment completion rate will be compared between the two treatment arms.

Subject who discontinued the study prematurely and the reasons will be listed.

6.2 Protocol Deviations

Protocol deviations will be categorized as follows:

- Entry Deviation: Subject entered study, but did not satisfy eligibility criteria.

- Withdrawal Deviation: Subject met withdrawal criteria during the study but was not withdrawn.

- Dosing Deviation: Subject received the wrong treatment or incorrect dose; including incorrect timing of a dose.

- Prohibited Medication Deviation: Subject received an excluded concomitant treatment.

- Operational Deviation: All other deviations; including, but not limited to: informed consent form-related deviations other than consent not obtained, IRB/IEC approval expired, study drug not stored under protocol-specified conditions, missing laboratory report, out-of window visit, etc.; includes missed visits, and subject refusal of a study procedure or procedures.

  Important deviations are defined as those that are likely to affect a) the safety or physical or mental integrity of the subject, b) the scientific value of the trial.

Deviations are summarized by subject count and by event count, by treatment group and by study site. All recorded protocol deviations are listed. Reported items that are not considered as protocol deviations are not included in summary tables and data listings, but remain in the database for reference.

6.3 Demographic and Baseline Characteristics

Demographics and baseline characteristics will be summarized for subjects in the Safety population using descriptive statistics.

6.3.1 Subject Demographics and Other Baseline Characteristics

Demographic variables will include age in years, gender, race, and ethnicity. Age is defined as the age on the day of signing inform consent:

age = INTCK('YEAR', Birth Date, Date of Informed Consent, 'C')

where INTCK is a SAS function.

Height and weighted collected at Screening are defined as baseline. Screening characteristics include height, weight, body surface area (BSA), body mass index (BMI), IPF confirmation, smoking history, rate of decline in FVC and DLCO prior to enrollment, medications used for treatment of IPF and medical history. Computation formulas:

$$BSA = [Weight^{0.425} (kg) * Height^{0.725} (cm)] \times 0.007184$$

$$BMI = Weight (kg) / (Height (m))^2$$

Time since the first IPF diagnosis (<1 year, 1 – 3 years, > 3years) is derived from of the onset date documented on the Medical History CRF to Day 1. If the day of first IPF diagnosis is missing, substitute it by 15; if month is missing, substitute by July 1st.

6.3.2   Medical History

Medical conditions, including allergies and surgeries, are coded in system organ class (SOC) and preferred term (PT) using MedDRA (version 18.0).

The medical conditions will be tabulated by SOC and PT. A subject with multiple medical conditions within an SOC is only counted once in this SOC. Similarly, a subject with multiple medical conditions within a PT is only counted once in this PT. The tabulation will be sorted alphabetically by SOC and by decreasing order of frequency of PT within each SOC based on the Safety Population.

6.3.3   Baseline Disease Characteristics

Baseline disease characteristics include FVC, DLCO, and other PFT parameters, lung fibrosis scores as well as other HRCT reading scores, and SGRQ scores. These baseline characteristics will be summarized based on the Safety Population.

Baseline assessments for safety evaluations (lab, vital signs, ECG, physical exam) are summarized along with the post-baseline evaluations.

6.3.4   Baseline GAP Score and Staging

The GAP score and staging were developed by Ley et al. (2012) as composite prognostic factors that can be used to predict expected survival in IPF patients. Four variables are included in the GAP score: gender (G), age (A), and 2 lung physiology variables (P) (FVC and DLCO). It is defined as the sum of the points of the 4 variables defined in Table 8. The range is 0 – 8. Staging is defined based on GAP score as shown in Table 9.

Table 8. Scoring of GAP Variables

	Predictor	Points
G (Gender)	Female	0
	Male	1
A (Age)	<=60	0

	61 - 65	1
	>65	2
P (Physiology)	FVC, % predicted	
	>75	0
	50 - 75	1
	< 50	2
	DLCO, % predicted Hb corrected	
	>55	0
	36 -55	1
	<=35*	2
	Cannot perform	3

*Value < 35.5 is rounded down to 35; value>=35.5 is rounded up to 36.

Table 9. GAP Staging

GAP Stage	GAP Score
I	0 – 3
II	4 – 5
III	6 – 8

6.3.5 Annualized Rate of Change in FVC and DLCO Prior to Study Treatment

Annualized rate of change in FVC (L) and DLCO (ml/min/mmHg) (without correction for Hb) prior to study treatment is estimated via linear random coefficient model historic data in the past 2 years as well as values at Screening and Day 1.

```
proc mixed;
  class subjid;
  model chg = time earliest_value  /noint s outp=pred_data;
  random time /subject=subjid type=UN;
run;
```

CHG in the model = Day 1 value - historical value;

Time in years = (Date of Day 1 - Date of historical assessment)/365.25.

All observed values prior to and on Day 1 will be included in the model. If the Day 1 value is missing, the last Screening value will be used to calculate the change. If the day portion of the historical assessment date is missing, day = 15 is assigned, Missing year or month will not be imputed. For multiple values on the same day, the highest value will be used.

6.3.6 Comparison between Treatment Arms in Baseline Parameters

Balance between the two treatment arms in baseline parameters is evaluated using the two sample t-test for continuous variables and the chi-square test for categorical variables. The p-values are used for evaluation of balance of randomization.

6.4 Summary of Prior and Concomitant Medications

The historical use of the medications listed on the Screening CRF is summarized using descriptive statistics for the Safety population.

Medications recorded on the Concomitant Medication CRF are classified in the analysis dataset in the following categories:

1. Medications that were stopped prior to the first infusion (PRIOR).

2. Medications that are used concomitantly with the study drug, which are defined as medications that were not stoped before the first infusion.

Prior and concomitant medications are summarized by ATC class and preferred term for the safety population. Subjects reporting more than one use of the same medication will be counted only once in the summary tables.

All medications captured in screening Medication CRF and Concomitant Medication CRF, as well as Non-Drug Therapies CRF are presented in data listings. Oxygen use will also be presented in data listing.

6.5 Summary of Study Drug Exposure and Treatment Compliance

Duration in days from first dose to last dose, number of doses, dose amount in mg, and compliance as defined in Section 4.2.9 will be summarized descriptively for the Safety population. Proportion of subjects being treated by study week will also be summarized. Study drug administration log will be listed. Dose amount prepared that differs from the intended amount (30 mg/kg times body weight except the first 2 doses which are 15 mg/kg for subjects in the substudy) by more than 10% will be flagged.

6.6 Efficacy Analyses

Analyses of efficacy parameters will be performed on the FAS population. Comparison between FG-3019 and placebo will be performed separately for main study, substudy pirfenidone background and substudy nintedanib background.

A primary analysis will be identified for each key efficacy parameter. Sensitivity analyses using different missing data handling rules and different analysis methods are also performed.

6.6.1 Analyses of the Primary Efficacy Endpoint (ΔFVCpp48)

For the primary analysis of ΔFVCpp48, missing data will be assumed to be MAR. The MAR assumption implies that the missing data can be reasonably predicted from the observed variables, and therefore can be estimated without bias using exclusively the observed data. As noted in the EMA Guideline on Missing Data in Confirmatory Clinical Trials (2011), the MAR assumption implies that the behavior of the post dropout observations can be predicted from the observed variables, and therefore that response can be estimated without bias using exclusively the observed data. For example, when a patient drops out due to lack of efficacy reflected by a series of poor efficacy outcomes that have been observed, it would be appropriate to impute or model poor efficacy outcomes subsequently for this patient. In the current study evaluating treatment effect in the IPF patient population, the random coefficient linear regression model based on the MAR assumption is an acceptable method, "since estimated slope based on available data from an individual patient before treatment discontinuation conservatively predicts the annual decline in lung function similar to linear extrapolation" (Nintedanib Statistical Review, page 42). This method will be defined as the primary analysis of ΔFVCpp48.

6.6.1.1 Analysis of ΔFVCpp48 Using a Random Coefficients Linear Regression Model (Main Study Only)

The random coefficient linear regression model includes treatment, visit (as a continuous variable), visit-by-treatment interaction, and baseline FVCpp as fixed effects and linear slope of visit as random effect. Observed data in all visits (including baseline visit with change = 0) from subjects in FAS will be included in the model. ΔFVCpp48 is compared between the treatment at visit = 48. An annual decline rate is projected at visit = 52. In addition to statistical significance level (p-value) of testing the treatment difference, least squares (LS) means (s.e.) and 95% confidence intervals (CIs) of ΔFVCpp48 for both treatment arms as well as the treatment difference are obtained from the model.

Model – LinearSlope (FVCpp)

proc mixed data = OC-main;

class trt subjid;

model ΔFVCpp = trt visitn trt*visitn FVCpp0 /solution covb;

random visitn /subject=subjid type=UN (option 'group=trt' will be used if it provides a better fit of the model at final analysis);

lsmeans trt / at visitn=48 diff CL;

lsmeans trt / at visitn=52 diff CL;

ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

Subject's age, sex, and height are not included in this model since the FVC (% predicted) value has been adjusted for these factors.

Relative treatment difference is defined as (Active – Placebo) / Placebo * 100.

6.6.1.2 ANCOVA Analysis of ΔFVCpp48 (Main Study only)

This analysis is to use the ANCOVA model to test whether the two treatment arms are different and estimate the treatment difference in ΔFVCpp48, under a different missing data mechnism. All subjects in the FAS population are included; missing data are imputed following the algorithm described in Section 5.4.1. From the ANCOVA model, LS mean (s.e.) of each treatment group as well as LS mean (s.e.) of treatment difference, p-value, and 95% confidence interval will be obtained.

Model – ANCOVA (FVCpp):

proc mixed data=ΔFVCpp48_imputed;

  class trt;

  model ΔFVCpp48 = trt FVCpp0;

  lsmeans trt / pdiff CL alpha=0.05;

run;

6.6.1.3 Analysis of Categorical Endpoints of ΔFVCpp48 (Main Study Only)

The categorical endpoints of ΔFVCpp48, subjects with no decline (change ≥0) and subjects with decline more than 10% (including deaths and disease progression), are defined in Section 4.2.7. These two endpoints will be compared between treatment arms using logistic regression controlling for FVCpp0. Odds ratio and its 95% CI will be estimated from this model.

Model – Logistic

proc logistic data = Event_of_interest;

  class TRT;

  model Event = TRT FVCpp0;

  oddsratio TRT / CL = PL;

run;

Absolute treatment differences in the proportions are also estimated. 95% CIs for the absolute differences are obtained from PROC FREQ using asymptotic approximation. Relative difference is defined as (FG-3019 – Placebo) / Placebo * 100.

6.6.1.4 MMRM Analysis of ΔFVCpp by Visit (Main and Substudy, Observed Cases Only)

Observed data in all post-baseline visits from subjects in FAS will be included in the model. From the MMRM model, LS means (s.e.) of each treatment group, as well as LS mean (s.e.) of treatment difference, p-value, and 95% confidence interval are obtained at each visit. The same MMRM model will be applied to the main study and the 2 subgroups in the substudy separately.

Model – MMRM (FVCpp)

proc mixed data = OC;

class trt subjid visitn;

model ΔFVCpp = trt visitn trt*visitn FVCpp0 FVCpp0*visitn /solution covb;

repeated avisit /subject=subjid type=UN; (option 'group=trt' will be used if it provides a better fit of the model at final analysis)

lsmeans trt*visitn / diff CL alpha=0.05 ;

ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

6.6.1.5   Categorical Summary of ΔFVCpp by Visit (Main and Substudy)

ΔFVCpp will be summarized by visit based on observed cases for subjects in the FAS population, in the following categories: ≤-10, (-10, -5], (-5, 0), [0, 5), [5, 10), and ≥10. Missing data will be included in a separate category.

