SPIRIT

Statistical Analysis Plan

Version No.: 1.0 (12NOV2019)

Study Title: A Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Efficacy and Safety of BG00011 in Patients With Idiopathic Pulmonary Fibrosis

Name of Study Treatment: BG00011

Protocol No.: 203PF203 V1.0 (20NOV2017) / NCT03573505

Study Phase: Phase 2b

APPROVAL

This document has been reviewed and approved by:

[redacted], Ph.D. SMT & CDT Statistician	[redacted] Signature	12-NOV-2019 Date
[redacted], Ph.D. [redacted]	[redacted] Signature	12-Nov-2019 Date
[redacted], MD SMT [redacted]	[redacted] Signature	12 Nov 2019 Date

Table of Contents

List of Abbreviations

AE	adverse event
ALT	alanine aminotransferase
ANCOVA	analysis of covariance
AST	aspartate aminotransferase
BUN	blood urea nitrogen
CBC	complete blood count
CI	confidence interval
CSR	clinical study report
CV	coefficient of variation
DL_{co}	carbon monoxide diffusion capacity
DL_{co}/Hb	carbon monoxide diffusion capacity, corrected for hemoglobin
DSMB	Data Safety Monitoring Board
EOT	End of Treatment
ECG	electrocardiogram
FEV_1	a ratio of the forced expiratory volume in one second
FVC	forced vital capacity
Hb	hemoglobin
HRCT	high resolution computed tomography
ICH	International Conference on Harmonization
IPF	idiopathic pulmonary fibrosis
LDH	lactate dehydrogenase
MCH	mean corpuscular hemoglobin
MCHC	mean corpuscular hemoglobin concentration
MCV	mean corpuscular volume
MedDRA	Medical Dictionary for Regulatory Activities
PFT	pulmonary function test
PK	pharmacokinetic
PT	preferred term
QLF	quantitative lung fibrotic reticulation score
qHC	quantitative honeycomb
qGG	quantitative ground-glass
RV	residual volume
SAP	Statistical Analysis Plan
SC	subcutaneous(ly)
SD	standard deviation
SLB	surgical lung biopsy
SOC	system organ class
TEAE	treatment emergent adverse event
TFLs	Tables, Figures, and Listings
UIP	usual interstitial pneumonia

1 INTRODUCTION

The intent of this document is to provide guidance for analyses of data for this Phase 2b study. This statistical analysis plan (SAP) has been developed after review of the Clinical Study Protocol 203PF203 (version 1 dated 20NOV2017).

On 16 August 2019, the BG00011 independent Data Safety and Monitoring Board (DSMB) met for a routinely scheduled bi-monthly meeting. During that meeting the DSMB recommended that study 203PF203 be stopped due to an imbalance in the prespecified safety composite endpoint. A separate unblinded team at Biogen has reviewed the data and accepted the DSMB recommendation. Because of the study early termination, the version 2 protocol dated on 02 JULY 2019 was not implemented, it is not used as reference for developing this SAP.

As the study was stopped early, all endpoints will be analyzed and summarized by descriptive statistics only with exception of primary and key secondary efficacy endpoints (i.e. Percent Predicted FVC, Absolute DLco corrected for hemoglobin and Percent Predicted Absolute DLco corrected for hemoglobin). The safety data will be analyzed in full. This version 1.0 SAP supersedes the statistical considerations identified in the protocol.

1.1 SCOPE OF WORK

This statistical analysis plan (SAP) covers the analyses includes all available subjects who enrolled in the study before the early termination due to the urgent safety measure.

All analyses of efficacy, safety, PK and exploratory endpoints described in this plan will be included in the final analysis.

The final analysis will employ a two-part database lock, with unblinding occurring after Part I data are available. The details of unblinding at the end of study is documented in the unblinding plan version 1.2 dated on 07NOV2019. The SAP will be finalized prior to the lock of Part I data. No changes to the SAP post finalization are expected. However, if any changes are necessary, the SAP will be updated to a new version with all revisions documented.

1.2 TABLES, FIGURES AND LISTINGS

A detailed description of the planned Tables, Figures, and Listings (TFLs) to be presented in the clinical study report (CSR) is provided in the accompanying TFLs shell document. When the SAP and TFL shells are agreed upon and finalized, they will serve as the template for this study's CSR.

2 STUDY OBJECTIVES AND ENDPOINTS

Primary Objective	**Primary Endpoints**
To evaluate the efficacy of BG00011 compared with placebo in subjects with IPF.	Yearly rate of change in forced vital capacity (FVC, expressed in mL over 52 weeks) in subjects randomized to BG00011 compared with placebo.
Secondary Objectives	**Secondary Endpoints**
To evaluate the efficacy of BG00011 compared with placebo in subjects with IPF as determined by change in percent predicted FVC.	Yearly rate of change in FVC, expressed in percent predicted, over 52 weeks.
To assess progression-free survival in subjects who receive BG00011 compared with placebo.	Time to progression, as defined by a composite endpoint, including any of the following events: • Absolute decline of 10% predicted in FVC (FVC percent predicted$_{baseline}$ – FVC percent predicted$_{progression}$ ≥10%). • Nonelective hospitalization for respiratory events. • Lung transplantation or death.
To assess the occurrence of IPF exacerbation (using modified diagnostic criteria for acute IPF exacerbation derived from [Collard 2007]) in subjects who receive BG00011 compared with placebo.	• Time to first acute exacerbation, measured in days. • Proportion of subjects with at least 1 acute exacerbation during the 52 weeks on study. • Number of exacerbations during 52 weeks.
To assess the incidence of absolute decline in FVC ≥10% in subjects who receive BG00011 compared with placebo.	Number of subjects with absolute decline of 10% predicted in FVC (FVC percent predicted$_{baseline}$ – FVC percent predicted$_{progression}$ ≥10%) over 52 weeks.
To assess the time to death or lung transplantation in subjects who receive BG00011 compared with placebo, and the transplant-free survival rate at Week 26 and Week 52.	Time to death or lung transplantation, measured in days.

To assess the time to nonelective hospitalizations in subjects who receive BG00011 compared with placebo,	Time to all nonelective hospitalizations and to nonelective respiratory hospitalizations, measured in days.
To assess additional PFT findings in subjects who receive BG00011 compared with placebo.	• Change in absolute and percent predicted FVC from baseline over time. • Carbon monoxide diffusion capacity (DL_{co}), absolute and percent predicted changes from baseline over time. • Total lung capacity, as measured by plethysmography, absolute and percent predicted changes from baseline over time.
To assess performance on the 6MWT in subjects who receive BG00011 compared with placebo.	Change from baseline in 6MWT parameters at Weeks 26 and 52.
To evaluate the safety and tolerability of BG00011.	• The incidence, severity, outcome, and relationship to study treatment of AEs and SAEs. • Change from baseline in clinical laboratory test results, vital signs, ECG, PFT results, and HRCT findings. • Immunogenicity (antibodies to BG00011).
To evaluate the serum concentration of BG00011.	Measurement of BG00011 serum concentrations using sparse PK sample collection at select timepoints during the study.
Exploratory Objectives	**Exploratory Endpoints**
[redacted]	[redacted]
[redacted]	[redacted]

[redacted]	[redacted] • [redacted] • [redacted] • [redacted]
[redacted]	[redacted]
[redacted]	• [redacted] • [redacted] • [redacted]
[redacted]	• [redacted] • [redacted] • [redacted]
[redacted]	• [redacted]

