Trial Statistical Analysis Plan

c11935728-03

BI Trial No.:	1199.247
Title:	A double blind, randomized, placebo-controlled trial evaluating the efficacy and safety of nintedanib over 52 weeks in patients with Progressive Fibrosing Interstitial Lung Disease (PF-ILD) Including Protocol Amendment 2 [c03736471-03]
Investigational Product:	Nintedanib
Responsible trial statisticians:	[redacted] Boehringer Ingelheim Pharma GmbH & Co. KG Global BDS Birkendorfer Straße 65 88397 Biberach an der Riss Germany Phone: +[redacted] Fax: +[redacted]
Date of statistical analysis plan:	14 DEC 2018 SIGNED
Version:	Final

1. TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

2. LIST OF ABBREVIATIONS

Term	Definition / description
AC	Adjudication Committee
AE	Adverse event
ALK	Alkaline Phosphatase
ALT	Alanine Aminotransferase
AST	Aspartate Aminotransferase
ATS	American Thoracic Society
AZA	Azathioprine
BI	Boehringer Ingelheim
BicDQ	BI Customised Drug Queries
bid	bis in die
BRA	Benefit-Risk Assessment
BRPM	Blinded Report Planning Meeting
CDG	Customized Drug Groupings
COHb	Carboxyhaemoglobin
CT	Computed Tomography
CTD	Connective Tissue Disease
CTP	Clinical Trial Protocol
CTR	Clinical Trial Report
DBL	Database Lock
DLCO	Carbon Monoxide Diffusion Capacity
DMARDs	Disease-Modifying Anti-Rheumatic Drugs
DNA	Desoxyribo Nucleic Acid
eCRF	electronic Case Report Form
EMA	European Medicines Agency
EOT	End of Treatment
ERS	European Respiratory Society (ERS)
FPI	First patient in (= first patient randomised)
FVC	Forced Vital Capacity
g	Gramm
Hb	Haemoglobin
HLT	High Level Term

Term	Definition / description
HRCT	High-Resolution Computer Tomography
ICH	International Conference on Harmonisation
IIP	Idiopathic Interstitial Pneumonia
ILD	Interstitial Lung Disease
INR	International Normalized Ratio
IPD	Important Protocol Deviation
IPF	Idiopathic Pulmonary Fibrosis
IRT	Interactive Response Technology
K-BILD	King's Brief Interstitial Lung Disease Questionnaire
kPa	Kilopascal
L-PF	Living with Pulmonary Fibrosis Symptoms and Impact Questionnaire
LPI	Last patient in (= last patient randomised)
LPLVPE	Last Patient Last Visit Primary Endpoint
m	Meter
MAR	Missing at Random
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligramm
min	Minute
ml	Milliliter
MMF	Mycophenolate Mofetil
mmHg	Millimeters of Mercury
mmol	Millimol
MAR	Missing at random
MCMC	Markov Chain Monte Carlo
MMRM	Mixed effect Model Repeat Measurement
MNAR	Missing not at random
MQRM	Medical Quality Review Meeting
N	Number
NAC	N-Acetylcysteine
OCS	Oral Corticosteroids
PD	Protocol Deviation
PF	Pulmonary Fibrosis

Term	Definition / description
PF-ILD	Progressive Fibrosing Interstitial Lung Disease
PF-IQOLS	Pulmonary Fibrosis Impact on Quality of Life Scale
PK	Pharmacokinetics
pred	Predicted
PT	Preferred Term
REML	Restricted Maximum Likelihood
REP	Residual Effect Period
RS	Randomised Set
SCS	Screened Set
SD	Standard deviation
SE	Standard Error
SEM	Standard Error of the Mean
SMQ	Standardised MedDRA Query
SOC	System organ class
SpO2	Saturation of oxygen
TS	Treated Set
TSAP	Trial Statistical Analysis Plan
TTE	Time-to-event
UIP	Usual Interstitial Pneumonia
ULN	Upper Limit of Normal
WHO-DD	World Health Organization Drug Dictionary

3. INTRODUCTION

As per ICH E9 (1), the purpose of this document is to provide a more technical and detailed elaboration of the principal features of the analysis described in the protocol, and to include detailed procedures for executing the statistical analysis of the primary and secondary variables and other data.

This Trial Statistical Analysis Plan (TSAP) assumes familiarity with the Clinical Trial Protocol (CTP), including Protocol Amendments. In particular, the TSAP is based on the planned analysis specification as written in CTP Section 7 "Statistical Methods and Determination of Sample Size". Therefore, TSAP readers may consult the CTP for more background information on the study, e.g., on study objectives, study design and population, treatments, definition of measurements and variables, planning of sample size as well as randomization. This TSAP follows Boehringer Ingelheims (BI) internal references (2, 3). To support the reader with the understanding of this document, please find a brief summary of the trial setup below.

This placebo-controlled, double-blind clinical trial 1199.247 investigates the efficacy and safety of nintedanib at a dose of 150 mg bid, in patients with Progressive Fibrosing Interstitial Lung Disease (PF-ILD) over 52 weeks. A total of approximately 600 patients will be randomised in a 1:1 ratio between the active treatment arm and the placebo arm. The study population will be enriched for patients with High-Resolution Computer Tomography (HRCT) with Usual Interstitial Pneumonia (UIP)-like fibrotic pattern only (referred to as "HRCT with UIP-like fibrotic pattern only" within this document), applying the HRCT criteria used in the Phase III Idiopathic Pulmonary Fibrosis (IPF) studies for nintedanib and confirmed by independent and blinded central readers, with the aim to have at least 400 patients with "HRCT with UIP-like fibrotic pattern only" and approximately 200 patients with "other HRCT fibrotic patterns".