6.6.1.6   Analysis of ΔFVC48 Using a Random Coefficients Linear Regression Model (Main Study Only)

For analysis of change from baseline in FVC (L), subject's age, sex, and height will be included as fixed effects in the random coefficient model, in addition to other factors specified in the random coefficient model for analysis of ΔFVCpp48.

Model – LinearSlope (FVC)

proc mixed data = OC-main;

class trt subjid;

model ΔFVC = trt visitn trt*visitn FVC0 age sex height /solution covb;

random visitn /subject=subjid type=UN (option 'group=trt' will be used if it provides a better fit of the model at final analysis);

lsmeans trt / at visitn=48 diff CL;

lsmeans trt / at visitn=52 diff CL;

ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

6.6.1.7   ANCOVA Analysis of ΔFVC48 (Main Study only)

For analysis of change from baseline in FVC (L) to week 48 with observed and imputed data, subject's age, sex, and height will be included, in addition to other factors specified in the in the ANCOVA model for analysis of ΔFVCpp48.

Model – ANCOVA (FVC):

proc mixed data=ΔFVC48_imputed;

  class trt;

  model ΔFVC48 = trt FVC0 age sex height;

  lsmeans trt / pdiff CL alpha=0.05;

run;

6.6.1.8   MMRM Analysis of ΔFVC by Visit (Main and Substudy, Observed Cases Only)

Same method described for MMRM analysis of ΔFVCpp will be applied to MMRM analysis of ΔFVC, except that subject's age, sex, and height are additionally added in the model as fixed effects

Model – MMRM (FVC)

proc mixed data = OC;

class trt subjid visitn;

model ΔFVC = trt visitn trt*visitn FVC0 FVC0*visitn age sex height /solution covb;

repeated avisit /subject=subjid type=UN; (option 'group=trt' will be used if it provides a better fit of the model at final analysis)

lsmeans trt*visitn / diff CL alpha=0.05 ;

ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

6.6.1.9   Analysis of Percent Relative Change in FVC (L) by Visit (Main and Substudy)

The percent relative change in FVC (L) will be summarized descriptively in mean (s.e.), median, range, and frequencies in the following categories: <-20%, -20% -< -10%, -10% -< 0, >=0. Observed and imputed values will be used in analysis of

6.6.1.10 Data Listing

Data listing will include observed FVC (L) and FVCpp as well as the change from baseline for all randomized subjects. Imputed data at Week 48 will also be included.

6.6.2   Analyses of Quantative HRCT Fibrosis Score

6.6.2.1   Analysis of Change from Baseline to Week 48 in HRCT Fibrosis Score (Main Study)

Change from baseline to Week 48 in quantitative HRCT fibrosis score (ΔFIB48) is defined in Section 4.2.2. Imputation for missing data is described in Section 5.4.2. The ANCOVA model will be used to analyze treatment difference based on the output datasets from the MI procedures. MIANALYZE procedure will be used to obtain the final estimates of LS mean (s.e.) and 95% CI as well as p-values.

Model – ANCOVA (ΔFIB – main):

proc mixed data= fib_mi2main;

  by _imputation_;

  class trt;

  model ΔFIB48 = trt fib0/solution covb;

  lsmeans trt / diff CL alpha=0.05 ;

  ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

LS mean (s.e.), 95% CI, and p-value of treatment difference are obtained using the following MIANALYZE procedure:

proc mianalyze data=mixdiffs;

    modeleffects estimate;

    stderr stderr;

ods output parameterEstimates = MIest;

run;

LS mean (s.e.) and 95% CI of each treatment are obtained using the similar MIANALYZE procedure based on the MIXLSMEANS output dataset from the ANCOVA model.

6.6.2.2 Responder Analysis of ΔFIB48 (Main Study)

For each imputation from the MI procedure, responders of are defined as ΔFIB48 < 0 (improved) or ΔFIB48 ≤ 0 (improved or stable) based on the observed + imputed data. Proportion of responders and 95% CI for the each treatment arm are obtained from the following PROC FREQ procedure followed by MIANALYZE.

PROC FREQ data = fib_mi2main;

TABLES responder48 / CL binomial;

BY trt _Imputation_;

ODS OUTPUT BINOMIAL=prop;

run;

PROC MIANALYZE DATA=propest (transpose of dataset PROP);

MODELEFFECTS _bin_;

STDERR e_bin;

BY trt;

ODS OUTPUT PARAMETERESTIMATES=prop_trt;

RUN;

Treatment difference and 95% CI are estimated using the MIANALYZE procedure with ModelEffects being (p1 – p2) and StdErr being sqrt (se1^2 + se2^2).

Model – Logistic

proc logistic data= fib_mi2main;

  model responder48 = trt fib0 / covb;

  by _Imputation_;

  ods output ParameterEstimates=fibparms

      CovB=fibcovb;

run;

```
proc mianalyze parms=fibparms
           covb(effectvar=stacking)=fibcovb;
    modeleffects intercept trt  fib0;
run;
```

6.6.2.3   Analysis of ΔFIB24 (Main and SubStudy)

The same MI – ANCOVA – MIANALYZE procedures described in the previous section are applied to analysis of ΔFIB24 in the main study.

For the substudy, the two subgroups are analyzed in the same model with subgroup and treatment-by-subgroup interaction.

<u>Model – ANCOVA (ΔFIB – sub)</u>:

```
proc mixed data= fib_mi2sub;
  by _imputation_;
  class trt  sub;
  model ΔFIB24  = trt  sub  trt*sub  fib0/solution covb;
  lsmeans trt*sub / diff CL alpha=0.05 ;
  ods output lsmeans=mixlsmeans diffs=mixdiffs;
run;
```

Other ΔFIB24 parameters are similarly estimated as the ΔFIB48 parameters.

6.6.2.4   Data Listing of Quantitative Fibrosis Score

The average of the imputed values from Models – FIB2 and – FIB4 are presented in the data listing for each subject.

6.6.3   Analysis of Change from Baseline in HRQoL Score

Change from baseline in SGRQ total and domain scores (ΔTOT, ΔSYM, ΔACT, and ΔIMP) are defined in Section 4.2.3. Missing data data in the total or domain scores are not imputed. Models MMRM (main) and MMRM (sub) described in Section 6.6.1.4 for analysis of the main and substudy, respectively, are used to estimate LS means (s.e.) of each treatment group, as well as LS mean (s.e.) of treatment difference, p-value, and 95% confidence interval at each visit for ΔTOT, ΔSYM, ΔACT, and ΔIMP.

Proportion of subjects with improved SGRQ domain or total score (change <0) will be compared between treatment arms using same logistic regression described in Section 6.6.1.5, controlling for BL_score. Odds ratio and its 95% CI will be estimated from the same model.

6.6.4   Analysis of Time to Progression of IPF

Time to progression (TTP) of IPF is defined in Section 4.2.4. Comparison between treatment arms as well as estimation of hazard ratio and its 95% CI will be performed using Cox regression model. Analysis will be performed on both Safety population and FAS poulation.

Model – Cox

```
proc phreg data = TTP;
  class TRT;
  model TTP * Censor(0) = TRT;
  hazardratio TRT / CL = PL;
run
```

Kaplan-Meier curves are plotted by treatment arm, without adjustment for other factors.

The rate of event at the end of year 1 and the rate of each component are tabulated. The corresponding exact 95% confidence interval will also be presented. Comparison between treatments is performed using logistic regression model based on subjects in the Safety population and FAS population.

Model – Logistic

```
proc logistic data = Event_of_interest;
  class TRT;
  model Event = TRT;
  oddsratio  TRT / CL = PL;
run;
```

Time to event and type of event will be presented in listing.

6.6.5   Analysis of Respiratory-Related Hospitalization and Respiratory-Related Death

Comparison between treatment arms as well as estimation of odds ratio and 95% CI will be performed using logistic regression (Model – Logistic in Section 6.6.4) based on subjects in the Safety population.

6.7   Exploratory Analyses

6.7.1   Sensitivity Analyses of the Primary Efficacy Endpoint

Sensitivity analyses are recommended by regulatory agencies. Different analyses based on different missing data assumptions will be performed.

6.7.1.1   Rank ANCOVA (Ranking Based on Time to Treatment Discontinuation)

Based on the study design, it is reasonable to assume that subjects who discontinued treatment due to death or progressive disease (PD) have worse clinical outcomes than those who are able to complete the study. Time to treatment discontinuation in the current context can be treated as a surrogate of time to clinical progression. Therefore, subjects who discontinued treatment due to death or PD (subset A) should be ranked lower than other subjects (Subset B), with ranking based on time to discontinuation and the lowest rank assigning to the earliest dropout. Time to treatment discontinuation is recorded on the EOT Disposition CRF. Subjects in subset B are ranked based on ΔFVCpp48, starting from adding 1 to the highest rank in subset A. Imputed data described in Section 5.4.1, Step 1, are included in ranking subjects in subset B, assuming missing

data in this subset is missing at random. Ranks for tied observations will be assigned based on the average of the corresponding ranks if there had been no ties.

The rank ANCOVA procedure is described by Stokes, Davis, and Koch (2012). The first step of the procedure is to fit a linear regression of the rank of the primary endpoint described above against the rank of baseline FVCpp. The residuals from this model are compared between treatment arms using the Mantel-Haenszel mean score chi-square test. The two SAS procedures are provided below.

PROC REG data=ranks noprint;

model R_ ΔFVCpp48 = R_ FVCpp0;

output out=res

      r=resid;

run;

PROC FREQ data=res;

tables TRT * resid/noprint cmh2;

run;

6.7.1.2 ANCOVA with Rank-based Imputation

The above rank ANCOVA procedure is used to evaluate qualitatively whether there is treatment difference between the two treatment groups. To evaluate quantitatively the magnitude of the treatment difference, missing data need to be imputed with quantative values. Following the ranking rationale described in the previous section, missing data in subset A can be imputed by adding additional declined value on top of the lowest ΔFVCpp48 in subset B. The amount of additional decline value is determined by the rank – the lower the rank, the larger decline value is added. The maximum additional decline (corresponding to the subject with the lowest rank) can be determined by clinical judgment. For this sensitivity analysis, an additional 10 points of decline (-10) is added to the earliest dropout from the lowest ΔFVCpp48 in subset B. Evenly distributed increment is added to ΔFVCpp48 for each increment of rank in subjects in subset A. The ANCOVA Model described in Section 6.6.1.1 is applied to rank-based imputed dataset.

6.7.1.3 ANCOVA with MI Imputation

In this approach there is no differentiation of subsets A and B. Missing values are imputed following MAR assumption for all subjects in FAS, using the MI models in Section 5.4.1.3. Model ANCOVA is used in conjunction with MYANALYZE to evaluate treatment difference. Based on the output dataset FVCpp_MI2 from Model in Section 5.4.1.3, the following ANCOVA model and MIANALYZE procedure are used.