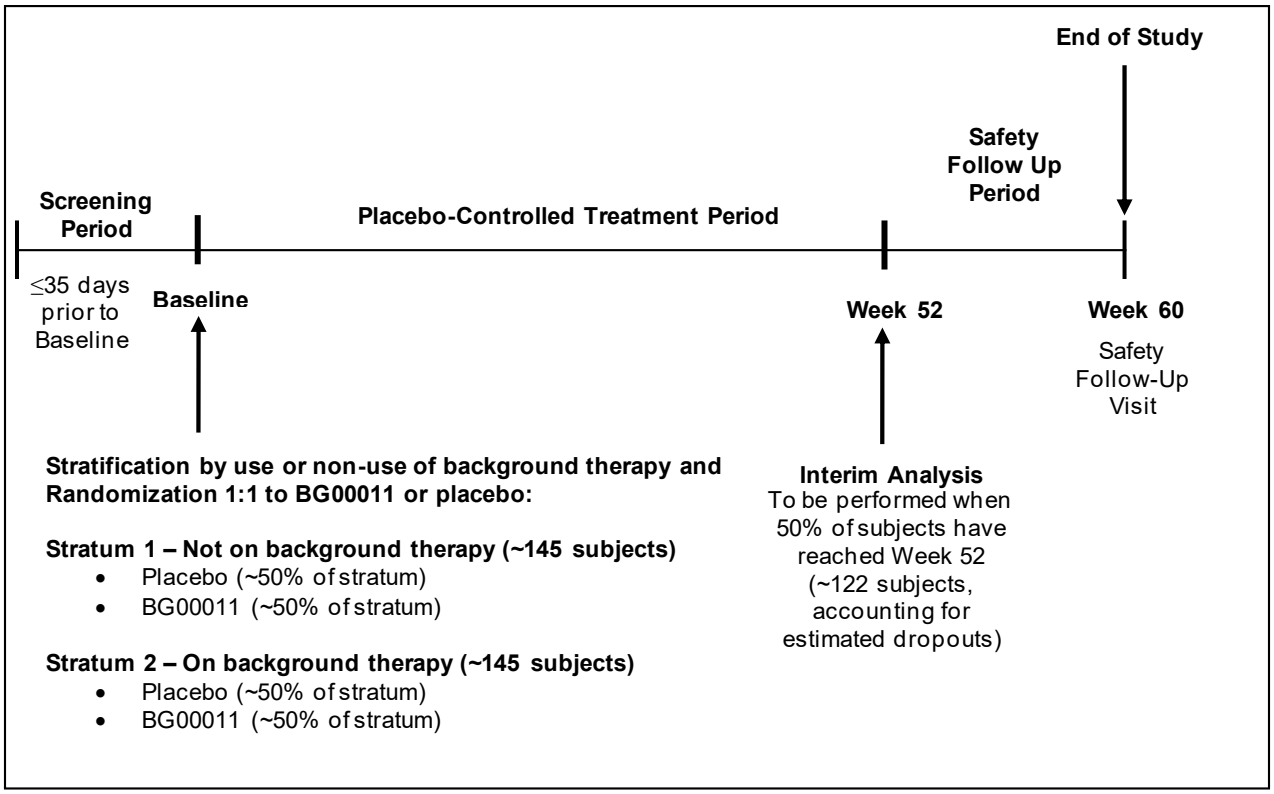
3 STUDY DESIGN

This is a Phase 2b randomized, double-blind, placebo-controlled study designed to evaluate the efficacy, safety, PK, and tolerability of BG00011 in subjects with IPF. Approximately 290 subjects who meet inclusion/exclusion criteria will be stratified by background therapy status (subjects receiving background therapy and subjects not receiving background therapy) and randomized in a 1:1 ratio of BG00011 to placebo (**Error! Reference source not found.** of the protocol).

Subjects will receive weekly injections of BG00011 or placebo for 52 consecutive weeks (a total of 52 doses). Study duration for an individual subject is approximately 65 weeks, including a Screening Visit up to 5 weeks prior to the first dose of study treatment, a Placebo-Controlled Treatment Period of 52 weeks, and a Safety Follow-Up Visit 8 weeks after the end of the Treatment Period.

A schematic of the study design is provided in Figure 1. Additional details on the study design and the schedule of study assessment is provided in the full protocol. Due to the early termination of the study, no interim analysis will be performed. This change supersedes the interim analysis planned in the schematic of the study design in the protocol.

Figure 1: Schematic for Study 203PF203



4 SAMPLE SIZE JUSTIFICATION

The sample size calculation is based on the annual declining rate in FVC (mL/year). With 145 subjects randomized per arm, and assuming a 15% dropout rate over a 1-year treatment period, the study has approximately 80% power to show that the BG00011 treatment arms differentiate from the arm without BG00011 at a 2-sided 15% α level, given a treatment effect of 82.5 mL/year with a common standard deviation of 280 mL. The calculation is based on the assumption that the FVC decline rate in subjects receiving placebo and no background therapy is 220 mL/year, and in subjects receiving background but no BG00011 therapy it is assumed to be 110 mL/year. A 60% reduction in the FVC decline rate in the subjects receiving placebo would be a treatment effect of 132 mL/year. A 30% reduction in the FVC decline rate in subjects receiving background therapy would be a treatment effect of 33 mL/year. With the assumption that 50% of each type of subject are randomized in the study, the combined treatment effect is approximately 82.5 mL/year. The 15% dropout rate for this study was estimated based on prior experience for the placebo arms in the Phase 3 studies for nintedanib and pirfenidone [King 2014; Richeldi 2014].

5 ANALYSIS POPULATIONS

5.1 Modified Intent-to-Treat (MITT) Population

The modified intent-to-treatment (MITT) population will include all subjects who are randomized and receive at least 1 dose of study treatment (BG00011 or placebo). Subject will be analyzed by randomized treatment, even if the subject did not receive the assigned treatment or did not follow the protocol until completion. The definition of MITT population is same as the intent-to-treat (ITT) population stated in the protocol.

5.2 Safety Population

The safety population will include the same subjects as defined for the MITT population. However, subjects will be analyzed by actual treatment received. If all subjects take the treatment as they were randomly assigned to, the safety population will be the same as the MITT population.

5.3 Immunogenicity population

The analysis population for immunogenicity is defined as all MITT subjects who have at least 1 postdose immunogenicity sample evaluated for anti-BG00011 antibodies and whose pre-dose anti-BG00011 samples are screened negative at baseline (Day 0) visit.

5.4 Pharmacokinetic (PK) population

The pharmacokinetic (PK) population will include all safety subjects who have at least one PK concentration measurement. Subjects will be analyzed by the actual treatment received.

6 General Analysis

6.1 General Consideration

Data listings will be provided by treatment group, subjects and visits, if applicable. Summary statistics and statistical analyses will only be presented for data where detailed in this SAP and will generally only be performed for subjects included in the relevant analysis population. Unless otherwise specified, all variables will be summarized for the MITT population.

Summary statistics will be presented throughout for all endpoints at all visits including follow-up visits. For continuous variables, summary statistics will generally include number of subjects with data, mean, standard deviation (SD), median, minimum and maximum. In addition, the 95% confidence interval (CI) for the mean will be presented. For log-normal data the geometric mean and geometric coefficient of variation (CV) will also be presented. For categorical endpoints, summary statistics will generally include number and percentage of subjects with data in each category.

The statistical software, SAS®, will be used for all summaries and statistical analyses.