The overall population and the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" are considered as co-primary populations. The subpopulation of patients with "other HRCT fibrotic patterns" is referred to as complementary population. The primary endpoint of the study – the annual rate of decline in Forced Vital Capacity (FVC) expressed in mL over 52 weeks – will be evaluated in both co-primary populations in a confirmatory manner. Secondary and further endpoints, as well as safety parameters, will also be shown and evaluated in both co-primary populations. For the primary endpoint a Hochberg procedure will be used in order to maintain an overall type 1 error rate of 5%. Statistical significance will be declared if the primary endpoint analyses in both co-primary populations are significant at the two-sided 5% level, or if the analyses in either co-primary population are statistically significant at the two-sided 2.5% level. No other endpoints i.e. secondary or further endpoints will be evaluated in a confirmatory manner.

For each patient, the study will consist of two parts: Part A and Part B. Part A of the study will consist of Visits 2 through 9, which will occur within 52 weeks of randomisation. Following completion of the week 52 visit (Visit 9), patients will continue receiving blinded study medication and have study visits every 16 weeks (Part B) until the end of the trial. See Figure 3: 1.

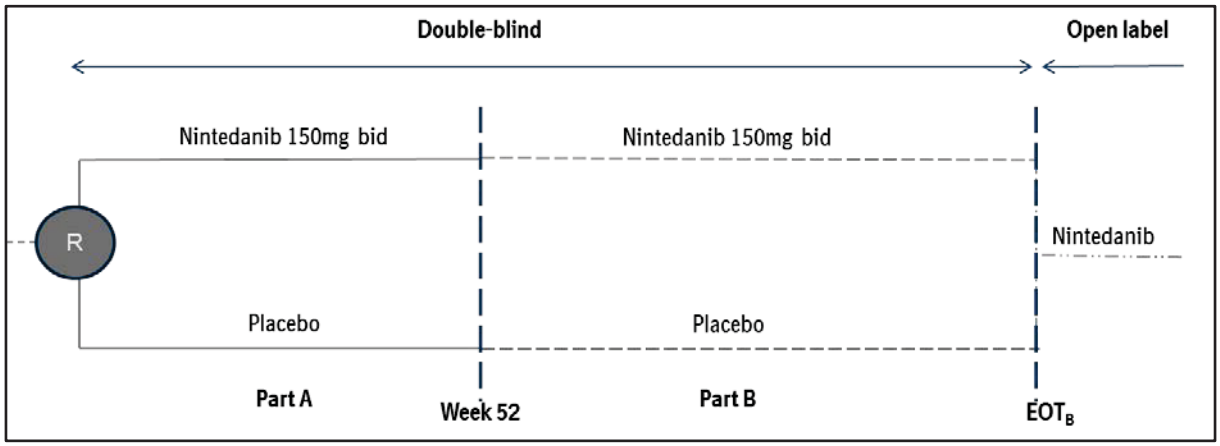
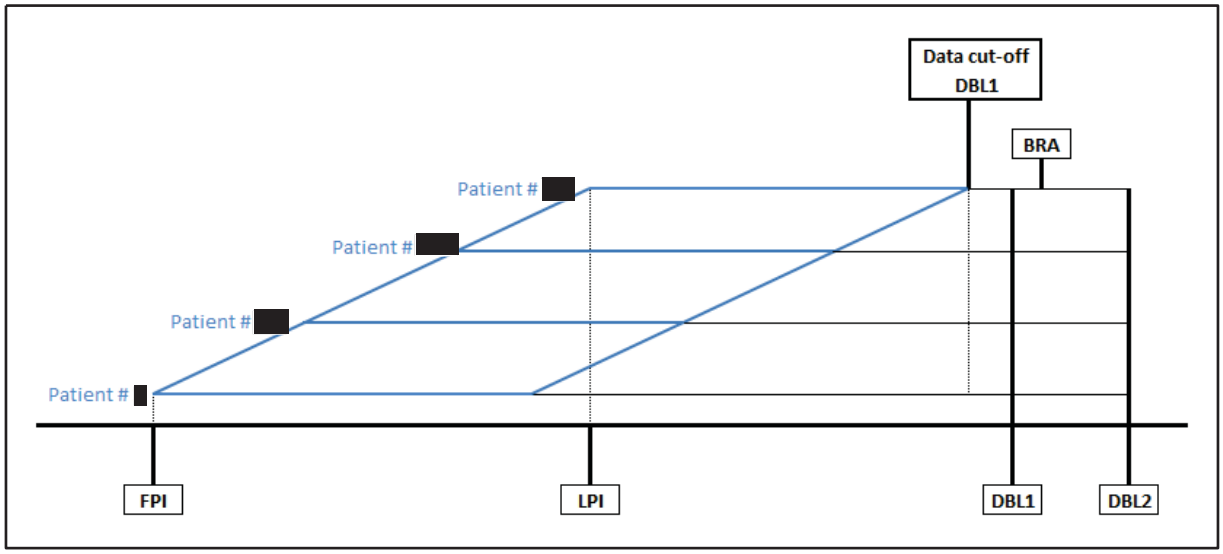


Figure 3: 1 Study design schematic of study 1199.247 followed by a separate open label extension trial 1199_0248 (optional): Part A is a fixed duration period of 52 weeks, and Part B has variable duration for each patient.