Model – ANCOVA_MI

proc mixed data=FVCpp_MI2;

class trt;

model ΔFVCpp48 = trt FVCpp0 /solution covb;

lsmeans trt / diff CL;

by _Imputation_;

ods output lsmeans=mixlsmeans diffs=mixdiffs;

run;

Treatment difference in ΔFVCpp48 is presented in LS mean (s.e.) and 95% CI using the following MIANALYZE procedure:

proc mianalyze data=mixdiffs;

    modeleffects estimate;

    stderr stderr;

ods output parameterEstimates = MIest;

run;

LS mean (s.e.) and 95% CI of ΔFVCpp48 in each treatment arm are obtained from the following MIANALYZE procedure:

proc mianalyze data=mixlsmeans;

    by trt;

    modeleffects estimate;

    stderr stderr;

  run;

Categorical analyses for categories of 'no decline' and 'decline 10 or more' using MI/MIANALYZE procedure to handle missing data will also be performed. Detailed procedure is the same as that described in Section 6.6.2.2.

6.7.1.4 ANCOVA with All Randomized Subjects

The same imputation algorithm described in Section 5.4.1 is applied to subjects in the FAS population. For subjects who are not in the FAS population, missing data are imputed as follows:

- For subjects who discontinued due to reasons other than death or PD, their missing data are imputed following Step 1 in Section 5.4.1, with all randomized subjects in the MI model;

- For subjects who discontinued due to death or PD, their missing data are imputed as the largest observed decline of FVCpp48.

Same Model – ANCOVA in Section 6.6.1.1 is used in analyzing treatment effect in the all randomized population.

6.7.2 Analyses of Other PFT Parameters

Other PFT parameters include change from baseline to scheduled visits in FEV1 (% predicted) and FEV1/FVC ratio. The MMRM (main) and MMRM (sub) models described in Section 6.6.1.4 will be used to obtain LS means (s.e.) of each treatment group, as well as LS mean (s.e.) of treatment difference, p-value, and 95% confidence interval at each visit.

6.7.3   Analyses of Other Quantitative HRCT Parameters

Other quantitative HRCT parameters include change from baseline to Weeks 24 or 48 in fibrosis score at the worst lung zone, the QILD score, GG, and HC scores.

The same analysis procedure described in Section 6.6.2 will be applied to analyzing change from baseline in fibrosis score at the worst lung zone and the QILD score. For GG and HC, descriptive summary based on observed cases will be provided at each visit.

6.7.4   Analysis of HRCT Visual Reading Scores

Baseline visual assessments will be summarized by lung lobe. A baseline whole lung score is defined as the average of the mid-point of the five lung lobes. Mid-points are: 0= none, 3 = 1-5, 15 = 6-25, 38 = 26-50, 63 = 51-75, 87 = '>75'. The average whole lung score (rounded to whole number) is summarized in the same categories as the individual lung lobes.

The whole lung average score in the follow-up visits is defined as the average of the five lung lobes. Categorical assessments at the follow-up visits are coded as: 1 = better, 0 = same; -1 = worse. The average score will be re-mapped to 'better' if average is positive, 'same' if average is zero, and 'worse' if average is negative. Responders will be defined based on the derived whole lung score

1. Better,
2. Better or no change.

The proportion of responders will be summarized at Weeks 24 and 48 for subjects who are in the FAS population. Analysis will be based on observed data ignoring missing data.

HRCT visual readings at each lung lobe are presented in a data listing.

6.7.5   Examination of Subgroups (Main Study)

Treatment effect on ΔFVCpp48 and ΔFIB48 will be examined in the following subgroups:

Baseline GAP staging (I vs. II+III)

Age (<65 vs. >=65)

Sex (male vs. female)

US sites versus ex-US sites

Time since IPF diagnosis (< 1 year vs. ≥1 year)

Baseline FVCpp (below vs. ≥ median)

Baseline FIB (below vs. ≥ median)

Baseline DLCOpp-hb (below vs. ≥ median)

Baseline FEV1/FVC (below vs. ≥ median)

Smoking status (never smoked vs. current/former).

The following ANCOVA model will be used to evaluate the effect of each subgroup individually.

Model – ANCOVA (subgroup):

proc mixed;

  class trt subgroup;

  model endpt = trt subgroup trt * subgroup BL;

  lsmeans trt * subgroup / pdiff CL alpha=0.05;

run;

The dependent variable endpt is ΔFVCpp48 and ΔFIB48; BL is the baseline value FVCpp0 or FIB0 corresponding to the dependent variable. When evaluating the baseline FVCpp subgroups, FVCpp0 will not be included in the model; same for evaluating baseline FIB subgroups, FIB0 will not be included. Subjects in FAS in the main study will be included in the model. Observed and imputed values are used. For ΔFIB48, the average of the imputed values from the MI procedure is used. (MIANALYZE is not used.)

Forest plot of the LS mean +/- standard error of the treatment difference in each subgroup will be presented.

6.7.6   Other Exploratory Analyses

Analysis plan for further exploratory analyses will be documented in a separate document.

6.8   Safety Analyses

Safety analyses will include summary of adverse events (including treatment emergent AEs, treatment emergent serious AEs, deaths, and infusion-associated AEs), prior and concomitant medication use, lab test results, vital signs, ECGs, and physical exams. In general, safety data will only be summarized descriptively and no rigorous inferential statistical procedures will be applied.

6.8.1   Adverse Events

All reported AEs are presented in listings. Treatment emergent adverse events (TEAEs), as defined in Section 4.2.13, are summarized by treatment arm, system organ class (SOC) and preferred term (PT).

An overall summary of TEAEs is provided. The summary includes the number (%) of subjects with at least one TEAE, at least one serious TEAE, deaths, TEAEs leading to study discontinuation, at least one TEAE related to study drug and TEAEs by severity.

The number and percent of subjects experiencing TEAEs will be tabulated by system organ class and preferred term and will be sorted alphabetically by SOC and by decreasing order of frequency of PT within each SOC. The decreasing order of frequency of preferred terms will be based on the overall population. A subject with multiple adverse events within a system organ class is only counted once in this system organ class. Similarly, a subject with multiple adverse events within a preferred term is only counted once in this preferred term.

TEAE will also be tabulated by severity grade. In the case of multiple events within the same preferred term, the event with the highest severity grade is included. Missing data in the severity grade will not be imputed and is ranked the lowest severity grade in the case of multiple events.

In addition to summary of all TEAEs, the following subgroups of AEs are summarized and listed separately:

- SAEs

- SAEs that are possibly related to FG-3019 (as assessed by investigators)

- SAEs that are IPF related or respiratory-related hospitalization

- AEs leading to study or treatment discontinuation

- TEAE with incidence >10% in either treatment arm within the main study or within each sub-group in the substudy

- AEs with severity grade >= 3

- AEs that are possibly related to FG-3019 (as assessed by investigators)

- AEs with severity grade >=3 that are possibly related to FG-3019 (as assessed by investigators)

- Possible infusion-associated AEs

The following events are presented in data listings.

- Deaths

- AEs that are possibly related to study procedure (as assessed by investigators)

- Possible infusion reactions

All reported adverse events will be presented in a data listing which will also include reference information such as the time (days) from the last prior infusion and from the first infusion to the onset of the event.

6.8.2 Mortality

Year 1 mortality rate will be summarized by GAP staging at baseline.

Respiratory-related and IPF related death will be identified in the data listing.

6.8.3 Summary of Central Laboratory Data

Central laboratory results and change from baseline are summarized descriptively by visit.

CTCAE grade 3 or higher lab test results will be considered potentially clinically significant. These results are presented in a data listing.

Shift tables to summarize changes from baseline to each visit in CTCAE categories are tabulated. Shift from baseline to most severe CTCAE category during the study is also summarized. In addition, shift-tables based on normal ranges provided by the central lab are also included.

Box-plots for selected lab tests by visit are presented to evaluate trend.

Lab data are summarized by nominal clinic visit. Data collected during the unscheduled visits are not included in the summary tables, but are included in data listings and in evaluations for potentially clinically significant abnormal changes.

6.8.4 Vital Signs

Vital sign observed values and change from baseline are summarized using descriptive statistics by cohort by visit for the Safety population. Pre-infusion and post-infusion assessments are summarized separately. In addition to summarizing the measurements by time point, the average measurement after the start of an infusion is defined as the post-infusion measurement.

In addition, potentially clinically significant changes from baseline are tabulated for each visit. Reference range limits for vital signs are provided in Table 5a.

All measurements and change from baseline are presented in data listings. Potential clinically significant changes from baseline are flagged.

Box-plots of vital sign measurements by visit, pre- and post-infusion separately, are presented to evaluate trend.

6.8.5 ECG Data

ECG findings are summarized in shift tables that include cross-tabulation of 'Normal', 'Abnormal NCS', 'Abnormal CS' at Weeks 24 or 48 versus Screening.

Detailed findings are presented in the data listing.

ECG parameters measured at each visit as well as change from baseline are summarized descriptively.

6.8.6 Physical Exams

The number and percent of subjects with "Normal", "Abnormal", "Not Done", and "Missing" physical examination results are summarized by body system, cohort, and visit. Shift table of changes from baseline will also be summarized by visit.

The data listing includes abnormal physical examination data.

6.9 PK, CTGF, HAHA, and Biomarkers

PK parameters defined in Sections 4.2.10 and 4.2.11 will be summarized descriptively at each sampling time. Parameters include peak and trough level of FG-3019, pirfenidone, and nintedanib as well as metabolite-to-parent ratio at the peak level of pirfenidone and nintedanib. Descriptive statistics include arithmetic mean (S.D.), 95% CI of the mean, CV, geometric mean, median, 25th and 75th percentiles, minimum and maximum.

Comparison of pirfenidone / nintedanib PK when given with FG-3019 vs placebo will be performed.

Analysis of CTGF, HAHA, DNA and other biomarker data will be documented in a separate report.

6.10 Changes from the Protocol Specified Analysis Methods

6.10.1 Change of the Primary Analysis Method for ΔFVCpp48

In the protocol rank ANCOVA is specified as the primary analysis method to test the hypothesis of whether FG-3019 is superior to placebo in the primary endpoint. This method was determined prior to FDA's marketing approval of pirfenidone and nintedanib. In FDA's statistical review, it was pointed out that the random coefficient linear regression model assuming missing data to be MAR is an acceptable method, "*since estimated slope based on available data from an individual patient before treatment discontinuation conservatively predicts the annual decline in lung function similar to linear extrapolation*" (Nintedanib Statistical Review, page 42). Based on this reason, the primary analysis is changed rank ANCOVA to random linear coefficient model. Rank ANCOVA will be performed as one of the sensitivity analyses to support the primary analysis.

6.10.2 Change in the Analysis Population for HRCT Parameters

In the protocol HRCT parameters are to be analyzed using the HRCT Evaluable population, which is defined as randomized subjects who have evaluable HRCT fibrosis scores at baseline and at Week 24 or at a later time point. Since the MI procedure can impute missing data for subjects who do not have post-baseline data when the fraction of missing data is relatively small comparing to the available, the MI procedure is used to analyze the HRCT fibrosis score based on FAS, which is a larger set than the HRCT Evaluable population and is the same population for analyzing other efficacy parameters.

6.10.3 Definition of Final Analysis of the Randomized Treatment Period

In the protocol, unblinded analysis after all subjects have completed the Randomized Treatment Period is defined as an interim analysis of the entire study including the Extended Treatment Period. It should be noted that, after all subjects (main and substudy) have completed the Randomized Treatment Period, the database for the Randomized Treatment Period portion will be locked and all analyses described in this document based on the locked database are considered final analyses for the Randomized Treatment Period.

7   VALIDATION AND QUALITY ASSURANCE

All datasets (SDTM, ADaM), tables, listings, figures are programmed by two programmers independently. The results must be 100% match.