6.1.1 Analysis Visit and Study Day Definition

6.1.1.1 Baseline

Unless stated otherwise, baseline value is defined as the data collected prior to the time and/or on the date of first dose, which is usually the same day as the Baseline visit (Day 0). If there is more than one value on/before the date of the first dose, the non-missing value closest to the first dose will be used as the baseline value.

6.1.1.2 Analysis Visit

Observations that are collected after the first randomized dose as scheduled visit will be used to the corresponding analysis week.

6.1.1.3 Study/Treatment Day

To facilitate the analysis of the data in the treatment period (TP), study/treatment day will be calculated for each day in the study as the number of days between the date of the first randomized dose and the specific day of interest. For a day that is on or after the day of the first randomized dose, the study/treatment day is a positive value and will be calculated as (date of study day – date of the first randomized dose +1). For a day that is prior to the date of the first randomized dose, the study/treatment day is a negative value and will be calculated as (date of study day – date of the first randomized dose). The study starts at study/treatment day 1.

6.1.1.4 Early Termination or Unscheduled Visits for Acute Exacerbation of IPF

In general, when data are summarized by analysis visit, the Early Termination Visit or Unscheduled Visits for Acute Exacerbation of IPF will be assigned to an appropriate scheduled analysis visit by using a window scheme provided in Table 1.

Table 1: Visit Windows for Early Termination or Unscheduled Visits

Scheduled Visit Time	Target study day	Analysis visit window (days)	
		Efficacy & Safety Data	PK Data
Baseline (Day 0)	1	Most recent non-missing value on/before the date of first dose	
Day 5*	6	N/A	[6,7]
Week 4	29	[2,43]	[8,43]
Week 8	57	[44,71]	
Week 12	85	[72,99]	
Week 16	113	[100,127]	
Week 20	141	[128,162]	
Week 26	183	[163,203]	
Week 32	225	[204,246]	
Week 38	267	[247,288]	
Week 44	309	[289,337]	
Week 52	365	[338,393]	
Safety Follow-up (Week 60)	421	• If subject who discontinued treatment due to reasons other than study termination by sponsor: > min (393, study day of each subject's week 52/EOT visit in treatment period) ** • If subject who discontinued treatment due to study early termination by sponsor: > min (393, study day of each subject's last dose)	

*The Day 5 Visit is for blood collection only (for PK analysis) and may be done in the clinic or at home (by appropriate personnel). Sample collection should occur on Day 5 (not sooner), with a +1 day window allowed (i.e., collection should occur on Day 5 but may occur on Day 6).
** For subjects who discontinued the treatment due to reasons other than study termination by sponsor will go on to complete all scheduled visits until either week 52 or the study was terminated by sponsor due to the urgent safety measure whichever occurs first. The Week 52/ EOT visit in treatment period is defined as the last visit scheduled for either week 52 or study termination due to the urgent safety measure.

The lower bound and the upper bound for the visit windows are defined as the midpoints of the scheduled visits. If the date and time from early termination visit or unscheduled visit for acute exacerbation falls in between the lower bound and the upper bound for a scheduled visit, then it will be assigned to that visit. If there are 2 or more assessments available in the same analysis window for a subject, the assessment that is closest to the target visit day will be used for analysis. If there are 2 or more assessments in the same analysis window with the same distance from the target visit day, the earlier/earliest assessment will be used.

6.2 Subject Disposition

The number and percentage of subjects who were randomized, dosed, randomized but not dosed, completed the study, discontinued study treatment (including reasons for discontinuation, specified in Sections 7.2.4) and withdrew from the study (including reasons for withdrawal, specified in Sections 7.2.5) will be summarized by treatment group and overall. The subject disposition will be summarized for the MITT and safety populations. If the safety population is the same as MITT population (i.e. all subjects receive their randomized treatment), the subject disposition will be summarized for MITT population only.

Additionally, a listing of subjects who discontinued treatment and/or withdrew from study and the associated reasons for discontinuation/withdrawal will be provided.

The number and percentage of subjects in each background stratum, treatment group and overall will be summarized for the MITT and safety populations. If the safety population is the same as MITT population, summary will only be presented for MITT population.

6.3 Demography and Baseline Disease Characteristic

6.3.1 Summary of Demographic information

The demographic information of age, gender, ethnicity, race, height, weight, and body mass index (BMI) at baseline will be summarized by treatment group and overall. In addition, the associated listing will be provided. The demographic will be summarized for the MITT population only – note that if the safety population differs from the MITT population (i.e. if not all subjects receive their randomized treatment) then the summaries will be repeated for the safety population.

6.3.2 Baseline Disease Characteristics

6.3.2.1 GAP score

As a disease characteristic, GAP score [Ley 2012] is an 8-point score (0-8) and calculated from age, gender, baseline percent predicted FVC and percent predicted DLco corrected for hemoglobin (DL_{co}/Hb) at baseline visit. The calculation is described in Table 2 as follows.

Table 2: GAP Score Calculation

	Calculation
Gender	• If female, add 0 point.

	• If male, add 1 point.
Age	• If Age $\leq$ 60, add 0 point • If 61 $\leq$ Age $\leq$ 65, add 1 point • If Age $>$ 65, add 2 points
Baseline percent (%) predicted FVC	• If % predicted FVC $>$ 75, add 0 point • If 50$\leq$ % predicted FVC $\leq$ 75, add 1 point • If % predicted FVC $<$ 50, add 2 points
Percent (%) predicted DL_{co}/Hb at baseline visit	• If % predicted DL_{co}/Hb $>$ 55, add 0 point • If 36$\leq$ % predicted DL_{co}/Hb $\leq$ 55, add 1 point • If % predicted DL_{co}/Hb $\leq$ 35, add 2 points • If DL_{CO} cannot be performed at baseline visit, add 3 points. If the % predicted DL_{co}/Hb is missing due to the unacceptable values through the ERT best test review (BTR) quality control process or hemoglobin missing, the defined baseline % predicted DL_{co}/Hb, the latest non-missing % predicted DL_{co}/Hb before first dosing, will be applied.

6.3.2.2 Summary of Baseline Disease Characteristics

Baseline disease characteristics information of ex-smoker status, randomization stratum (on/off background therapy), region, GAP score, GAP score category (0-3, 4-5 and 6-8), FVC (mL), percent predicted FVC (%), forced expiratory volume over 1 second (FEV_1, L), FEV_1/FVC, absolute DL_{co} corrected for hemoglobin (DL_{co}/Hb, mL/min/mmHG), percent Predicted DL_{co}/Hb (%), total lung capacity (TLC, L), residual volume (RV, L) will be summarized by treatment group and overall.

The baseline characteristics will be summarized for the MITT population only – note that if the safety population differs from the MITT population, then the summaries will be repeated for the safety population.

6.4 Concomitant Therapy

A concomitant therapy is any drug or substance administered between the time a subject is screened for the study and the final study visit.

In order to define concomitant for therapies with missing start or stop date, the following additional criteria are defined:

- If both the start and stop dates of a therapy are missing, then the therapy will be considered concomitant;
- If the start date is missing and the stop date of the therapy falls on/after the screening visit, then the therapy will be considered concomitant;

- If the start date of a therapy is prior to the date of the screening and the stop date of the therapy is missing and the therapy is listed as ongoing, that therapy will be considered concomitant;
- If the start date of a therapy is prior to the date of screening and the stop date of that therapy is missing and the therapy is not listed as ongoing, that therapy will be considered non-concomitant.