FPI: First Patient In / LPI: Last Patient In / DBL1: Primary Database Lock / BRA: Benefit-Risk Assessment / DBL2: Final Database Lock

Figure 3: 2 Patient level example of study design and trial part duration for an individual patient – including the time point when the benefit-risk analysis occurs (BRA). The blue area represents Part A of the study on patient level, and the subsequent part in the black area Part B.

Figure 3: 2 depicts Part A and Part B of the trial on patient level, and shows that the assessment of benefit-risk (BRA) of nintedanib in PF-ILD will occur when the last randomised patient (LPI) reaches the end of Part A. At that time a database lock will occur (DBL1), and the data will be unblinded. Efficacy and safety analyses will be performed on the data from Part A of the trial to assess the benefit-risk of nintedanib over 52 weeks. In addition, data collected in Part B of the trial (after 52 weeks) and available at the time of data cut-off for DBL1 will provide supportive longer term information for further endpoints and on safety.

Patients who stopped the trial medication prematurely (before week 52) are asked to continue following the visit schedule and come to the sites for assessments, in order to have a dataset for the confirmatory primary analyses that is as complete as possible. Per study design, patients will have a different amount of data collected in Part B of the trial, as the duration of study participation for each patient is different depending on their entry date into the study. This is the reason why evaluations over the whole trial (Part A and B combined) include only descriptive analyses of efficacy parameters, assessments of time-to-event endpoints and safety analyses.

Analyses over the whole trial (Part A and Part B) will be repeated after the final database lock (DBL2) to include newly emerging data in the descriptive analyses of efficacy parameters and to assess further time-to-event efficacy endpoints and selected safety outputs over the whole trial. An overview of all trial endpoints can be found in Section 4.3.

Unless stated otherwise, SAS® Version 9.4 or later will be used for all analyses.

4. CHANGES IN THE PLANNED ANALYSIS OF THE STUDY

4.1 ADDITIONS / NEW ANALYSES

In addition to the evaluations depicted in the CTP, the following evaluations will be added:

- Sensitivity Analyses for the primary endpoint
    - A tipping point analysis for the primary endpoint in both co-primary populations will be implemented. Details can be found in Section 9.12.1.

4.2 CHANGES

The following changes compared to the plans outlined in the CTP will be implemented:

- In Section 7.3.2 (Secondary endpoint analyses) of the CTP it is written that convergence issues observed in the analysis of continuous endpoints via a restricted maximum likelihood (REML) based repeated measures approach would be tackled by applying different covariance structures. The exact wording is: "If this analysis fails to converge, the following covariance structures will be tested in order: heterogeneous toeplitz (TOEPH), toeplitz (TOEP) and autoregressive 1 (AR[1]). The first model to converge will be used." It is not intuitive why a different approach should be used for a secondary or further endpoint compared to what is done when the primary analysis shows similar issues. Hence, the analysis approach for convergence issues for continuous secondary or further endpoints will be modified to follow the one used for the primary analysis.

- In Section 7.3.2 (Secondary endpoint analyses) of the CTP it is written that Time-to-event (TTE) endpoints will be analysed using a Cox proportional hazards model. The exact wording is: "Time-to-event endpoints will be analysed using a Cox proportional hazards model. The equality of the hazard rates will be tested by the Wald test for the treatment effect at the two-sided 5% significance level. The model will include the treatment effect and HRCT pattern (except for the analyses on the co-primary analysis population assessing only patients with HRCT with UIP-like fibrotic pattern only) as covariates. Breslow's method for handling ties will be used. Kaplan-Meier plots by treatment group will also be presented. If the proportion of patients experiencing at least 1 acute ILD exacerbation over 52 weeks is less than 2%, then only frequencies of exacerbations by treatment group will be provided for the main secondary endpoint of time to first acute ILD exacerbation or death over 52 weeks and no statistical analysis will be performed." While the Cox Model is still employed to compute the effect size, i.e. the hazard ratio, a stratified log-rank test will be used to evaluate the effect of nintedanib on TTE endpoints compared to placebo.

4.3 CLARIFICATIONS

The following points warrant further clarification:

- Within this document, "drug intake" refers to the intake of the trial medication – either placebo or nintedanib.