Both primary and validation programmers will develop programs independently based on the specifications and/or SAP. If the outputs are datasets, the final outputs are compared. If the outputs are TLFs, benchmark results are generated and compared. The validation findings and resolutions are documented on a Validation Worksheet.

The detail process for validation and quality assurance is documented in the Standard Operating Procedures of Q2, Inc.

8 REFERENCES

EMA 2011, Guideline on Missing Data in Confirmatory Clinical Trials.

Fingerlin TE, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013 June ; 45(6): 613–620. doi:10.1038/ng.2609.

Goldman HI, Becklake MR. Respiratory function tests: normal values at median altitudes and the prediction of normal results. Am Rev Tuberculosis 1959; 79:457-471.

Hankinson JL, Odencrantz JR, Fedan KB, Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999; 159:179 – 187.

Jones P. St. George's Respiratory Questionnaire Manual, Version 2.3. St. George's, University of London, June 2009.

King TE, et al. A Phase 3 Trial of Pirfenidone in Patients with Idiopathic Pulmonary Fibrosis. (2014) N Engl J Med

Kupferberg DH, Kaplan RM, Slymen DJ, Ries AL. Minimal Clinically Important Difference for the UCSD Shortness of Breath Questionnaire. J Cardiopulm Rehabil. 2005, 25(6):370-7.

Ley B., et al. (2012) A Multidimensional Index and Staging System for Idiopathic Pulmonary Fibrosis. Ann Intern Med. 2012;156:684-691.

Little RJ (2008) Selection and Pattern-Mixture Models. Chapter 18 in Advances in Longitudinal Data Analysis, G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs, eds., pp. 409-431, London: CRC Press.

Mallinckrod CH, et al. (2008) Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Therapeutic Innovation & Regulatory Science, Vol 42, Issue 4, pp. 303 - 31

Maruish ME (Ed.). *User's Manual for the SF-36v2 Health Survey (3rd Edition).* Lincoln, RI: QualityMetric Incorporated.

Miller A, Thornton JC, Warshaw R, et al. Single breath diffusing capacity in a representative sample of the population of Michigan, a large industrial state. Predicted values, lower limits of normal, and frequencies of abnormality by smoking history. Am Rev Respir Dis 1983; 127: 270 – 277.

Nintedanib FDA Statistical Review for Basis of Approval (2015).
Pirfenidone FDA Statistical Review for Basis of Approval (2015).

Richeldi, L, et al. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis (2014). N Engl J Med

SF-36v2 Administration Guide, version 2009, provided by QualifyMetric Inc.

SF-36v2 Manual and Interpretation Guide (Ware et al., 1993).

Stokes M, Davis C, Koch G, *Categorical Data Analysis Using SAS*, third edition, 2012.

Swigris JJ, et al. The UCSD shortness of breath questionnaire has longitudinal construct validity in idiopathic pulmonary fibrosis. Respiratory Medicine (2012) 106, 1447 - 1455

Swigris JJ, et al. The SF-36 and SGRQ: validity and first look at minimum important differences in IPF. Respir Med. 2010 February ; 104(2): 296–304.

Quanjer PH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur Respir J 2012; 40: 1324–1343.

Witek TJ, Mahler DA. Minimal important difference of the transition dyspnoea index in a multinational clinical trial. Eur Respir J 2003, 21: 267–272)

Toxicity Grade Report for Set: CTCAE ver 4.03, ICON Central Laboratories, Inc.

9 PLANNED TABLES, LISTINGS, AND FIGURES

9.1 List of Tables and Figures

TLF Number	Title	Analysis Population
14.1	**Demographic Data Summary figures and tables**	
14.1-1	Subject Enrollment by Site	Randomized
14.1-2	Subject Disposition	Randomized
Figure 14.1-1	Time to Early Withdrawal from Study Treatment	Randomized
14.1-3	Subject Demographics and Subject Characteristics	Safety
14.1.4-1	Baseline Data: Pulmonary Function Test	Safety
14.1.4-2	Baseline Gap Score and Staging	Safety
14.1.4-3	Annualized Rate of Change in FVC and DLCO Prior to Study Treatment	Safety
14.1-5	Baseline Data: HRCT CAD Reading	Safety
14.1.5-1	Baseline Visual Fibrosis Assessment by Lung Zone	Safety
14.1.5-2	Baseline Visual Ground Glass Assessment by Lung Zone	Safety
14.1.5-3	Baseline Visual Honeycomb Assessment by Lung Zone	Safety
14.1-6	Baseline Data: SGRQ Domain and Total Scores	Safety
14.1-7	Medical History	Safety
14.1.8-1a	Medications for IPF Used and Stopped 4 Weeks Prior to Screening	Safety
14.1.8-1b	Medications for IPF Continued Into Study	Safety
14.1.8-2	Medications Stopped Prior to the First Infusion	Safety
14.1.8-3	Concomitant Medications Used After the First Infusion	Safety
14.1.8-4	Oxygen Use	Safety
14.1.9-1a	Summary of Protocol Deviations by Event and by Site	Safety
14.1.9-1b	Summary of Protocol Deviations by Event and by Treatment	Safety

14.1.9-2a	Summary of Protocol Deviations by Subject and by Site	Safety
14.1.9-2b	Summary of Protocol Deviations by Subject and by Treatment	Safety
14.2	Efficacy Data Summary Figures and Tables	
14.2.1	Analysis of Change from Baseline in FVC (% predicted) and FVC (L)	
T14.2.1-1	Analysis of Change from Baseline to Week 48 in FVC (% predicted): Random Coefficients Linear Regression Model based on Observed Cases	FAS - Main
T14.2.1-2	Analysis of Change from Baseline to Week 48 in FVC (% predicted): ANCOVA based on Observed and Imputed values	FAS - Main
T14.2.1-3	Analysis of Categorical Endpoints of Change from Baseline to Week 48 in FVC (% predicted): Classification based on Observed and Imputed Values	FAS-Main
T14.2.1-4	Analysis of Change from Baseline in FVC (% predicted) by Visit: MMRM based on Observed Cases	FAS
T14.2.1-5	Categorical Summary of Change from Baseline in FVC (% predicted) by Visit: Classification based on Observed Cases Only	FAS
F14.2.1-1	Analysis of Change from Baseline to Week 48 in FVC (% predicted): Random Coefficient Model based on Observed Cases	FAS - Main
F14.2.1-2	LS Mean (s.e.) of Change from Baseline to Week 48 in FVC (% predicted): ANCOVA based on Observed and Imputed Values	FAS - Main
F14.2.1-3	Analysis of Categorical Endpoints of Change from Baseline to Week 48 in FVC (% predicted): Classification based on Observed and Imputed Values	FAS - Main
F14.2.1-4	Cummulative Distribution of Change from Baseline to Week 48 in FVC (% predicted) (Observed and Imputed Values Included)	FAS - Main
F14.2.1-5	LS Mean (s.e.) of Change from Baseline in FVC (% predicted) by Visit: MMRM based on Observed Cases	FAS
T14.2.1-11	Analysis of Change from Baseline to Week 48 in FVC (L): Random Coefficients Linear Regression Model	FAS - Main

	based on Observed Cases	
T14.2.1-12	Analysis of Change from Baseline to Week 48 in FVC (L): ANCOVA based on Observed and Imputed values	FAS - Main
T14.2.1-13	Analysis of Change from Baseline in FVC (L) by Visit: MMRM based on Observed Cases	FAS
F14.2.1-11	Linear Slope Analysis of Rate of Decline in FVC (L): Random Coefficient Model based on Observed Cases	FAS - Main
F14.2.1-12	LS Mean (s.e.) of Change from Baseline to Week 48 in FVC (L): ANCOVA based on Observed and Imputed values	FAS - Main
F14.2.1-13	Cummulative Distribution of Change from Baseline to Week 48 in FVC (L) (Observed and Imputed Values Included)	FAS - Main
F14.2.1-14	LS Mean (s.e.) of Change from Baseline in FVC (L) by Visit: MMRM based on Observed Cases	FAS
14.2.1-21	Descriptive Summary of Percent Relative Change from Baseline in FVC (L) by Visit (Observed and Imputed Values Included)	FAS
F14.2.1-21	Cumulative Distribution of Percent Relative Change from Baseline to Week 48 in FVC (L) (Observed and Imputed Values Included)	FAS - Main
14.2.2	Analysis of Change in Quantitative HRCT Fibrosis Score	
T14.2.2-1	Analysis of Change from Baseline to Week 48 in Quantitative HRCT Fibrosis Score: ANCOVA with MI/MIANALYZE	FAS – Main
T14.2.2-2	Analysis of Change from Baseline to Week 24 in Quantitative HRCT Fibrosis Score: ANCOVA with MI/MIANALYZE	FAS
F14.2.2-1	LS Mean (s.e.) of Change from Baseline in Quantitative HRCT Fibrosis Score by Visit: ANCOVA with MI/MIANALYZE	FAS
F14.2.2-2	% of Subjects with Improved Quantitative HRCT CAD Fibrosis Score by Visit: ANCOVA with MI/MIANALYZE	FAS
F14.2.2-3	Waterfall Plot of Change from Baseline in Quantitative HRCT Fibrosis Score by Visit (Observed and Imputed Values Included)	FAS

F14.2.2-4	Cummulative Distribution of Change from Baseline to Week 48 in HRCT Fibrosis Score (Observed and Imputed Values Included)	FAS - Main
14.2.3	Analysis of Change from Baseline in SGRQ Domain and Total Scores	
T14.2.3-1	Analysis of Change from Baseline in SGRQ Domain and Total Scores by Visit: MMRM based on Observed Cases	FAS
T14.2.3-2	Summary of Proportion of Subjects with Improved SGRQ Domain and Total Scores: Classification based on Observed Cases	FAS
F14.2.3-1	LS Mean (s.e.) of Change from Baseline in SGRQ Domain and Total Scores by Visit: MMRM Model on Observed Cases	FAS
14.2.4	Analysis of Progression of IPF	
14.2.4a	Analysis of Time to Progression of IPF, Respiratory-Related Hospitalization, and Respiratory-Related Death	FAS
14.2.4b	Analysis of Time to Progression of IPF, Respiratory-Related Hospitalization, and Respiratory-Related Death	Safety
F14.2.4-1a	Kaplan-Meier Curve of Time to Progression of IPF	FAS
F14.2.4-1b	Kaplan-Meier Curve of Time to Progression of IPF	Safety
14.2.10	Sensitivity Analyses of the Primary Endpoint	
14.2.10-1	Sensitivity Analysis of Change from Baseline to Week 48 in FVC (% predicted): Rank ANCOVA	FAS – Main
14.2.10-2	Sensitivity Analysis of Change from Baseline to Week 48 in FVC (% predicted): ANCOVA with Rank-based Imputation	FAS – Main
14.2.10-3	Sensitivity Analysis of Change from Baseline to Week 48 in FVC (% predicted): ANCOVA with MI/MIANALYZE	FAS – Main
14.2.10-4	Sensitivity Analysis of Change from Baseline to Week 48 in FVC (% predicted): ANCOVA with All Randomized Subjects	Safety – Main
14.2.11	Additional PFT Parameters	