For a record with a partial start/end date, the year/month of the partial date will be compared to screening to determine whether it is concomitant.

All concomitant medications will be coded using the World Health Organization (WHO) medication dictionary. All concomitant non-drug treatments will be coded using the MedDRA dictionary.

The number and percent of subjects taking concomitant medication, non-drug treatments will be summarized by treatment group.

6.5 Study Drug Exposure and Study Drug Compliance

Number of weeks on study treatment (BG00011 or placebo), calculated as

Duration of treatment exposure (weeks) = (date of the last dose – date of the first dose +1)/7,

will be summarized as a categorical variable with the following cumulative duration of exposure categories: at least 4, 8, 12, 16, 20, 26, 32, 38, 44, 52 weeks.

In addition, the duration of treatment exposure (weeks) will be summarized as continuous variable with summary statistics.

The first dose of BG00011 or placebo (study treatment) will be administered at the study site during the Baseline Visit. All subsequent doses should be administered by the subject or the subject's caregiver. Compliance with study treatment dosing is to be documented by the subject or the subject's caregiver in the electronic study diary and monitored at each study visit by study site personnel.

Overall compliance for each subject will be assessed based on the dosing recorded in the electronic study diary along with the first dose recorded in the eCRF. The compliance will be calculated and summarized as a continuous variable for each treatment group and background status. The compliance will be calculated using the following formula:

Percent Compliance = (actual number of doses administered)/(number of doses expected to be administered) *100

For all subjects who did not reach week 52 visits whether due to the study early termination and/or treatment discontinuation, "number of doses expected" will be relative to the time of discontinuation.

6.6 Protocol Deviations

Protocol deviations identified during site monitoring will be captured in a Protocol Deviation Log and will be categorized as major or minor deviations based on protocol deviation

classification prior to the final database lock. The major protocol deviations will be summarized and listed. The minor protocol deviations will also be listed. All summaries for protocol deviations will be presented by background therapy status and treatment group.

6.7 Multiplicity Adjustment

No multiple comparison adjustments are planned and all inferential analyses performed will be interpreted nominally.

7 Efficacy Analysis

7.1 General considerations

All efficacy analyses will be performed on the MITT population, unless otherwise specified. For each endpoint, additional conditions may apply to the definition of the population for the analysis. Subjects will be analyzed in the groups to which they were randomized.

7.2 Analysis of Primary Endpoint

7.2.1 Primary Analysis

The estimand of the primary analysis is the mean difference of the change from baseline in FVC (expressed in mL, at each week 26 or week 38) between treatment and placebo group in the MITT population [ICH E9 (R1) Addendum 2014, 2017]. Data will be censored when subject experienced intercurrent event (i.e. death and/or lung transplantation and/or treatment discontinuation) and prematurely discontinue the study treatment. The estimate of this analysis reflects the treatment effect of BG00011 if the drug were taken as directed [de jure efficacy estimand].

Due to the study early termination, the estimand will assess the mean difference of the change from baseline in FVC (mL) at week 26 and 38 based on the data availability and visit times of interest. Both change from baseline in FVC (mL) at weeks 26 and 38 are the continuous secondary efficacy endpoints. If less than 20 subjects (i.e. <20 subjects) had non-missing measurements at either visit time of interest, the mean difference comparison will not be performed at that visit time due to the insufficient data.

A random coefficients linear regression model will be used as the primary analysis where the change from baseline in FVC (mL) is the outcome variable assuming linear decline in lung function over time. The model will include random coefficients for intercept and slope (time), and fixed-effect terms for treatment, randomization stratum (background therapy), age, and baseline FVC along with treatment-by-time interaction and background therapy status-by-time interaction. In this model, time is considered as a continuous variable.

An unstructured variance covariance structure will be used to model the within-patient measurements variability. If the unstructured covariance structure matrix results in a lack of convergence, first-order autoregressive, AR(1), covariance structure will be used. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

If a subject who experiences unscheduled visit(s) for acute exacerbation of IPF and has non-missing FVC measurement(s), those unscheduled FVC measurement(s) will be included in the primary analysis.

In the primary analysis, missing data are assumed to be missing at random (MAR) [Rubin 1976]. For subjects who experienced intercurrent event (i.e. death and/or lung transplantation and/or treatment discontinuation), subject's post intercurrent event data will be censored and then imputed based on individual's estimated rate of worsening of lung function prior to treatment discontinuation, similar to linear extrapolation because of the linear assumption of the statistical model.

7.2.2 Supplementary Analysis

A mixed model of repeated measures (MMRM) model will be used as supplementary analyses to analyse the change from baseline in FVC using the same fixed effects, for treatment, randomization stratum (background therapy), age, and baseline FVC along with treatment-by-time interaction and background therapy status-by-time interaction, as in the random coefficients linear regression model. MMRM assume a nonlinear decline in lung function over time. An unstructured variance covariance structure, same as used in the primary analysis, will be used to model the within-patient measurements variability. If the unstructured covariance structure matrix results in a lack of convergence, first-order autoregressive, AR(1), covariance structure will be used. The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom.

The comparison will be conducted to evaluate the treatment difference in absolute FVC at week 26 and 38 between BG00011 cohort versus placebo cohort, respectively.

- If a subject who experiences unscheduled visit(s) for acute exacerbation of IPF and has non-missing FVC measurement(s), such unscheduled FVC measurement(s) may be included in the analysis by using the follow rule.
- If the unscheduled FVC measurement is between 2 consecutives post-baseline scheduled visits with both FVC measurements non-missing, the unscheduled FVC measurement(s) will not be included in the model. If more than one unscheduled FVC measurements are between the same visit window (between the same 2 consecutives post-baseline scheduled visits), all unscheduled FVC measurements will not be included in the model.
- If the unscheduled FVC measurement is between 2 consecutives post-baseline scheduled visits with one of the two FVC measurements missing, the unscheduled FVC measurement will be included in the model and imputed as the schedule visit with missing FVC value. If more than one unscheduled FVC measurements are between the same visit window, the unscheduled FVC measurement which is closest to the scheduled visit with missing FVC measurement will be included in the model and imputed as the schedule visit with missing FVC value.
- If the unscheduled FVC measurement is between 2 consecutives post-baseline scheduled visits with FVC measurements missing at both visits, the unscheduled FVC measurement will be included in the model and imputed to the scheduled visit which is closer to the unscheduled visit. If the unscheduled visit is in the middle of the 2 schedule visits, the unscheduled FVC measurement will be included in the model and imputed to the earlier scheduled visit. If more than one unscheduled FVC measurements are between the same visit window, the earliest unscheduled FVC measurement and the latest unscheduled FVC in the same visit window will be included in the model and the earliest unscheduled FVC measurement will be imputed to the earlier scheduled visit and the latest unscheduled FVC measurement will be imputed to the later scheduled visit.

Since the study stops early, Table 1 will be applied to define the visit window for FVC measurement at the early termination visit.

7.2.3 Complementary Analysis

The estimand of this complementary analysis is the mean difference of the change from baseline in FVC (expressed in mL over 52 weeks, i.e. mL/year) between treatment and placebo group in the MITT. All observed on-treatment and the retrieved dropout data including those collected after experienced intercurrent event (i.e. death and/or lung transplantation and/or treatment discontinuation) and prematurely discontinue the study treatment will be applied as observed, no imputation will be implemented [ICH E9 (R1) Addendum 2017]. The estimate of this analysis reflects the treatment effect of BG00011 seen as actually taken [*de facto* effectiveness estimand]. All data after intercurrent events are included, the same random coefficients linear regression model described in Section 7.2.1 will be applied. In this complementary analysis, missing data are assumed to be missing at random [Rubin 1976] and no imputation will be applied.