- This study comprises of the following "populations": The "overall" population and the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" are considered to be co-primary populations. The subpopulation of patients with "other HRCT fibrotic patterns" is considered to be the complementary population. Unless stated otherwise in the TSAP, evaluations are only done for the co-primary populations.
- In order to determine whether the results for the primary endpoint are consistent between the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" and the subpopulation of patients with "other HRCT fibrotic patterns" the following steps will be taken
    - The numerical results, including the 95% CI will be inspected visually
    - HRCT pattern will be treated as a subgroup in one sensitivity analysis, and the obtained interaction p-value will be depicted on a forest plot in addition to the results of the "overall" population, the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" and the subpopulation of patients with "other HRCT fibrotic patterns". See Section 7.4.3 for additional details.
- Due to the possible occurrence of mis-stratifications during the randomisation process, the HRCT pattern as reported in the eCRF will be used in all analyses. In case this occurs in more than 3% of all randomised patients, an additional sensitivity analysis is implemented utilising the values used during the randomisation process in the Interactive Response Technology (IRT) system. See Section 7.4.2.2.2 for additional details.
- Two patients were randomised by mistake, whose disease extent was ≤ 10% and whose HRCT pattern was not determined by the central readers. Both patients were randomised under "other HRCT fibrotic patterns". Given that the HRCT pattern was actually not determined, those patients do not have an entry for this variable in the eCRF. They will be counted in the subpopulation of patients with "other HRCT fibrotic patterns" in all analyses within the CTR except for the sensitivity analysis implemented to evaluate the mis-stratifications – there they will be counted in the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only". See Section 6.6 for further details.
- This study is only powered and adequately sized to evaluate the efficacy of nintedanib as determined by the primary endpoint analysis. To avoid confusion and misinterpretation, the depiction of p-values in the outputs will be limited to the following evaluations:
    - Primary analysis performed in the "overall population" and the co-primary population of patients with "HRCT with UIP-like fibrotic pattern only"
    - Sensitivity analyses of the primary endpoint analysis in both co-primary populations
    - Nominal p-values for the evaluation of the main secondary endpoints, evaluated in both co-primary populations

  - Interaction p-values for the efficacy evaluations in the subgroup analyses (see Section 6.4 for information about the devised subgroups)

- In Section 5.1.3 of the CTP, the following statement was made: "More details will be provided and additional further endpoints may be defined in the trial statistical analysis plan (TSAP)"
  - Details about further endpoints over the whole trial can be found in Section 5.3.2.
  - No additional endpoints apart from those already mentioned in the CTP were defined for this trial.
- Due to the layout of this document, it may not be as straight forward to determine the different endpoints. To help with this and show the order of endpoints for CTR creation, below is an overview of the endpoints as depicted in Sections 5.1.1, 5.1.2 and 5.1.3 of the CTP:
  - Primary Endpoint
    - The primary efficacy endpoint is the annual rate of decline in FVC (expressed in mL over 52 weeks).
  - Secondary Endpoints
    - Main secondary efficacy endpoints
      - Absolute change from baseline in King's Brief Interstitial Lung Disease Questionnaire (K-BILD) total score at week 52
      - Time to first acute Interstitial Lung Disease (ILD) exacerbation or death over 52 weeks
      - Time to death over 52 weeks
    - Other secondary efficacy endpoints
      - Time to death due to respiratory cause over 52 weeks
      - Time to progression (defined as a ≥ 10% absolute decline in FVC % predicted (pred)) or death over 52 weeks
      - Proportion of patients with a relative decline from baseline in FVC % pred of >10% at week 52
      - Proportion of patients with a relative decline from baseline in FVC % pred of >5% at week 52
      - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Symptoms dyspnea domain score at week 52
      - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Symptoms cough domain score at week 52

- Further Endpoints
  - Further efficacy endpoints over 52 weeks (Part A)
    - Time to first non-elective hospitalization or death over 52 weeks
    - Absolute change from baseline in FVC (mL) at week 52
    - Absolute change from baseline in FVC % pred at week 52
    - Proportion of patients with an absolute decline from baseline in FVC % pred of >10% at week 52
    - Proportion of patients with an absolute decline from baseline in FVC % pred of >5% at week 52
    - Absolute change from baseline in DLCO % pred at week 52
    - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Total score at week 52
    - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Impact score at week 52
    - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Symptoms total score at week 52
    - Absolute change from baseline in Living with Pulmonary Fibrosis (L-PF) Symptoms Fatigue domain score at week 52
    - Absolute change from baseline in Pulmonary Fibrosis Impact on Quality of Life Scale (PF-IQOLS) summary score at week 52
  - Further efficacy endpoints over the whole trial (Part A and Part B)
    - Time to first acute ILD exacerbation or death over the whole trial
    - Time to death over the whole trial
    - Time to death due to respiratory cause over the whole trial
    - Time to progression (defined as a ≥ 10% absolute decline in FVC % pred) or death over the whole trial
    - Time to first non-elective hospitalization or death over the whole trial

5. ENDPOINTS

In this section, more details are given regarding endpoints. Note that for all endpoints and analyses, Section 6.7 should be consulted for baseline value definition. For analyses that are specified over 52 weeks (Part A), unless otherwise specified, all data (including baseline) up to week 52 (inclusive – as defined in Section 6.1) will be taken into account, including follow-up visits before week 52 for patients who discontinue the trial medication prematurely. Analyses over the whole trial (Part A and Part B) reported at the time of the benefit-risk analysis will be based on all data collected until the data cut-off for DBL1. Analyses over the whole trial (Part A and Part B) reported after the final database lock (DBL2) will be based on all data collected within the trial.

Handling of missing data points is described in Section 6.6.

For endpoints where the "date of last contact" is utilised, the following will apply:

- The last contact date when the patient was known to be alive is defined as the latest date recorded in the electronic Case Report Form (eCRF) from the dates listed below (in case a date is planned to be imputed for the analysis, the imputed date will also be used for the definition of "date of last contact"):
    - Date of last visit (e.g. for spirometry), data of last reported Adverse Event (AE) (excluding censored dates), date of last reported concomitant treatment, date of last laboratory sample, date of last drug intake, date of last reported dose change / interruption, last contact date (as documented on the termination of trial eCRF page if the reason for not completing the planned observation period is NOT "Death"), vital status date (from the vital status eCRF page if the patient is known to be alive) and the latest of vital status date / last successful contact date (from the vital status eCRF page if the patient was lost to follow-up).