T14.2.11-1	Analysis of Change from Baseline in FEV1 (% predicted) by Visit: MMRM based on Observed Cases	FAS
F14.2.11-1	LS Mean (s.e.) of Change from Baseline in FEV1 (% predicted) by Visit: MMRM based on Observed Cases	FAS
T14.2.11-2	Analysis of Change from Baseline in FEV1/FVC (%) by Visit: MMRM based on Observed Cases	FAS
F14.2.11-2	LS Mean (s.e.) of Change from Baseline in FEV1/FVC (%) by Visit: MMRM based on Observed Cases	FAS
14.2.12	Additional Quantitative HRCT Parameters	
T14.2.12-1	Analysis of Change from Baseline in Quantitative HRCT Fibrosis Score at the Worst Lung Zone by Visit: ANCOVA with MI/MIANALYZE	FAS
F14.2.12-1	Analysis of Change from Baseline in Quantitative HRCT Fibrosis Score at the Worst Lung Zone by Visit: ANCOVA with MI/MIANALYZE	FAS
T14.2.12-2	LS Mean (s.e.) of Change from Baseline in Quantitative HRCT QILD Score by Visit: ANCOVA with MI/MIANALYZE	FAS
F14.2.12-2	LS Mean (s.e.) of Change from Baseline in Quantitative HRCT QILD Score by Visit: ANCOVA with MI/MIANALYZE	FAS
T14.2.12-3	Analysis of Change from Baseline in Quantitative HRCT Ground Glass Score by Visit: MMRM based on Observed Cases	FAS
T14.2.12-4	Analysis of Change from Baseline in Quantitative HRCT Honeycomb Score by Visit: MMRM based on Observed Cases	FAS
14.2.13	Summary of HRCT Visual Reading Scores	
14.2.13-1	Change from Baseline in Visual Fibrosis Assessment by Lung Zone and Visit	FAS
14.2.13-2	Change from Baseline in Visual Ground Glass Assessment by Lung Zone and Visit	FAS
14.2.13-3	Change from Baseline in Visual Honeycomb Assessment by Lung Zone and Visit	FAS
F14.2.13-4	% of Subjects with Improved HRCT Fibrosis Visual Score in Whole Lung by Visit	FAS

14.2.14	Examination of Subgroups	
14.2.14.1	Analysis of ΔFVCpp48 by Subgroup: ANCOVA based on Observed and Imputed values	FAS - Main
14.2.14.2	Analysis of ΔFIB48 by Subgroup: ANCOVA based on Observed and Imputed values	FAS - Main
F14.2.14.1	Forest Plot of LS Means (s.e.) of ΔFVCpp48 by Subgroup: ANCOVA based on Observed and Imputed values	FAS - Main
F14.2.14.2	Forest Plot of LS Means (s.e.) of ΔFIB48 by Subgroup: ANCOVA based on Observed and Imputed values	FAS - Main
14.2.15	Other Exploratory Analyses	
14.2.15.1	Pearson Correlation between Change from Baseline to Week 48 in FVC (% predicted) and HRCT Fibrosis Score (Observed + MI Imputed for Both Measures)	FAS
14.2.15.2	Pearson Correlation between Change from Baseline to Week 48 in FVC (% predicted) and SGRQ Domain/Total Scores (Observed Cases Only)	FAS
14.3.1	Summary of Adverse Events	
14.3.1-1	Overall Summary of Treatment Emergent Adverse Events	Safety
14.3.1-2	Summary of Treatment Emergent Adverse Events by System Organ Class and Preferred Term	Safety
14.3.1-2a	Summary of Treatment-Emergent Adverse Event Incidence > 10% by System Organ Class and Preferred Term	Safety
14.3.1-3	Summary of Treatment-Emergent Adverse Events Leading to Treatment/Study Discontinuation by System Organ Class and Preferred Term	Safety
14.3.1-4	Summary of Treatment Emergent Adverse Events by System Organ Class, Preferred Term, and Maximum Severity	Safety
14.3.1-5	Summary of Treatment Emergent Adverse Events with Severity Grade ≥ 3 by System Organ Class and Preferred Term	Safety

14.3.1-6	Summary of Treatment Emergent Adverse Events Assessed as At Least Possibly Related to FG-3019/Placebo by System Organ Class and Preferred Term	Safety
14.3.1-6a	Summary of Treatment Emergent Adverse Events Assessed as At Least Possibly Related to FG-3019/Placebo or Background Therapies by System Organ Class and Preferred Term	Substudy Safety
14.3.1-7	Summary of Infusion-Associated Treatment Emergent Adverse Events by System Organ Class and Preferred Term	Safety
14.3.2-1	Listing of Deaths	Safety
14.3.2-1a	Summary of Mortality Rate	Safety
14.3.2-1b	Summary of Mortality Rate by Baseline GAP Staging	Safety
14.3.2-2a	Listing of Serious Adverse Events	Safety
14.3.2-2b	Summary of Treatment Emergent Serious Adverse Events by System Organ Class and Preferred Term	Safety
14.3.2-3a	Listing of Serious Adverse Events Assessed as At Least Possibly Related to FG-3019/Placebo	Safety
14.3.2-3b	Summary of Treatment Emergent Serious Adverse Events as At Least Possibly Related to FG-3019/Placebo by System Organ Class and Preferred Term	Safety
14.3.2-4a	Listing of IPF Related or Respiratory Related Hospitalization Serious Adverse Events	Safety
14.3.2-4b	Summary of Treatment Emergent IPF Related or Respiratory Related Hospitalization Serious Adverse Events by System Organ Class and Preferred Term	Safety
14.3.2-4c	Summary of Treatment Emergent Serious Adverse Events Not Related to IPF or Respiratory Related Hospitalization by System Organ Class and Preferred Term	Safety
14.3.2-5a	Listing of Serious Adverse Events Assessed as At Least Possibly Related to Background Therapy	Safety
14.3.2-5b	Summary of Treatment Emergent Serious Adverse Events Assessed as At Least Possibly Related to Background Therapy by System Organ Class and Preferred Term	Safety
14.3.3	Narratives (not part of TLFs)	

14.3.4	Listing of Clinically Significant Abnormal Laboratory Values (CTCAE Grade >= 3)	Safety
F14.3.4	Lab Test Over Time for Subjects with Clinically Significant Abnormal Laboratory Values (CTCAE Grade >= 3)	Safety
14.3.5.1-1	Laboratory Value and Change from Baseline in Complete Blood Count	Safety
14.3.5.1-2	Laboratory Value and Change from Baseline in Complete Blood Count: Absolute	Safety
14.3.5.1-3	Laboratory Value and Change from Baseline in Complete Blood Count: WBC Differential	Safety
14.3.5.1-4	Summary of CTCAE Grades by Visit in CBC Tests	Safety
14.3.5.1-5	Shift Table Summary of CTCAE Grade in Baseline versus Post-baseline Visits in CBC Tests	Safety
14.3.5.1-6	Shift Table Summary of CTCAE Grade in Baseline versus Maximum Grade During Study in CBC Tests	Safety
14.3.5.2-1	Laboratory Value and Change from Baseline in Serum Chemistry Results: Liver Function Tests	Safety
14.3.5.2-2	Laboratory Value and Change from Baseline in Serum Chemistry Results: Kidney Function and Metabolic Tests	Safety
14.3.5.2-3	Laboratory Value and Change from Baseline in Serum Chemistry Results: Coagulation	Safety
14.3.5.2-4	Laboratory Value and Change from Baseline in Serum Chemistry Results: Electrolytes	Safety
14.3.5.2-5	Laboratory Value and Change from Baseline in Other Serum Chemistry Results	Safety
14.3.5.1-6	Summary of CTCAE Grades by Visit in Serum Chemistry Tests	Safety
14.3.5.1-7	Shift Table Summary of CTCAE Grade in Baseline versus Post-baseline Visits in Serum Chemistry Tests	Safety
14.3.5.1-8	Shift Table Summary of CTCAE Grade in Baseline versus Maximum Grade During Study in Serum Chemistry Tests	Safety

14.3.5.3-1	Shift Table Summary of Complete Blood Count	Safety
14.3.5.3-2	Shift Table Summary of Complete Blood Count: Absolute	Safety
14.3.5.3-3	Shift Table Summary of Complete Blood Count: WBC Differential	Safety
14.3.5.4-1	Shift Table Summary of Serum Chemistry Results: Liver Function Test	Safety
14.3.5.4-2	Shift Table Summary of Serum Chemistry Results: Kidney Function and Metabolic	Safety
14.3.5.4-3	Shift Table Summary of Serum Chemistry Results: Coagulation	Safety
14.3.5.4-4	Shift Table Summary of Serum Chemistry Results: Electrolytes	Safety
14.3.5.4-5	Shift Table Summary of Other Serum Chemistry Results	Safety
F14.3.5-1a	Box Plot of CBC Tests by Treatment and Visit	Safety
F14.3.5-1	Box Plot of Change from Baseline in CBC Tests by Treatment and Visit	Safety
F14.3.5-2a	Box Plot of Serum Chemistry Tests by Treatment and Visit	Safety
F14.3.5-2	Box Plot of Change from Baseline in Serum Chemistry Tests by Treatment and Visit	Safety
14.3.6	Vital Signs	
14.3.6-1	Vital Sign Measurements and Change from Baseline	Safety
14.3.6-1a	Vital Sign Measurements and Change from Baseline by Time Point	Safety
14.3.6-2	Potentially Clinically Meaningful Changes in Vital Sign Measurements	Safety
F14.3.6-1a	Box Plot of Systolic Blood Pressure by Visit	Safety
F14.3.6-1b	Box Plot of Change from Baseline in Systolic Blood Pressure by Visit	Safety
F14.3.6-2a	Box Plot of Diastolic Blood Pressure by Visit	Safety
F14.3.6-2b	Box Plot of Change from Baseline in Diastolic Blood	Safety

	Pressure by Visit	
F14.3.6-3a	Box Plot of Pulse Rate by Visit	Safety
F14.3.6-3b	Box Plot of Change from Baseline in Pulse Rate by Visit	Safety
F14.3.6-4a	Box Plot of Respiratory Rate by Visit	Safety
F14.3.6-4b	Box Plot of Change from Baseline in Respiratory Rate by Visit	Safety
14.3.7-1	Shift Table Summary of ECG Findings at Weeks 24/48 versus at Screening	Safety
14.3.7-2	Summary of ECG Parameters	Safety
14.3.7-3	Potentially Clinically Meaningful Changes in ECG Parameters	Safety
14.3.8	Shift Table Summary of Physical Examinations	Safety
14.4	Summary of Exposure	
14.4-1	Duration of Treatment	Safety
14.4.1a	Number (%) of Subjects on Treatment by Study Week	Safety
14.4-2	Number of infusions	Safety
14.4-3	FG-3019 Dose Amount (mg) per Infusion	Safety
14.4-4	Summaryof Treatment Compliance	Safety
14.5	Summary of PK, PD, and Biomarker Parameters	
14.5.1	Summary of Peak and Trough Concentration Level of FG-3019, Pirfenidone, Nintedanib and Metabolite-to Parent Ratio by Time Point	Substudy

9.2   List of Listings

16.1.7	Randomization scheme and codes	Randomized
16.2.1	Discontinued Subjects	Randomized
16.2.1a	Subject Disposition	Randomized
16.2.2-1	Subjects Who Did Not Meet Entry Criteria	Randomized
16.2.2-2	Protocol deviations	Randomized