Similar to the primary analysis, the complementary analysis will be evaluated at week 26 and week 38, respectively.

7.3 Analysis of Secondary Efficacy Endpoints

7.3.1 Analysis of Continuous Endpoints

7.3.1.1 Pulmonary Function Tests

The continuous secondary efficacy endpoint of change from baseline in percent predicted FVC, at week 26 and week 38 will be applied using the primary analysis, supplementary analysis and complementary analysis described in Section 7.2.

The continuous secondary efficacy endpoints listed as follows will be applied using MMRM model described in Section 7.2.2. Similarly, if less than 20 subjects (i.e. <20 subjects) had non-missing measurements for any endpoint at a given visit time of interest, the mean difference comparison will not be performed at that visit time for that endpoint due to the insufficient data.

- Change from baseline in DL_{co}, corrected for hemoglobin (DL_{co}/Hb) at weeks 26 and 38, respectively.
- Change from baseline in percent (%) predicted DL_{co} /Hb at weeks 26 and 38, respectively.
- Change from baseline in TLC at weeks 26
- Change from baseline in percent (%) predicted TLC at weeks 26

Because the study stopped early, the following continuous secondary efficacy endpoints could not be evaluated with insufficient data.

- Changes from baseline in percent (%) predicted FVC over 52 weeks
- Changes from baseline in DL_{co}/Hb absolute over 52 weeks.
- Changes from baseline in percent (%) predict DL_{co}/Hb over 52 weeks.
- Changes from baseline in total lung capacity (TLC), as measured by plethysmography, over 52 weeks.

- Changes from baseline in percent (%) predicted TLC over 52 weeks.

All the continuous secondary endpoints listed below will be summarized by descriptive statistics at each scheduled visit time. The measurement and change from baseline will be presented. Plots of mean value with standard error for each of the following continuous secondary endpoints at each visit will also be presented.

- FVC
- Percent (%) predicted FVC
- DL_{co}/Hb
- Percent (%) predicted DL_{co}/Hb
- Total lung capacity (TLC), as measured by plethysmography
- Percent (%) predicted TLC

7.3.1.2 6-Minute Walk Test (6MWT)

6MWT distance and change from baseline in 6MWT distance will be summarized and presented by descriptive statistics at each scheduled visit time.

7.3.2 Analyses of Binary Endpoints

The non-responder imputation (NRI) approach will be applied for all binary endpoints, unless otherwise specified. Subjects who discontinued treatment will be classified as non-responders for all the visits following treatment discontinuation.

The odds ratio (OR), relative risk (RR) and absolute risk reduction (ARR) with their confidence interval and associated p-value for BG00011 compared to placebo will be presented.

7.3.2.1 IPF Acute Exacerbation

The secondary efficacy endpoint, proportion of subjects with at least 1 acute exacerbation (using modified diagnostic criteria for acute IPF exacerbation derived from [Collar 2007]) during the 52 weeks on study, will be analyzed using the binary endpoint analyses described in Section 7.3.2.

7.3.2.2 Pulmonary Function Tests

The secondary efficacy endpoint, number of subjects with absolute decline of 10% predicted in FVC (i.e. $\text{FVC percent predicted}_{\text{baseline}} - \text{FVC percent predicted}_{\text{progression}} \geq 10\%$) over 52 weeks will be analyzed using the binary endpoint analyses described in Section 7.3.2. Subject with absolute decline of 10% predicted in FVC at any scheduled visits will be considered as subject with absolute decline of 10% predicted in FVC regardless whether the percent predicted FVC decline is less than 10% afterward and before the study early termination.

7.3.3 Analysis of Time-to-Event Endpoints

The time-to-event endpoints will be analyzed using survival analyses methods. The Kaplan-Meier curve and number of subjects at risk at each visit will be presented. The proportion of subjects who progressed in each treatment group will be summarized.

Time to event will be analyzed by a cox proportional hazards model with terms for treatment (BG00011 vs. placebo) and randomization stratification factor. A stratified log-rank test will be used to compare the 2 treatment groups using randomization stratus as the stratification factor. The progression will be summarized descriptively for each stratum.

Censoring rules will be applied for all time-to-event endpoints, unless otherwise specified. Subjects who do not experience an event by early termination, will be censored at the date of last assessment during the treatment period. The start date for calculation of day to censor or event will be the date of first dose. If a subject who was not on background therapy at randomization initiate treatment with background therapy during the study, the study will be censored at the time of new background therapy initiation. If a subject who was on background therapy at randomization stop background therapy during the study, the study will be censored at the time of the background therapy stopped.

7.3.3.1 Composite endpoint

The secondary efficacy endpoint, time to progression defined as a composite endpoint will be analyzed using the method described in Section 7.3.3. The composite endpoint is defined by including any of the following events.

- Absolute decline of 10% predicted in FVC (i.e. $\text{FVC percent predicted}_{\text{baseline}} - \text{FVC percent predicted}_{\text{progression}} \geq 10\%$)
- Nonelective hospitalization for respiratory events
- Lung transplantation or death

7.3.3.2 Other Time-to-Event endpoints

The time-to-event types secondary efficacy endpoints listed as follows will be applied only by the Kaplan-Meier curve method where number of subjects at risk at each visit and the proportion of subjects who progressed in each treatment group will be summarized and presented.

- Time to first acute exacerbation (using modified diagnostic criteria for acute IPF exacerbation derived from [Collar 2007]), measured in days.
- Time to death or lung transplantation measured in days.
- Time to all nonelective hospitalizations, measured in days.
- Time to all nonelective respiratory hospitalizations, measured in days.

7.3.4 Analysis of Other Endpoints

7.3.4.1 Number of Occurrences

The secondary efficacy endpoint, number of exacerbations during 52 weeks, defined by using the 2007 modified diagnostic criteria [Collard 2007], a listing will be provided if less than 4 subjects in each treatment group in which each subject experiences more than one exacerbation during the study period. If no fewer than 4 subjects in either treatment group in which each subject experience more than one exacerbation during the study period, a table of summary descriptive statistics will be provided with its associated listing.

7.4 Analysis of Exploratory Efficacy Endpoints

[redacted]

- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

8 Safety Analysis

8.1 General Consideration

All safety analyses will be performed for the safety population unless otherwise specified. No statistical testing will be performed on the safety data.

8.2 Analysis of Adverse Events (AEs)

All AEs will be coded by system organ class and preferred term using the Medical Dictionary for Regulatory Activities (MedDRA) dictionary (Version 22.0). In this study, all AEs will be analyzed based on the principle of treatment emergence. Therefore, whenever an analysis of summary of AEs is mentioned, it is intended that this is in reference to treatment emergence AEs (TEAEs). A TEAE is defined as an AE that has an onset date on or after the date of first dosing, or if it was present prior to the first dose and subsequently worsened until last follow up visit.

To define treatment emergence for AEs with missing start or stop date or time the following additional criteria will be used:

- if both the start and stop dates for an AE are missing, then this AE is considered treatment emergent;
- if the start date for an AE is missing and the stop date/time falls after the first dose date/time, then this AE is considered treatment emergent;
- if the start date for an AE was the same as the first dose date, and the start time was missing, then that event is considered treatment emergent.