5.1 PRIMARY ENDPOINT

The primary efficacy endpoint is the annual rate of decline in Forced Vital Capacity (FVC) over 52 weeks expressed in ml. The analysis will be based on FVC values obtained at pre-specified visits over 52 weeks (Part A). Handling of missing data points is described in Section 6.6.1.

5.2 SECONDARY ENDPOINTS

5.2.1 Key secondary endpoints

This section is not applicable as no key secondary endpoint has been specified in the CTP.

5.2.2 Secondary endpoints

The secondary endpoints include change from baseline and time-to-event endpoints at or over 52 weeks. Please note: Additional information about the 52 weeks analysis period can be found in Section 6.1.1, according to the type of endpoint.

5.2.2.1 Main secondary efficacy endpoints

5.2.2.1.1 Absolute change from baseline in King's Brief Interstitial Lung Disease Questionnaire (K-BILD) Total Score at 52 weeks

The K-BILD is a self-administered health status questionnaire that was developed and validated specifically for patients with ILD (12). The questionnaire consists of 15 items and 3 domains: breathlessness and activities, psychological, and chest symptoms. Possible score ranges from 0-100, with a score of 100 representing the best health status.

The K-BILD evaluation will primarily focus on the absolute change from baseline in the total score at week 52. To score the K-BILD, the Likert response scale weightings for individual items are combined, if necessary for certain items, to ensure they detect progressive change in health status. Then the scores are transformed to a range of 0–100 by using logit values and a look-up table to yield interval-level scores, where higher scores indicate better health status. The specific instructions for deriving the K-BILD total score are provided in Section 9.1.

Specific rules of handling of missing items are detailed in Section 6.6.2.1.1.

5.2.2.1.2 Time to death over 52 weeks

Date of death for an individual patient will be obtained from either the AE reporting page for patients with AEs leading to death or from the vital status assessment eCRF page.

For patients with known date of death (regardless of the cause of death) within the first 52 weeks the derivation will be as follows:

- Time to death [days] = Date of death – date of first drug intake + 1

Patients who did not experience any event within the first 52 weeks will be censored according to the mechanism for censoring as described in Table 5.2.2.1.2: 1 below.

Table 5.2.2.1.2: 1 Censoring Rules for Time to death over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient died within the first 52 weeks and date of Death is known	Event	Date of event
2	Patient died within the first 52 weeks and date of Death is unknown	Event	Imputed date of event
3	Patient is alive after 52 weeks	Censored	Day 373 (372 days after first drug intake)
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive

5.2.2.1.3 Time to first acute ILD exacerbation or death over 52 weeks

Acute ILD exacerbation is defined in the CTP as an acute, clinically significant, respiratory deterioration characterized by evidence of new widespread alveolar abnormality with all of the following:

- Previous or concurrent diagnosis of ILD
- Acute worsening or development of dyspnoea typically less than one month duration
- Computed tomography with new bilateral ground-glass opacity and/or consolidation superimposed on a background pattern consistent with fibrosing ILD
- Deterioration not fully explained by cardiac failure or fluid overload

Events that are clinically considered to meet the definition of acute exacerbation but fail to meet all four diagnostic criteria due to missing CT data should be termed "suspected acute exacerbations". Please note that only acute ILD exacerbations that fulfil all the criteria depicted above will be taken into account and contribute to the endpoint. Information on acute ILD exacerbations will be obtained from the Adverse Events page.

For those patients who either experience acute ILD exacerbations or who die due to any cause within the first 52 weeks, time to first acute ILD exacerbation or death [days] will be computed as:

- Earliest of date of first documented acute ILD exacerbation or death – date of first drug intake + 1

Patients who did not experience any ILD exacerbation event within the first 52 weeks will be censored according to the mechanism for censoring as described in Table 5.2.2.1.3: 1.

Table 5.2.2.1.3: 1 Censoring Rules for Time to first acute ILD exacerbation or death over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented acute ILD exacerbation or died within the first 52 weeks and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented acute ILD exacerbation or died within the first 52 weeks and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented acute ILD exacerbation and is alive after 52 weeks	Censored	Day 373 (372 days after first drug intake)
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.2.2.2 Other secondary efficacy endpoints

5.2.2.2.1 Time to death due to respiratory cause over 52 weeks

Analysis of time to death due to respiratory cause will be based on the adjudicated cause of death as determined by an independent Adjudication Committee (AC). The AC will review all fatal cases and adjudicate all deaths to either cardiovascular, respiratory or other causes.

For those patients who died, and whose death is attributed to respiratory causes within the first 52 weeks by the AC, time to death due to respiratory cause [days] will be computed as:

- Date of death – date of first drug intake + 1

Patients who did not experience any event within the first 52 weeks will be censored according to the mechanism for censoring as described in Table 5.2.2.2.1: 1.