16.2.3-1	Subjects Excluded from Efficacy Analysis	Randomized
16.2.3-2	Subjects Included in Efficacy Analysis	FAS
16.2.4.1	Demographic Data	Randomized
16.2.4.2	Medical History	Randomized
16.2.4.3	Baseline Characteristics	Randomized
16.2.4.4	Baseline Gap Score and Staging	Randomized
16.2.4.5	Smoking History	Randomized
16.2.4.6	Previous PFT Values and Estimated Annualized Rate of Change	Randomized
16.2.4.7	Medications for IPF in the Past	Randomized
16.2.4.8	Prior Medications	Randomized
16.2.4.9	Concomitant Medications	Randomized
16.2.4.10	Procedures and Non-Drug Therapies	Randomized
16.2.4.11	Oxygen Use	Randomized
16.2.4.12	IPF Confirmation	Randomized
16.2.5.1	Study Drug Administration	Safety
16.2.5.1-1	Treatment Compliance	Safety
16.2.6.1-1	FVC Observed, % predicted, Change from Baseline, and Imputation Values	Safety
16.2.6.1-2	DLCO Observed, % predicted, and Change from Baseline	Safety
16.2.6.1-3	FEV1 Observed, % predicted, and Change from Baseline	Safety
16.2.6.1-4	FRC Observed, % predicted, and Change from Baseline	Safety
16.2.6.1-5	TLC Observed, % predicted, and Change from	Safety

	Baseline,	
16.2.6.1-6	FEV1/FVC and Change from Baseline	Safety
16.2.6.2-1	Time to Progression of IPF	Safety
16.2.6.3-1	HRCT CAD Scores and Change from Baseline	Safety
16.2.6.3-2	HRCT Visual Reading Scores	Safety
16.2.6.4-1	SGRQ Domain and Total Scores and Change from Baseline	Safety
16.2.6.4-2	UCSD Shortness of Breath Total Score and Change from Baseline	Safety
16.2.7-1	Listing of All Reported Adverse Events	Safety
16.2.7-2	Listing of Adverse Events Assessed as At Least Possibly Related to FG-3019/Placebo	Safety
16.2.7-2a	Listing of Adverse Events Assessed as At Least Possibly Related to FG-3019/Placebo or Background Therapies	Substudy Safety
16.2.7-3	Listing of Adverse Events Leading to Treatment/Study Discontinuation	Safety
16.2.7-4	Listing of Adverse Events with Severity Grade ≥ 3	Safety
16.2.7-4a	Listing of Possible Treatment Related Adverse Events with Severity Grade ≥ 3	Safety
16.2.7-5	Listing of Possible Infusion Associated Adverse Events	Safety
16.2.7-5a	Listing of PI Identified Possible Infusion Associated Adverse Events	Safety
16.2.7-6	Listing of Possible Procedure Related Adverse Events	Safety
16.2.8.1-1	Complete Blood Count	Safety
16.2.8.1-2	Complete Blood Count: Absolute	Safety

16.2.8.1-3	Complete Blood Count: WBC Differential	Safety
16.2.8.1-4	Complete Blood Count: Morphology	Safety
16.2.8.2-1	Serum Chemistry Results: Liver Function Test	Safety
16.2.8.2-2	Serum Chemistry Results: Kidney Function and Metabolic	Safety
16.2.8.2-3	Serum Chemistry Results: Coagulation	Safety
16.2.8.2-4	Serum Chemistry Results: Electrolytes	Safety
16.2.8.2-5	Serum Chemistry Results: Other	Safety
16.2.8.3-1	Urinalysis Results	Safety
16.2.8.3-2	Urinalysis Results: Casts	Safety
16.2.8.3-3	Urinalysis Results: Cellular	Safety
16.2.8.3-4	Urinalysis Results: Crystals	Safety
16.2.8.4	Pregnancy Results – Females	Safety
16.2.9	Physical Examinations	Safety
16.2.10-1	Vital Signs	Safety
16.2.10-2	Vital Signs (Average over Post-Infusion Time Points)	Safety
16.2.11	Electrocardiogram	Safety

APPENDICES

Appendix I Computation Formulae for Predicted PFT Parameters

Predicted PFT values are computed based on Age, Height (Ht), Weight (Wt), and BSA. Measurements at Screening will be used in computing the predicted values throughout the study. Specifically,

Age_d = (Date when PFT is performed – Date of birth)/365.25

Ht = height in centimeters measured at Screening

Wt = weight in kilograms measured at Screening

BSA = Body Surface area in meter^2 calculated by the formula of DuBois and DuBois = (Wt $^{0.425}$ x Ht $^{0.725}$) x 0.007184

Note: Age_d, without rounding to integer, is used only in calculating predicted values for the PFT parameters.

1 Predicted FVC and FEV1

The predicted values of FVC and FEV1 are calculated using the following algorithms for the corresponding gender-age groups (Hankinson JL, et al.)

1.1 Predicted FVC

Gender-Race-Age Group	Intercept	Age_d (yr)	Age_d^2 (yr)	Ht^2 (cm)
Men, Caucasian, $Age_d \geq 20$	-0.1933	0.00064	-0.000269	0.00018642
Men, African American, $Age_d \geq 20$	-0.1517	-0.01821	0	0.00016643
Women, Caucasian, $Age_d \geq 18$	-0.3560	0.01870	-0.000382	0.00014815
Women, African American, $Age_d \geq 18$	-0.3039	0.00536	-0.000265	0.00013606

Men, Caucasian , $Age_d \geq 20$

Predicted FVC (L) = -0.1933 + 0.00064* Age_d -0.000269* Age_d^2 + 0.00018642*Ht^2

Men, African American, $Age_d \geq 20$

Predicted FVC (L) = -0.1517 - 0.01821* Age_d + 0.00016643*Ht^2

Women, Caucasian , $Age_d \geq 18$

Predicted FVC (L) = -0.3560 + 0.01870* Age_d - 0.000382* Age_d^2 + 0.00014815*Ht^2

Women, African American, $Age_d \geq 18$

Predicted FVC (L) = -0.3039 + 0.00536* Age_d - 0.000265* Age_d^2 + 0.00013606*Ht^2

Subjects of Southeast Asian ethnicity, as reported by the investigator, were applied a correction factor of 0.88 to the Caucasian equations based on more pertinent demographic characteristics, as per Quanjer et al. (2012) and the study Pulmonary Function Tests experts.

For all other cases the Caucasian predicted values will be used.

% predicted FVC = measured FVC (L)/predicted FVC (L) *100.

1.2 Predicted FEV1

Gender-Race-Age Group	Intercept	Age_d (yr)	Age_d^2 (yr)	Ht^2 (cm)
Men, Caucasian, $Age_d \geq 20$	0.5536	-0.01303	-0.000172	0.00014098
Men, African American, $Age_d \geq 20$	0.3411	-0.02309	0	0.00013194
Women, Caucasian, $Age_d \geq 18$	0.4333	-0.00361	-0.000194	0.00011496
Women, African American, $Age_d \geq 18$	0.3433	-0.01283	-0.000097	0.00010846

Men, Caucasian and other races, $Age_d \geq 20$

Predicted FEV1 (L) = 0.5536 - 0.01303* Age_d - 0.000172* Age_d^2 + 0.00014098*Ht^2

Men, African American, $Age_d \geq 20$

Predicted FEV1 (L) = 0.3411 - 0.02309* Age_d + 0.00013194*Ht^2

Women, Caucasian and other races, $Age_d \geq 18$

Predicted FEV1 (L) = 0.4333 - 0.00361* Age_d - 0.000194* Age_d^2 + 0.00011496*Ht^2

Women, African America, $Age_d \geq 18$

Predicted FEV1 (L) = 0.3433 - 0.01283* Age_d - 0.000097* Age_d^2 + 0.00010846*Ht^2

Subjects of Southeast Asian ethnicity, as reported by the investigator, were applied a correction factor of 0.88 to the Caucasian equations based on more pertinent demographic characteristics, as per Quanjer et al. (2012) and the study Pulmonary Function Tests experts.

For all other cases the Caucasian predicted values will be used.

% Predicted FEV1 = measured FEV1 (L) / predicted FEV1 (L)

2 Predicted DLCO Corrected for Hb

Predicted DLCO comes from Miller A, et al. Predicted value for non-smoking (NS) status will be used for all calculations in this study.

Gender	Intercept	Age$_d$ (year)	Ht (cm)
Men	12.9113	-0.2290	0.4180 / 2.54
Women	2.2382	-0.1111	0.4068 / 2.54

For men:

Predicted DLCO (ml/min/mmHg) = 12.9113 - 0.2290* Age$_d$ + 0.4180* Ht / 2.54

For women:

Predicted DLCO (ml/min/mmHg) = 2.2382 - 0.1111* Age$_d$ + 0.4068* Ht / 2.54

<u>Predicted DLCO Corrected for Hb</u>

For men:

Predicted DLCO corrected for Hb (ml/min/mmHg)

= DLCO predicted * 1.7 * Hb / (10.22 + Hb)

For women,

Predicted DLCO corrected for Hb (ml/min/mmHg)

= DLCO predicted * 1.7 * Hb / (9.38 + Hb)

where Hb = hemoglobin in g/dL measured at the closest point to the time when DLCO is measured. If two Hb values are available with the equal distance, the earlier measurement is used.

For men

% predicted DLCO (corrected for Hb)

= measured DLCO / predicted DLCO corrected for Hb *100

= measured DLCO / {predicted DLCO * 1.7 * Hb / (10.22 + Hb)} *100

For women

% predicted DLCO (corrected for Hb)

= measured DLCO / predicted DLCO corrected for Hb *100

= measured DLCO / {predicted DLCO * 1.7 * Hb / (9.38 + Hb)} * 100

3 Predicted TLC and FRC

Predicted TLC and predicted FRC are based Goldman HI and Becklake MR as shown below.

Gender	Parameter	Intercept	Age$_d$ (year)	Ht (cm)	Wt (kg)	BSA (m^2)
Men	TLC	-9.167	-0.015	+0.094	0	0

	FRC	-7.11	0	+0.081	0	-1.792
Women	TLC	-7.49	-0.008	+0.079	0	0
	FRC	-4.74	0	+0.053	-0.017	0

For men:

Predicted TLC (L) = -9.167 - 0.015* Age_d + 0.094*Ht

Predicted FRC (L) = -7.11 + 0.081*Ht - 1.792*BSA

For women:

Predicted TLC (L) = -7.49 - 0.008* Age_d + 0.079*Ht

Predicted FRC (L) = -4.74 + 0.053*Ht - 0.017*Wt

% predicted TLC = measured TLC (L) / predicted TLC (L) *100.

% predicted FRC = measured FRC (L) / predicted FRC (L) *100.

Appendix II   Scoring of the SGRQ Domain and Total Scores

The description in this section is extracted from *St. George's Respiratory Questionnaire Manual*, Version 2.3 (June 2009), by Jones, Paul, St. George's, University of London.

1   Overview of SGRQ

The Saint George's Respiratory questionnaire (SGRQ) is designed to measure health impairment in patients with asthma and COPD. The questionnaire is in two parts. Part 1 is to assess the patient's perception of their recent respiratory problems (Symptoms). Part 2 is to measure disturbances of patients' daily physical activity (Activity) and disturbances of psycho-social function (Impacts). The domain and the total scores are percentages of sums of weights divided by the maximum possible score in the corresponding domain. The scores range from 0 to 100, where 0 indicates the best and 100 the worst possible health status.

The questions covered by each domain and the maximum possible scores are listed below.

Domain	Questions	Maximum Possible Score
Symptoms	1 – 8	662.5
Activity	11 and 15	1209.1
Impacts	9 – 10, 12 – 14, 16 – 17	2117.8
Total	All	3989.4

2   SGRQ Scoring Algorithm

The component scores and the total score are derived in the following steps:

1. The weights for all items with positive responses are summed.

2. The weights for missed items are deducted from the maximum possible weight for each component. The weights for all missed items are deducted from the maximum possible weight for the total score.