For AEs with a partial start date, the year/month of the event date will be compared to that of the first dosing date to determine whether the event is treatment emergent.

Overall TEAE summary table will include the number of subjects with any TEAE, the number of subjects with any TEAE by maximum severity (mild, moderate, severe), the number of subjects with any related TEAE, the number of subjects with Serious adverse events (SAE), the number of subjects with related SAE, the number of subjects with TEAE leading to withdrawal of study drug, the number of subjects with TEAE leading to study withdrawal, and the number of deaths. The overall TEAE summary table will be provided for all subjects in the safety population and for subjects in each randomization stratification factor.

Incidence is defined as the number of subjects who experienced an event and each subject will be counted only once within each category. Therefore, to count the number of subjects who experience TEAEs within each preferred term (PT), a subject experiencing the same TEAE multiple times will only be counted once for that PT. Similarly, if a subject experience multiple TEAEs within the same system organ class (SOC), that subject will be counted only once for that SOC category. For subjects experiencing the same AE more than once within a same SOC and PT, that subject will be counted only once at the maximum severity in the calculation of incidence by severity.

Incidence proportion (i.e. percentage, %) is defined as the number of subjects who experienced an event divided by total number of subjects in the analysis population times 100.

The following incidence tables of TEAEs will be presented. The Incidence and incidence proportion will be provided in the incidence tables. The sorting order for TEAE incidence tables, unless otherwise specified, will be by decreasing frequency order of total column within each category. For the table of TEAEs by SOC and PT sorted by decreasing frequency presented by treatment group, SOCs will be presented in decreasing frequency order of total column, and within each SOC, PTs will be presented in decreasing frequency order of total column.

- TEAEs by SOC and PT sorted by decreasing frequency
- TEAEs by SOC and PT sorted by alphabetical order
- TEAEs at least 5% higher in incidence by SOC and PT for BG00011 compared to placebo
- TEAEs by PT
- TEAEs by PT with an incidence of 5% or more in any treatment group
- Severe TEAEs by SOC and PT
- Severe TEAEs by PT
- TEAEs by maximum severity by SOC and PT
- TEAEs by maximum severity by PT
- Related TEAEs by SOC and PT
- SAE by SOC and PT
- SAE by PT
- Related SAE by SOC and PT
- TEAEs that led to discontinuation of study drug by SOC and PT
- TEAEs that led to withdrawal from study by SOC and PT
- Incidence rates of TEAEs by SOC and PT
- Suspected unexpected serious adverse reactions (SUSAR) by SOC and PT

The following listings will be provided:

- Listing of all TEAEs
- Listing of SAEs
- Listing of TEAEs that led to discontinuation of study drug
- Listing of TEAEs that led to withdrawal from study
- Listing of deaths

8.2.1 Subgroup Analysis of AE

The analysis of AE will be evaluated on pre-defined subgroups which includes:

- Regions (Australia, Asia, Europe, North America, South America)
- Usage of background therapy: Nintedanib or Pirfenidone.

The following incidence tables will be presented by pre-defined subgroups using the same methods as mentioned in Section 8.2.

- SAE by SOC and PT by region
- SAE by PT by region
- SAE by SOC and PT by usage of background therapy
- SAE by PT by usage of background therapy

The region and background therapy (Nintedanib or Pirfenidone) will also be including in the following listings.

- Listing of SAEs
- Listing of AE of exacerbation

8.3 Analysis of Laboratory Tests

The following laboratory safety assessments will be conducted:

- Hematology: complete blood count (CBC) including red blood cell count, white blood cell counts with differentials (neutrophils, lymphocytes, monocytes, eosinophils, and basophils), hemoglobin, hematocrit, mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC).
- Serum chemistry: albumin, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), calcium, chloride, carbon dioxide, creatinine, direct bilirubin, gamma-glutamyl transpeptidase, glucose, lactate dehydrogenase (LDH), phosphorus, potassium, sodium, total bilirubin, total protein
- Urinalysis: including urine protein, glucose, ketones, occult blood, and white blood cells by dipstick, with microscopic examination if indicated

The clinical laboratory parameters include safety assessment listed above will be summarized by treatment group and visit. Descriptive statistics will be presented for continuous variables at each visit by treatment group, including number of subjects, mean, standard deviation, median, 25%, 75% quartiles, minimum, and maximum values. Plots of mean value with standard error for numeric laboratory parameters at each visit will be presented. For categorical endpoints, summary statistics will be presented at each visit by treatment group include number and percentage of subjects with data in each category.

Changes from baseline and percent changes from baseline in quantitative laboratory values will be summarized using descriptive statistics by treatment group and visit.

The clinical laboratory parameters by worst post-baseline grade will be provided for all safety assessment. The worst post-baseline cumulative grade will be provided for the safety assessments for Hematology and Urinalysis.

Individual subject listings of clinical laboratory parameters at each visit will be provided. Values for hematology, serum chemistry, and urinalysis values outside the central laboratory reference ranges will be flagged on the individual subject data listings.

8.3.1 Shift Analysis for Lab Tests

Laboratory abnormalities will be summarized with shift from baseline tables. Each hematology and chemistry value will be flagged as "low", "normal", or "high" relative to the normal ranges of the laboratory that performed the assay, or as "unknown" if no results are available. Each urinalysis value will be flagged as "positive", "negative", or "unknown" if no values are available.

Shift from baseline to high/low status for hematology and chemistry parameters and shifts from baseline to high/positive for urinalysis will be presented. In each summary, the denominator for the percentage is the number of patients at risk for the shift. The number at risk for a shift to low is the number of subjects whose baseline value was not low and who had at least one post-baseline value. The number at risk for a shift to high is the number of subjects whose baseline value was not high and who had at least one post-baseline value. Subjects will be counted only once for each parameter and each type of shift regardless of how many post-dosing assessments had that type of shift. A shift to high includes normal to high, low to high, and unknown to high; a shift to low includes normal to low, high to low, and unknown to low. A shift to positive includes 'negative' to 'positive' and 'unknown' to 'positive'.

8.3.2 Analysis of Clinically Significant Abnormalities

The number and percentage of subjects with clinically significant abnormal values for laboratory parameters will be summarized. A listing of clinically significant abnormal values will also be presented. Clinical significance will be determined for selected parameters based on the criteria listed below in Table 3.

Table 3: Criteria for clinically significant abnormal laboratory value

Parameter name	Unit	Low	High
Hematology			
White blood cells	$x10^9$ cells/L	<3.0	>16
Lymphocytes	$x10^9$ cells/L	<0.8	>12
Neutrophils	$x10^9$ cells/L	<1.5	>13.5
Monocytes	$x10^9$ cells/L	N/A	>2.5
Eosinophils	$x10^9$ cells/L	N/A	>1.6
Basophils	$x10^9$ cells/L	N/A	>1.6
Red blood cells (RBC)	$x10^{12}$ cells/L	≤ 3.5	≥ 6.4
Hemoglobin - Females	g/L	≤ 95	≥ 175
- Males		≤ 115	≥ 190
Hematocrit - Females	%	≤ 32	≥ 54
- Males		≤ 37	≥ 60
Platelet count	$x10^9$ cells/L	≤ 75	≥ 700
Blood Chemistry			