Table 5.2.2.2.1: 1 Censoring Rules for Time to death due to respiratory cause over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient died within the first 52 weeks due to respiratory cause and the date of the event is known	Event	Date of event
2	Patient died within the first 52 weeks due to respiratory cause and the date of the event is unknown	Event	Imputed date of event
3	Patient is alive after 52 weeks	Censored	Day 373 (372 days after first drug intake)
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive

5.2.2.2.2 Time to progression or death over 52 weeks

Progression is defined as the date when ≥ 10% of absolute decline in FVC % pred compared to baseline occurs for the first time.

For those patients who fulfil the criteria for progression or death due to any cause within the first 52 weeks, time to progression or death due to any cause [days] will be computed as:

- Earliest of date of documented progression event or death – date of first drug intake + 1

Patients who did not experience any event within the first 52 weeks will be censored according to the mechanism for censoring as described in Table 5.2.2.2.2: 1.

Table 5.2.2.2.2: 1 Censoring Rules for Time to progression or death over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented progression or died within the first 52 weeks and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented progression or died within the first 52 weeks and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented progression and is alive after 52 weeks	Censored	Day 373 (372 days after first drug intake)
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.2.2.2.3 Relative decline from baseline in FVC % pred at 52 weeks

Patients will be categorised based on the relative decline from baseline in FVC % pred greater than 5% or greater than 10% at week 52. The categorisations will be handled as separate endpoints, and the proportion of patients with a decline of greater than 5% as well as the proportion of patients with a decline of greater than 10% will be evaluated separately between the two treatment groups. Please note that patients who fulfil the criterion of a relative decline of greater than 10% also automatically fulfil the criterion of a relative decline of greater than 5%, and will therefore be counted in both endpoint evaluations. Specific rules of handling of missing items are detailed in Section 6.6.2.3.

5.2.2.2.4 Change from baseline in Living with Pulmonary Fibrosis Symptoms and Impact Questionnaire (L-PF) Dyspnea and Cough Domain Scores at week 52

The Living with Pulmonary Fibrosis (L-PF) questionnaire is a 44 item questionnaire with two modules: 1) symptoms (23 items) and 2) impacts (21 items). L-PF was developed with the input of patients with pulmonary fibrosis (PF) and thus is intended to capture perceptions specific to PF patients. The Symptoms module yields three domain scores: 1) dyspnea, 2) cough and 3) fatigue as well as a total Symptoms score. The Impacts module yields a single Impacts score. Symptoms and Impacts scores are summed to yield a total L-PF score. Scoring is performed as a summary score, the mean of the dimension ratings multiplied by 100. Summary score range from 0-100, the higher the score the greater the impairment.

The absolute change from baseline in the L-PF symptoms dyspnea domain score at week 52 as well as the absolute change from baseline in the L-PF symptoms cough domain score at week 52 will be computed and evaluated separately. Scoring instructions for L-PF are described in Section 9.2. Specific rules of handling of missing items are detailed in Section 6.6.2.1.2.

5.3 FURTHER ENDPOINTS

This section includes endpoints over 52 weeks (Part A) as well as further endpoints over the whole trial (Part A and Part B).

5.3.1 Endpoints over 52 weeks

Please note: Additional information about the 52 weeks analysis period can be found in Section 6.1.1, according to the type of endpoint.

5.3.1.1 Time to first non-elective hospitalisation or death over 52 weeks

Time to non-elective hospitalization assessment will be based on the date of hospitalization collected on a specific hospitalization eCRF page. The eCRF page will capture whether the non-elective hospitalization was due to respiratory cause, and the primary admission diagnosis according to the investigator. Cause of non-elective hospitalizations will not be adjudicated by the AC.

For those patients who either experience a non-elective hospitalisation or who die due to any cause within the first 52 weeks, time to first non-elective hospitalisation or death [days] will be computed as:

- Earliest of date of first non-elective hospitalisation or death – date of first drug intake + 1

Patients who did not experience any event within the first 52 weeks will be censored according to the mechanism for censoring as described in Table 5.3.1.1: 1.

Table 5.3.1.1: 1 Censoring Rules for Time to first non-elective hospitalisation or death over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented non-elective hospitalisation or died within the first 52 weeks and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented non-elective hospitalisation or died within the first 52 weeks and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented non-elective hospitalisation and is alive after 52 weeks	Censored	Day 373 (372 days after first drug intake)
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.3.1.2 Absolute change from baseline in FVC at week 52

The absolute change from baseline in FVC at week 52 will be computed and shown. The evaluation will be done by assessing the absolute change in FVC in ml between week 52 and baseline; and in addition to that, by assessing the absolute change from baseline in FVC in % pred between week 52 and baseline. The changes in these two parameters between the two treatment groups will be depicted and described.

Specific rules of handling of missing items are detailed in Section 6.6.2.1.

5.3.1.3 Absolute decline from baseline in FVC % pred at week 52

Patients will be categorised based on the absolute decline from baseline in FVC % pred greater than 5% or greater than 10% at week 52. The categorisations will be handled as separate endpoints, and the proportion of patients with a decline of greater than 5% as well as the proportion of patients with a decline of greater than 10% will be evaluated separately between the two treatment groups. Please note that patients who fulfil the criterion of an absolute decline of greater than 10% also automatically fulfil the criterion of an absolute decline of greater than 5%, und will therefore be counted in both endpoint evaluations. Specific rules of handling of missing items are detailed in Section 6.6.2.3.