3. The score is calculated by dividing the summed weights by the adjusted maximum possible weight for that component and multiplying the result by 100 (rounded to the nearest integer).

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 is acceptable.

The Symptoms component is calculated from the summed weights for the positive responses to questions 1-8. This component will tolerate a maximum of two missed items. More than two missing items will result in a missing Symptoms score.

The Activity component is calculated from the summed weights for the positive responses to questions 11 and 15. This component will tolerate a maximum of four missed items. More than four missing items will result in a missing Activity score.

The Impacts component is calculated from the summed weights for the positive responses to questions 9-10, 12-14, and 16-17. This component will tolerate a maximum of six missed items. More than six missing items will result in a missing Impacts score.

The Total score is calculated from the summed weights for the positive responses in the questionnaire and expressing the result as a percentage of the total possible weight for the non-missing items. If more than 24% of the items are missing then the Total score will be missing.

3 Weight of Each Question

PART 1

1) Over the past 3 months, I have coughed:

      Most   80.6

      Several 63.2

      A few 29.3

      Only   28.1

      Not   0.0

2) Over the past 3 months, I have brought up phlegm (sputum):

      Most 76.8

      Several 60.0

      A few 34.0

      Only 30.2

      Not 0.0

3) Over the past 3 months, I have had shortness of breath:

      Most 87.2

      Several 71.4

      A few 43.7

      Only 35.7

      Not 0.0

4) Over the past 3 months, I have had attacks of wheezing:

      Most 86.2

      Several 71.0

      A few 45.6

      Only 36.4

      Not 0.0

5) During the past 3 months, how many severe or very bad unpleasant attacks of chest trouble have you had?

More than three 86.7

3 attacks 73.5

2 attacks 60.3

1 attack 44.2

None 0.0

6) How long did the worst attack of chest trouble last?

a week or more 89.7

3 or more days 73.5

1 or 2 days 58.8

less than a day 41.9

7) Over the last year, in an average week, how many good days (with little chest trouble) have you had?

None 93.3

1 or 2 76.6

3 or 4 61.5

nearly every day 15.4

every day 0.0

8) If you have a wheeze, is it worse in the morning?

No 0.0

Yes 62.0

PART 2

9) How would you describe your chest condition?

The most important problem I have 83 .2

Causes me quite a lot of problems 82.5

Causes me a few problems 34.6

Causes no problem 0.0

10) If you have ever had paid employment?

My chest trouble made me stop work 88.9

My chest trouble interferes with my work or made me change my work 77.6

My chest trouble does not affect my work 0.0

11) Questions about what activities usually make you feel breathless.

Sitting or lying still 90.6

Getting washed or dressed 82.8

Walking around the home 80.2

Walking outside on the level 81.4

Walking up a flight of stairs 76.1

Walking up hills 75.1

Playing sports or games 72.1

12) More questions about your cough and breathlessness.

My cough hurts 81.1

My cough makes me tired 79.1

I get breathless when I talk 84.5

I get breathless when I bend over 76.8

My cough or breathing disturbs my sleep 87.9

I get exhausted easily 84.0

13) Questions about other effects your chest trouble may have on you.

My cough or breathing is embarrassing in public 74.1

My chest trouble is a nuisance to my family, friends or neighbours 79.1

I get afraid or panic when I cannot get my breath 87.7

I feel that I am not in control of my chest problem 90.1

I do not expect my chest to get any better 82.3

I have become frail or an invalid because of my chest 89.9

Exercise is not safe for me 75.7

Everything seems too much of an effort 84.5

14) Questions about your medication.

My medication does not help me very much 88.2

I get embarrassed using my medication in public 53.9

I have unpleasant side effects from my medication 81.1

My medication interferes with my life a lot 70.3

15) Questions about how activities may be affected by your breathing.

I take a long time to get washed or dressed 74.2

I cannot take a bath or shower, or I take a long time 81.0

I walk more slowly than other people, or I stop for rests 71.7

Jobs such as housework take a long time, or I have to stop for rests 70.6

If I walk up one flight of stairs, I have to go slowly or stop 71.6

If I hurry or walk fast, I have to stop or slow down 72.3

My breathing makes it difficult to do things such as walk up hills, carry things up stairs, light gardening such as weeding, dance, play bowls or play golf 74.5

My breathing makes it difficult to do things such as carry heavy loads, dig the garden or shovel snow, jog or walk at 5 miles per hour, play tennis or swim 71.4

My breathing makes it difficult to do things such as very heavy manual work, run, cycle, swim fast or play competitive sports 63.5

16) We would like to know how your chest trouble usually affects your daily life.

I cannot play sports or games 64.8

I cannot go out for entertainment or recreation 79.8

I cannot go out of the house to do the shopping 81.0

I cannot do housework 79.1

I cannot move far from my bed or chair 94.0

17) Tick the statement which you think best describes how your chest affects you.

It does not stop me doing anything I would like to do 0.0

It stops me doing one or two things I would like to do 42.0

It stops me doing most of the things I would like to do 84.2

It stops me doing everything I would like to do 96.7

APPENDIX III GENERAL SPECIFICATIONS FOR SUBMISSION DATA

1. Study Data Tabulation Model (SDTM)

Raw datasets will be mapped to SDTM datasets following variable names and attributes specified in the SDTM Implementation Guide (version 3.2 and subsequent update). Data coding is also mapped to the SDTM controlled terminology. OpenCDISC module will be run to confirm compliance to the SDTM guidelines. Deviations from the guidelines will be documented. Variables that exist in different source datasets, such as lab data collection date, will be consolidated and included once in the SDTM datasets. Unscheduled 'Not Done' records are excluded from the SDTM datasets. Detailed mapping specifications are documented in Define.xml and annotation CRFs.

CRF variables that are not included in the SDTM datasets will be marked as 'Not Submitted', since the SDTM datasets will be submitted as the raw datasets and the raw datasets that match the CRFs will not be submitted to regulatory agencies.

The SDTM datasets are the basis to derive the analysis datasets. Summary tables and listings that do not require complex derived variables may be generated directly from SDTM datasets.

Table A contains a list of SDTM datasets to be created for the study. The corresponding supplemental datasets are not included. This set of SDTM datasets includes only data for the Randomized Treatment Period as described in Section 1.

Table A.   Study Data Tabulation Model Datasets (SDTM)

SDTM Domain	SDTM Domain Description	SDTM Domain Structure	General Observation Class	Source Data Used	Key Variables
AE	Adverse Events	One record per adverse event per subject	Events	AE	STUDYID, USUBJID, AESTDTC, AEDECOD, AESPID
CM	Concomitant Medications	One record per recorded medication occurrence per subject	Interventions	CM, CMIPF	STUDYID, USUBJID, CMCAT, CMSPID, CMTRT, CMSTDTC
DD	Death Details	One record per subject	Findings	DTH	STUDYID, USUBJID, DDTESTCD
DM	Demographics	One record per subject	Special Purpose Domains	DM, EX, ICF, DTH, SITE_INV	STUDYID, USUBJID

DS	Disposition	One record per disposition status or protocol milestone per subject	Events	DS, ICF, DTH, EX	STUDYID, USUBJID, DSSTDTC, DSDECOD, DSSPID, EPOCH
DV	Protocol Deviations	One record per protocol deviations per subject	Events	DV	STUDYID, USUBJID, DVTERM
EC	Exposure as Collected	One record per protocol-specified study treatment per collected-dosing interval per subject	Interventions	EX	STUDYID, USUBJID, ECTRT, ECSTDTC, ECGRPID
EG	ECG Test Results	One record per ECG observation per visit per subject	Findings	EG	STUDYID, USUBJID, EGTESTCD, VISITNUM
EX	Exposure	One record per constant dosing interval per subject	Interventions	EX	STUDYID, USUBJID, EXTRT, EXSTDTC
IE	Inclusion/Exclusion Criteria Not Met	One record per inclusion/exclusion criterion not met per subject	Findings	IE	STUDYID, USUBJID, IETESTCD
LB	Laboratory Test Results	One record per lab test per specimen per method per LOINC code per reason not done per visit per subject	Findings	eDT LB, RAND, LBC, LBPREG	STUDYID, USUBJID, LBCAT, LBTESTCD, VISITNUM, LBMETHOD, LBREASND, LBLOINC
MH	Medical History	One record per medical history event per time interval per subject	Events	MH	STUDYID, USUBJID, MHSPID, MHDECOD, MHSTDTC, MHENDTC
PC	Pharmacokinetic Concentrations	One record per time-point concentration or sample characteristic per analyte per subject	Findings	Source documents, PKC	STUDYID, USUBJID, PCTESTCD, VISITNUM, PCTPTNUM

PE	Physical Examination	One record per body system or abnormality per visit per subject	Findings	PE	STUDYID, USUBJID, PETESTCD, VISITNUM
PP	Pharmacokinetic Parameters	One record per PK parameter per time-concentration profile per modeling method per subject	Findings	Source documents	STUDYID, USUBJID, PPTESTCD, PPCAT, VISITNUM, PPTPTREF
PR	Procedures	One record per recorded procedure per occurrence per subject	Interventions	NDT, OU	STUDYID, USUBJID, PRCAT, PRTRT, PRSPID
QS	Questionnaires	One record per question per questionnaire per visit per subject	Findings	SGRQ, UCSD	STUDYID, USUBJID, QSTESTCD, VISITNUM
SC	Subject Characteristics	One record per characteristic per subject	Findings	DM	STUDYID, USUBJID, SCTESTCD
SE	Subject Elements	One record per actual Element per subject	Special Purpose Domains	SDTM.DM, SDTM.EX, SDTM.DS, SDTM.SV	STUDYID, USUBJID, TAETORD, SESTDTC
SU	Substance Use	One record per substance type per subject	Intervention	SUTOB	STUDYID, USUBJID, SUTRT
SV	Subject Visits	One record per actual visit per subject	Special Purpose Domains	All datasets including visits	STUDYID, USUBJID, VISITNUM
TA	Trial Arms	One record per planned Element per Arm	Trial Design	N/A	STUDYID, ARMCD, TAETORD
TE	Trial Elements	One record per planned Element	Trial Design	N/A	STUDYID, ETCD
TI	Trial Inclusion/Exclusion Criteria	One record per I/E criterion per protocol criteria version	Trial Design	N/A	STUDYID, TIVERS, IETESTCD
TS	Trial Summary	One record per Trial Summary parameter	Trial Design	N/A	STUDYID, TSPARMCD,

		per occurrence			TSSEQ
TV	Trial Visits	One record per planned Visit per Arm	Trial Design	N/A	STUDYID, VISITNUM, ARMCD
VS	Vital Signs	One record per vital sign measurement per time point per visit per subject	Findings	VS, SC	STUDYID, USUBJID, VSTESTCD, VISITNUM, VSDTC
ZH	High Resolution Computed Tomography	One record per examination per location per time point per visit per subject	Findings	HRCT, TEXTCAD, TEXTVIS, IMQUAL	STUDYID, USUBJID, ZHCAT, ZHTESTCD, VISITNUM, ZHDTC, ZHLOC, ZHEVAL
ZI	IPF Confirmation	One record per IPF Confirmation per test per subject	Findings	IPFD	STUDYID, USUBJID, ZITESTCD
ZP	Pulmonary Function Tests	One record per test per time point per visit per result per subject	Findings	eDT PFT, RAND, ZP, MHPFT	STUDYID, USUBJID, ZPCAT, ZPTESTCD, VISITNUM, ZPDTC, ZPORRES

2 Analysis Data Model (ADaM)

The ADaM datasets are created based on the SDTM datasets following the ADaM Implementation Guide (version 1.0 and subsequent update). They include derived parameters and flags that are needed to generate tables/listings/figures. ADSL and ADAE are created using the specifications provided in the ADaMIG. Study endpoints are included in analysis files in ADaM Basic Data Structure (BDS). Horizontal analysis files are created for analyses on relationship of multiple endpoints. Variable names and labels of the horizontal files are from PARAMCD and PARAM of the corresponding BDS files. Detail derivations of each variable in the ADaM datasets are documented in the Metadata.