Parameter name	Unit	Low	High
Alanine aminotransferase (ALT)		N/A	> 3 x ULN
Aspartate aminotransferase (AST)		N/A	> 3 x ULN
Alkaline phosphatase (ALP)		N/A	>3 x ULN
Total bilirubin		N/A	>2 x ULN
Blood urea nitrogen (BUN)		N/A	≥10.7 mmol/L
Creatinine		N/A	≥176.8 umol/L
Sodium	mmol/L	≤ 126	≥ 156
Potassium	mmol/L	≤ 3	≥ 6
Chloride	mmol/L	≤ 90	≥ 118
Bicarbonate	mmol/L	≤ 16	≥ 35
Glucose (non-fasting)	mmol/L	≤ 2.2	≥ 9.7
HbA1C	%	N/A	≥ 6.5%
Calcium	mmol/L	≤ 2	≥ 3
Phosphorus	mmol/L	≤ 0.5491	≥ 1.7119
Albumin	g/L	≤ 25	≥ 100
Total protein	g/L	≤ 45	≥ 100

8.4 Vital Signs

Vital sign parameters include oral temperature, systolic blood pressure, diastolic blood pressure, pulse rate, respiration rate and weight. Subjects evaluated will be safety subjects with a baseline and at least one post-baseline assessment for that vital sign. The descriptive statistics for actual values and change from baseline will be summarized at each visit. Plots of mean vital sign values over time will be provided.

The analysis of vital signs will also focus on the incidence of clinically relevant outliers based on the following criteria. The incidence as described in Section 8.2 and percentage of clinically relevant outliers determined by each criterion will be summarized. A listing of all vital sign values will be presented for subjects with any vital sign abnormalities. Clinically relevant abnormal vital signs will be flagged in the listing.

Vital Sign	Criteria to determine clinically relevant abnormal vital signs
Temperature	>38°C and an increase from pre-dose of at least 1°C
Pulse	>100 beat per minute (bpm) and an increase from baseline of more than 20 bpm, or <50 bpm and a decrease from baseline of more than 20 bpm.
Systolic Blood Pressure	>160 mmHg and an increase from baseline of more than 40 mmHg, or <90 mmHg and a decrease from baseline of more than 30 mmHg

Diastolic Blood Pressure	>100 mmHg and an increase from pre-dose of more than 30 mmHg, or <45 mmHg and a decrease from pre-dose of more than 20 mmHg

8.5 Electrocardiogram (ECG)

The ECG data will be obtained directly from the 12-lead ECG traces. These data include the PR interval (milliseconds), QTcB interval (milliseconds), QRS duration (milliseconds), and heart rate (beats/min).

Descriptive statistics will be used to summarize ECG measurements at each visit for each treatment group. Individual subject listings of ECG parameters at each timepoint will also be provided by treatment group.

A shift table of ECG interpretation results (i.e. normal; abnormal, not adverse event; abnormal, adverse event from baseline to post-baseline will be presented at each visit by treatment group. If there is no ECG result at baseline, the subject will be counted in the unknown category. A shift to abnormal will include a shift from "unknown" or "normal" at baseline to "abnormal" post-baseline.

8.6 Physical Examination

Abnormal findings which are noted after subjects received study treatment and are deemed by the investigator as clinically significant will be reported as AEs and included in AE analyses.

8.7 Immunogenicity (Anti-BG00011 Antibodies)

Blood samples for the presence and titer of anti-BG00011 antibodies will be collected at scheduled visits of baseline, weeks 4, 12, 26, 38 and 60 (safety follow-up). The analysis visit should be defined using window in the table below in Table 4. Each blood sample will initially be tested for the presence of anti-BG00011 antibodies using a validated electrochemiluminescence (ECL) method. Confirmed positive samples using the screening assay will be titered and subsequently tested using a validated neutralization assay.

Table 4: Visit Windows for Anti-BG00011 Antibodies

Analysis Visit	Target visit day	Analysis visit window
Baseline (Day 0)	1	Most recent non-missing pre-dose value
Week 4	29	[2,57]
Week 12	85	[58,134]
Week 26	183	[135,225]
Week 38	267	[226,344]

Safety Follow-up (Week 60)	421	• If subject who discontinued treatment due to reasons other than study termination by sponsor: > min (393, study day of each subject's week 52/EOT visit in treatment period) * • If subject who discontinued treatment due to study early termination by sponsor: > min (393, study day of each subject's last dose)

* For subjects who discontinued the treatment due to reasons other than study termination by sponsor will go on to complete all scheduled visits until either week 52 or the study was terminated by sponsor due to the urgent safety measure whichever occurs first. The Week 52/ EOT visit in treatment period is defined as the last visit scheduled for either week 52 or study termination due to the urgent safety measure.

Antibody results will be presented by treatment group and subject in listings. The incidence of antibodies to BG00011 will be summarized over time by treatment group using descriptive statistics.

Since the study early termination, the antibody results will not be included in the Part 1 data of the two-part database lock. The antibody data will be included in the Part 2 data. In order to keep the data integrity, the reconciliation of the antibody samples will be conducted prior to the Part 1 data being cleaned and finalized.

9 Pharmacokinetics (PK) Analysis

BG00011 concentrations in serum will be summarized for PK population using descriptive statistics at the scheduled PK sample collection visits. Number of evaluable subjects, arithmetic mean, standard deviation, median, 25% and 75% quartiles, minimum, maximum, geometric mean, and CV will be presented at each visit.

This study will collect only sparse PK samples; thus, no noncompartmental or compartmental methods will be used to analyze the PK data for presentation in the clinical study report. The mean serum concentrations (+/- SD) will be plotted over time by treatment group on the linear and log linear scale.

No dose-proportionality assessments will be conducted because of the sparse PK data sampling. Atypical drug concentrations (e.g., very low or very high) will be excluded from the analysis, if no apparent explanation exists. Concentration observations will also be removed from the data set if corresponding dosing or sampling times are missing or cannot be reconstructed. Measurements that are below the level of quantification (BLQ) will be set to missing and removed from analysis before statistics are calculated, if the BLQ occurs before the first measurable concentration (such as pre-dose), it should be set to zero. If the value of the concentration is "Non-Determinable", then the concentration value will be set to missing.

A listing of individual serum concentration data will be provided. The BLQ and all atypical drug concentration excluded from the analysis will be flagged in the listing.

[redacted]

[redacted]

[redacted]

11 Reference

Berry, J. D., Miller, R., Moore, D. H., Cudkowicz, M. E., Van Den Berg, L. H., Kerr, D. A., Dong, Y., Ingersoll, E. W., and Archibald, D. The combined assessment of function and survival (CAFS): A new endpoint for ALS clinical trials. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2013; 14(3):162-168.