5.3.1.4 Absolute change from baseline in Carbon Monoxide Diffusion Capacity (DLCO) % pred over 52 weeks

Single-breath DLCO measurements will be carried out according to the American Thoracic Society (ATS) / European Respiratory Society (ERS) guideline on DLCO measurements (13) DLCO and the corresponding alveolar volume will be measured at time points given in the Flowcharts in the CTP.

DLCO values will be adjusted for altitude, carboxyhaemoglobin (COHb) and the most recent haemoglobin value. For predicted normal values of DLCO, different sites may use different prediction formulas, based on the method used to measure DLCO. In any case, the calculation method used must be in compliance with the ATS/ERS guideline on DLCO measurements and the prediction formula appropriate for that method.

For analyses DLCO % pred at week 52 will be computed as follows:

- (mean DLCO) / (DLCO predicted) × 100%,

where mean DLCO is the mean of the two valid measurements entered in the eCRF (in case only one valid measurement then this one will be used) and

- DLCO predicted (men) = (11.114 × height [m]) – (0.066 × age [years]) – 6.03
- DLCO predicted (women) = (8.176 × height [m]) – (0.049 × age [years]) – 2.746

The value of DLCO % pred will always be presented after correction for Haemoglobin. Percent predicted DLCO corrected for haemoglobin (Hb) expressed in g/dL (13) can be calculated as:

- For men:

$$\text{Percent predicted DLCO corrected for Hb} = \frac{\text{Percent predicted DLCO} \times (10.22 + Hb)}{1.7Hb}$$

- For women:

$$\text{Percent predicted DLCO corrected for Hb} = \frac{\text{Percent predicted DLCO} \times (9.38 + Hb)}{1.7Hb}$$

Please note that different units for DLCO are dealt with the following way:

- If unit of DLCO measured value is given as "mmol/min/kPa", then the value of DLCO is used for the computations
- If unit of DLCO measured value is given as "ml/min/mmHg", then the value of DLCO is divided by 2.987 before it is used in the computations

The absolute change from baseline in DLCO % pred at week 52 will be depicted and compared between the two treatment groups.

5.3.1.5 Living with Pulmonary Fibrosis Symptoms and Impact Questionnaire (L-PF)

A more detailed description of the L-PF can be found in Section 5.2.2.2.4 or in the CTP.

The absolute change from baseline in the L-PF total score, the L-PF impact score, the L-PF symptoms total score and the L-PF symptoms fatigue domain score all at week 52 will be computed and evaluated. Scoring instructions for L-PF are described in Section 9.2. Specific rules of handling of missing items are detailed in Section 6.6.2.1.2.

5.3.1.6 Change from baseline in Pulmonary Fibrosis Impact on Quality of Life Scale (PF-IQOLS) summary score at week 52

The PF-IQOLS scale was developed to measure the impact of a disease and its treatment on the patient's quality of life. The IQOLS includes 16 dimensions and a 5-point standard Likert-type response scale on which patients rate the impact of the disease on their quality of life (14, 15). Scoring of IQOLS is performed as a summary score, the average of the individual dimension ratings . Summary scores range from 1.0-5.0 with greater scores reflecting a worse quality of life.

The absolute change from baseline in the PF-IQOLS summary score at week 52 will be computed and evaluated between the two treatment groups. Scoring instructions for PF-IQOLS are described in Section 9.3. Specific rules of handling of missing items are detailed in Section 6.6.2.1.3.

5.3.2 Endpoints over the whole trial

Part B of the trial (a variable treatment period beyond 52 weeks on patient level) will start after Part A has concluded, and is conducted in order to collect supportive longer term data on the effect of nintedanib in patients with PF-ILD in a controlled manner, especially for the early enrollers. Due to the varying length of follow-up in Part B of the trial, the efficacy measures incorporating data from Part A and Part B will focus on time-to-event endpoints.

5.3.2.1 Time to death over the whole trial

Date of death for an individual patient will be obtained from either the AE reporting page for patients with AEs leading to death or the information from the vital status assessment page. The start date of time at risk is the date of first drug intake, and, in general, patients who did not die during the trial will be censored at the last known date when the patient was alive.

For patients with known date of death (regardless of the cause of death) at any point within the trial, the derivation will be as follows:

- Time to death = Date of death – date of first drug intake + 1

Patients who did not experience any event during their trial participation will be censored according to the mechanism for censoring as described in Table 5.3.2.1: 1.

Table 5.3.2.1: 1 Censoring Rules for Time to death over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient died and date of Death is known	Event	Date of event
2	Patient died and date of Death is unknown	Event	Imputed date of event
3	Patient is alive	Censored	Date of last contact when the patient was known to be alive
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive

5.3.2.2 Time to first acute ILD exacerbation or death over the whole trial

For those patients who either experience acute ILD exacerbations or who die due to any cause at any point in the study, time to first acute ILD exacerbation or death [days] will be computed as:

- Earliest of date of first documented acute ILD exacerbation or death – date of first drug intake + 1

Patients who did not experience any event during their trial participation will be censored according to the mechanism for censoring as described in Table 5.3.2.2: 1.

Table 5.3.2.2: 1 Censoring Rules for Time to first acute ILD exacerbation or death over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented acute ILD exacerbation or died and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented acute ILD exacerbation or died and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented acute ILD exacerbation and is alive	Censored	Date of last contact when the patient was known to be alive and event free
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.3.2.3 Time to first non-elective hospitalisation or death over the whole trial

Time to non-elective hospitalization assessment will be based on the date of hospitalization collected on a specific hospitalization eCRF page.