Table B.   Analysis Data Model (ADaM)

Dataset	Description	Structure	Keys
ADSL	Subject-Level	One record per subject	STUDYID, SUBJID

	Analysis Dataset		
ADAE	Analysis Dataset Adverse Events	One record per subject per adverse event	STUDYID, SUBJID, AESTDTC, AEDECOD, AESPID
ADCM	Analysis Dataset Concomitant Medications	One record per subject per recorded medication occurrence	STUDYID, SUBJID, CMCAT, CMSPID, CMTRT, CMSTDTC
ADDD	Analysis Dataset Death	One record per subject per paramter	STUDYID SUBJID PARAMCD
ADEG	Analysis Dataset for ECG Test Results	One record per subject per visit	STUDYID, SUBJID, PARAMCD, AVISITN, ADT
ADEX	Analysis Dataset Exposure	One record per subject per parameter per visit	STUDYID, SUBJID, PARAMCD, AVISITN
ADHRCT	Analysis Dataset HRCT	One record per subject per measurement method per Lung Lobe per parameter per visit	STUDYID, SUBJID, PARAMCD, AVISITN, DTYPE ALOC
ADLB	Analysis Dataset Laboratory Test Results	One record per subject per category per parameter per visit per normal range per result	STUDYID, SUBJID, PARCAT1, PARCAT2, PARAMCD, AVISITN, ANRHI, AVAL
ADMH	Analysis Dataset Medical History	One record per subject per medical history per time interval of medical history	STUDYID, SUBJID, MHSPID, MHDECOD, MHSTDTC, MHENDTC
ADPC	Analysis Dataset PK Concentrations	One record per subject per visit per time point per parameter	STUDYID, SUBJID, AVISITN, ATPTN, PARAMCD
ADPE	Analysis Dataset Physical Examination	One record per subject per parameter per visit	STUDYID, SUBJID, AVISITN, PARAMCD
ADPF	Analysis Dataset Pulmonary Function Tests	One record per subject per parameter per time point per visit	USUBJID, SUBJID, PARAMCD, AVISITN, DTYPE, ADT, ATM

ADPP	Analysis Dataset Pharmacokinetic Parameters	One record per subject per parameter per visit	STUDYID, SUBJID, AVISITN, PARAMCD
ADPR	Analysis Dataset Procedures	One record per recorded procedure per occurrence per subject	STUDYID, SUBJID, PRTRT, PRSPID
ADQS	Analysis Dataset for Questionnaire	One record per subject per category per parameter per visit	STUDYID, SUBJID, PARCAT1, PARAMCD, AVISITN, VISITNUM
ADTTE	Analysis Dataset Time-to-Event	One record per subject per parameter	STUDYID, SUBJID, PARAMCD
ADVS	Analysis Dataset Vital Signs	One record per subject per parameter per time point per visit	STUDYID, SUBJID, PARAMCD, AVISITN, DTYPE, ADT, ATM, ATPTN

APPENDIX IV GENERAL SPECIFICATIONS FOR TABLES, LISTINGS, FIGURES

1 Software Used

All programming of tables, listings and figures (TLFs) will be performed using the statistical software package SAS® version 9.3 or greater.

2 General

All TLFs are based on SDTM and/or ADaM datasets. By default, data listings reflect the actual values captured in SDTM and ADaM datasets, including date/time variables and missing values. Except for concatenation of some variables for compact display purpose, data are presented directly with minimum manipulation. In general, the character standard result variables, such as –STRESC, are presented in data listings. Date are presented in listings in format yyyy-mm-dd. For incomplete date, CDISC presentation convension is followed.

For continuous variables that are recorded as "<X" or ">X", the value of "X" will be used in the calculation of summary statistics. The value "X" is also captured in the numeric variable in the SDTM datasets as well as in the ADaM datasets for consistency, although SDTMIG recommends capturing missing values in the numeric variables.

In general, reported verbatim, such as terms of AE, medical history, medication names, specifications to the 'Other' fields, findings, etc., are presented in upper case. However, when reported fields are long, such as comments and protocol deviation descriptions, listing in lower case enhances readability.

3 Table/Listing/Figure Output File Type and Organization

In general, the final set of TLFs will include both PDF and RTF files. Outputs are combined in several large PDF files, eg, all tables, all listings, and all figures, in the order as in the planned TLF in Section 12. A table of content should be included with hyperlinking to individual outputs. True RTF files (in-text format) will be created for tables and listings. SAS outputs for statistical procedures used in analysis of primary, secondary, and exploratory efficacy endpoints will also be included.

4 Page Layout

All column headers (consisting of one or several words) will start with uppercase and thereafter only lowercase characters, except for acronyms and abbreviations. In case values from the database will be displayed in column headers, they may be displayed as in the database. Pages will be numbered as 'Page x of y', where 'y' is the total number of pages of the corresponding table or listing. The page specifications are presented in Table A.

Table A.   Specifications for Page Layout

Paper Size	Letter
Orientation	Landscape
Alignment	Center
Font size	9
Font type	Courier New (default)
Margins	
Top	0.75"
Bottom	0.38"
Left	0.75"
Right	0.38"

The margin sizes and font size for listings may be flexible to provide sufficient information on a single page to facilitate review and comparison.

When created using SAS, tables and listings will be created using ODS, and output files will be produced in RTF. When RTF files are produced, titles and footnotes will appear as document headers/footers.

5 Titles and Footnotes

All tables and listings will have a header showing "FibroGen, Inc.", the protocol number, database cutoff date or 'Final Database', and Page x of y. A footer will show the program file path/name, output file path/name, run date and time.

All titles are written in title format, with uppercase at the beginning of each word; articles, prepositions, and conjunctions, which are of three characters length or less will start with lowercase letters (Mixed Case). Footnotes are in regular text format.

Titles

In total there are up to 10 titles available, defined as following:

first title "FibroGen, Inc." (left aligned) and "Database extraction date: ddMMMyyy" or "Final Database" (right aligned)

second title protocol number + "Clinical Study Report" (left aligned) and "Page x of y" (right aligned)

third title blank

fourth title: table/listing/figure number

fifth title: table/listing/figure title

sixth title:	population names if provided in SAP, or brief definition of specific analysis set

Footnotes

Up to 10 footnote lines are available for tables, listings and figures. Footnotes 1, 9 and 10 are standard. Footnotes 2 to 8 (left aligned) might be used as needed. They are to be specified in the Shell.

first footnote	is a separating horizontal line.
second – eighth	are free text which can be used for explanations. Footnotes will be referenced using numbers in square brackets, starting with [1], followed by [2] etc.
ninth footnote	left blank; in case needed may also be used as for explanations.
tenth footnote	the program name (left aligned); the date and time in the format ddMMMyyyy hh:mm when the output was created; the version (e.g. draft or final); and the word "Confidential".

Footnotes are denoted by [1], [2], and so on.

If footnotes take more than 30% of the space of a long listing, they may be presented only on a standalone first page.

TLF numbers and titles should be inputted from an external file that can be directly copy-and-pasted from the SAP planned TLFs, rather than including in the body of the program. This is to ensure consistency between the SAP and the actual outputs.

Footnotes may be inputted from an external file as well for ease of managing changes.

For summary tables, the corresponding listings with the parameters being summarized should be footnoted as reference. For figures, the corresponding summary table should be footnoted as reference.

6 Table, Listing, Figure Metadata

The table, listing, and figure (TLF) metadata will include the TLF numbers, titles, analysis populations, program names, input dataset names. For tables and figures, PARAMCD, PARAM, and other conditions will be specified. TLF numbers, titles, and footnotes will be imported from this master spreadsheet. In addition, this spreadsheet will record the names of the original programmer and the validator/reviewer and the date of validation approval.

7 Significant Digits of Summary Statistics

- All percentages will be rounded to one decimal place and aligned by the decimal place.

- If the count is zero, the percentage will be suppressed and only '0' will be presented.

- Any p-values will be rounded to four decimal places and will be presented as '<.0001' if they are less than 0.0001 after rounding.
- For variables of direct measurements, summary statistics are displayed with the following specifications of decimal places in Table B.

Table B.   Significant Digits of Summary Statistics

Description	Characteristic	Number of decimal places
Count	N	0
Mean	Mean	As in source + 1
Standard deviation	Std	As in source + 1
Standard error of the mean	SEM	As in source + 2
Confidence Interval	CI	As in source + 1
Minimum	Min	As in source
Median	Median	As in source
Maximum	Max	As in source
Q1 / Q3	Q1/Q3	As in source
10% / 90%	10%/90%	As in source
Percentage	%	All percentages will be rounded to one decimal place and lined up by the decimal place. The percentage will be suppressed when the count is zero
Coefficient of variation	CV (%)	1
p-value	p-value	p-values will be rounded to four decimal places and will be presented as '<.0001' if they are less than 0.0001 after rounding

N=number; Std=Standard deviation; CI=Confidence Interval; Min=minimum; Max=maximum; CV=Coefficient of variation

As a general guideline for derived parameters, 3 significant digits may be displayed for a parameter with an overall mean less than 100; otherwise, 1 decimal place may be used. If a derived parameter is in the same scale as some related measured parameters, such as MAP, QTc, the same display format may be used as the measured parameters.

Summary Statistics are to be displayed in the following order: Count, Mean, Standard Deviation, <Coefficient of Variation, Standard Error of the Mean, Confidence Interval>, Minimum, <10%>, <Q1>, Median, <Q3>, <90%>, Maximum.

For categorical variables the categories will be displayed in the TLFs in the same order they appear in the CRF.

8 Figure Specifications

- In general, figures should include annotation of key summary statistics: n, mean, SD or SE, median for continuous variables; n and percent for categorical variables; number of subjects at risk and cumulative number of events as well as median and 95% CI for time-to-event data. Other statistics such as quartiles, ranges may be included depending on need and space.

- P-values should be presented if comparisons are of interest.

- For scatter plots, linear or non-linear trend lines should be included if the association of the two variables is of interest. Correlation coefficient or regression coefficients as well as corresponding p-values should be presented.

- For box plots, 'BOXSTYLE=SCHEMATIC' should be used. The whiskers are drawn to the most extreme points in the group that lie within the fences. The upper fence is defined as the third quartile (represented by the upper edge of the box) plus 1.5 times the interquartile range. The lower fence is defined as the first quartile (represented by the lower edge of the box) minus 1.5 times the interquartile range. Observations outside the fences are identified with a special symbol.

9 Unit Conversion

Units Presented in TLFs	Units Reported or Derived from CRF	Conversion Formula
Kilogram (kg)	Pound (lb)	kg = lb/2.2
Centimeter (cm)	Inch (in)	cm = 2.54 * in
Celsius C°	Fahrenheit (F°)	C° = (5/9) * (F° – 32)
Year	Day	1 year = 365.25 days
Months	Day	1 month = 30.4375 days