[redacted]

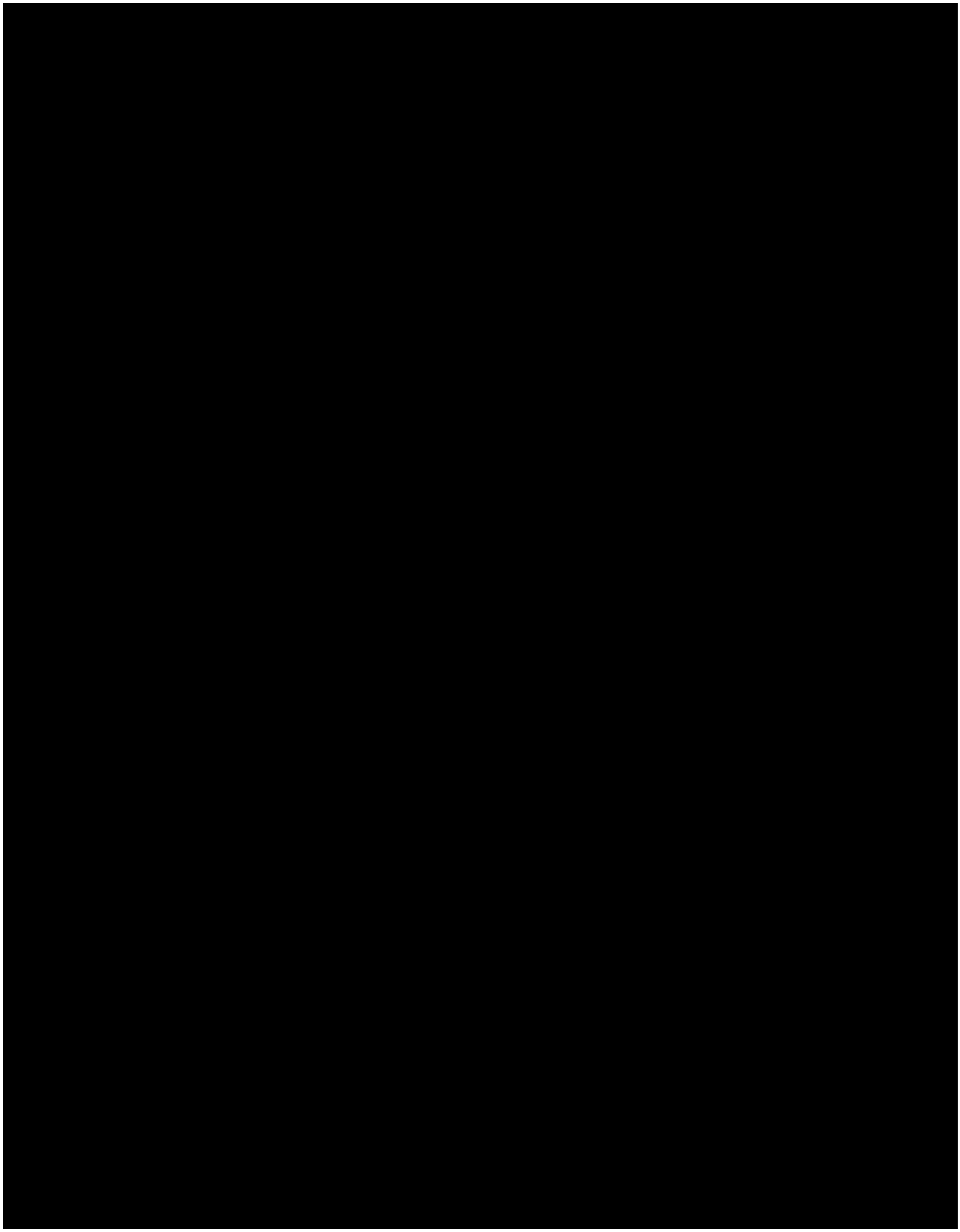
Finkelstein DM, Schoenfeld DA. Combining mortality and longitudinal measures in clinical trials. Stat Med. 1999; 18:1341–54

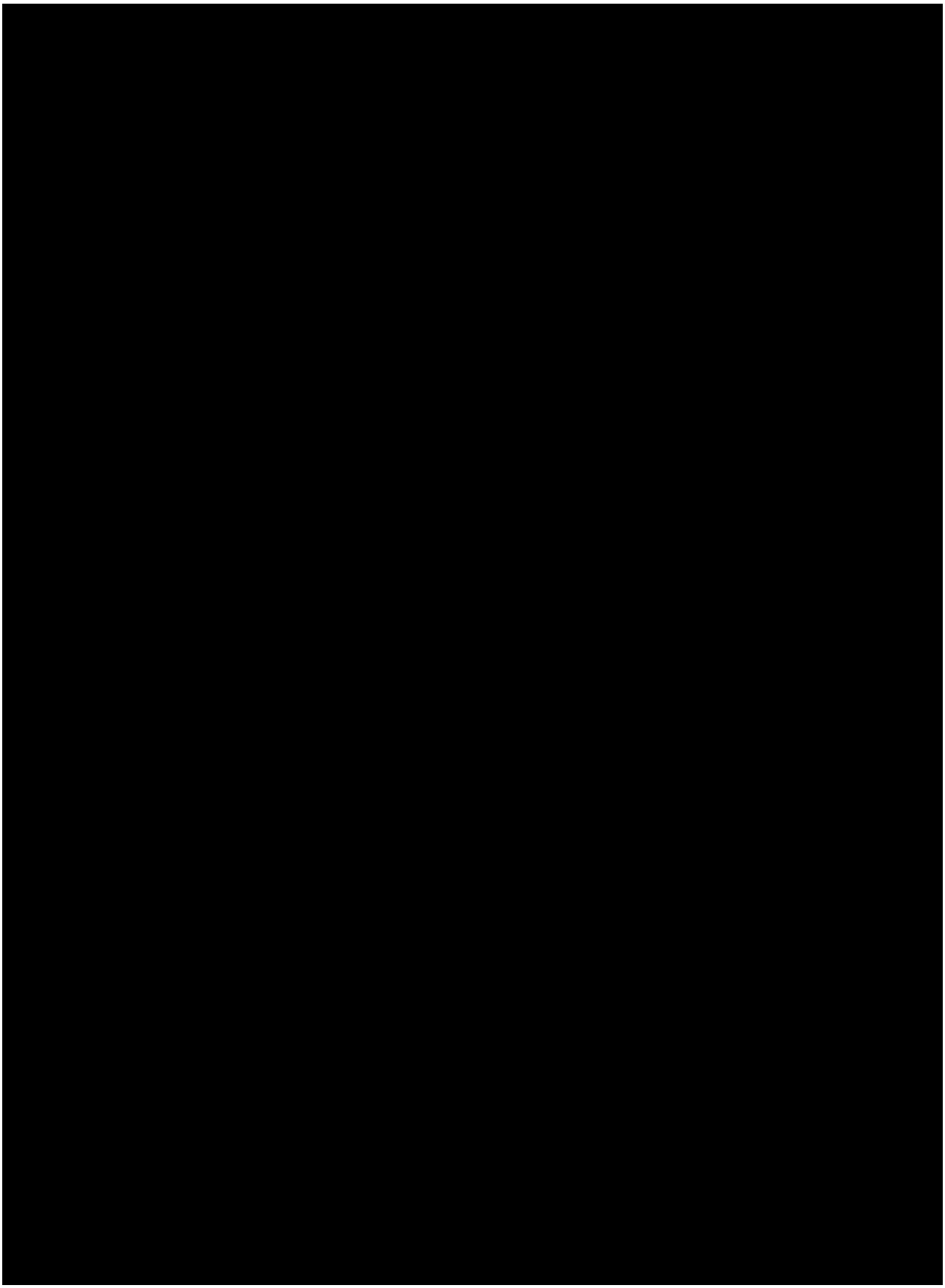
ICH E9 (R1): Addendum to statistical principles for clinical trials on choosing appropriate estimands and defining sensitivity analyses in clinical trials. Endorsed by the ICH Steering Committee. 2014.

ICH E9 (R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials to the Guideline on Statistical Principles for Clinical Trials. European Medicines Agency. 2017.

Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med. 2012;156(10):684-69122586007

Rubin, D. B. (1976) “Inference and missing data.” *Biometrika*, 63(3), 581-592.





[redacted]