For those patients who either experience a non-elective hospitalisation or die due to any cause within the trial, time to first non-elective hospitalisation or death [days] will be computed as:

- Earliest of date of first non-elective hospitalisation or death – date of first drug intake + 1

Patients who did not experience any event during their trial participation will be censored according to the mechanism for censoring as described in Table 5.3.2.3: 1.

Table 5.3.2.3: 1 Censoring Rules for Time to first non-elective hospitalisation or death over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented non-elective hospitalisation or died and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented non-elective hospitalisation or died and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented non-elective hospitalisation and is alive	Censored	Date of last contact when the patient was known to be alive and event free
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.3.2.4 Time to death due to respiratory cause over the whole trial

Analysis of time to death due to respiratory cause will be based on the adjudicated cause of death as determined by an independent Adjudication Committee (AC). For those patients who died, and whose death is attributed by the AC to respiratory causes at any time within the trial, time to death due to respiratory cause [days] will be computed as:

- Date of death – date of first drug intake + 1

Patients who did not experience any event during their trial participation will be censored according to the mechanism for censoring as described in Table 5.3.2.4: 1.

Table 5.3.2.4: 1 Censoring Rules for Time to death due to respiratory cause over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient died due to respiratory cause and the date of the event is known	Event	Date of event
2	Patient died due to respiratory cause and the date of the event is unknown	Event	Imputed date of event
3	Patient is alive	Censored	Date of last contact when the patient was known to be alive
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.3.2.5 Time to progression or death over the whole trial

Progression is defined as the date when ≥ 10% of absolute decline in FVC % pred occurs for the first time. For those patients who progressed or died due to any causes within the trial, time to progression or death due to any cause [days] will be computed as:

- Earliest of date of documented progression event or death – date of first drug intake + 1

Patients who did not experience any event will be censored at the date of last contact when the patient was known to be alive and was known not to have progressed.

Patients who did not experience any event during their trial participation will be censored according to the mechanism for censoring as described in Table 5.3.2.5: 1.

Table 5.3.2.5: 1 Censoring Rules for Time to progression or death over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented progression or died and the date of the event is known	Event	Earliest Date of event
2	Patient had a documented progression or died and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented progression and is alive	Censored	Date of last contact when the patient was known to be alive and event free
4	Patient status is unknown	Censored	Date of last contact when the patient was known to be alive and event free

5.4 OTHER VARIABLES

5.4.1 Demographics and baseline characteristics

5.4.1.1 Demographic data

- Gender
- Ethnicity: overall and by race
- Race: single race respondents, multiple race respondents (all combinations ticked), and all race categories regardless of how many race categories were ticked
- Age [years] at time of informed consent (transferred by IRT)
- Age in categories [years] (<30; ≥30-<45; ≥45 and <60;≥60 and <75; ≥75)

- Weight [kg] as continuous variable and in classes (<30; ≥30 - <60; ≥60 - <90; ≥90)
- Height [cm]
- Body mass index [kg/m^2]: Weight[kg] / Height [m]*Height[m], as a continuous variable and in classes (< 18.5; ≥18.5-<25; ≥25-<30, ≥30)
- Tobacco consumption (Never, Current, Former)
- Pack years (for patients who tick *"Current" or "Former"* in the variable "Tobacco consumption") as continuous variable and in classes (< 20; ≥20-<40; ≥40)

5.4.1.2 Trial indication

- Time elapsed since the first diagnosis of ILD [years] will be calculated as: (date of randomisation – date of first diagnosis of ILD) / 365.25
- Time elapsed since the first diagnosis of ILD will also be coded in classes (≤ 1 year; > 1 year to ≤ 3 years; > 3 years to ≤ 5 years; > 5 years)
- Underlying ILD diagnosis ("*Idiopathic nonspecific interstitial pneumonia", "Unclassifiable idiopathic interstitial pneumonia", "Hypersensitivity pneumonitis", "Rheumatoid Arthritis associated ILD", "Mixed connective tissue disease", "Systemic sclerosis associated ILD","Exposure-related ILD", "Sarcoidosis", "Other fibrosing ILD"*)
- Underlying ILD diagnosis in groups
  - *Hypersensitivity pneumonitis (HP)*
  - *Idiopathic nonspecific interstitial pneumonia (iNSIP)*
  - *Unclassifiable idiopathic interstitial pneumonia (uIIP)*
  - *Autoimmune ILDs (auto)*
  - Other ILDs (other)
- Autoimmune ILD (Sub-classification based on the underlying ILD diagnosis categories *"Rheumatoid Arthritis associated ILD", "Mixed connective tissue disease" and "Systemic sclerosis associated ILD" and the classification of "Other fibrosing ILDs" shown in Table 9.13: 1 with the flag "autoimmune ILD"*)
- Exposure-related ILD (Sub classification based on the underlying ILD diagnosis category *"Exposure related ILD" and the classification of "Other fibrosing ILDs" shown in Table 9.13: 1 with the flag "exposure related ILD"*)
- Other ILDs (Sub-classification based on the underlying ILD diagnosis category "Sarcoidosis", exposure-related ILDs and the classification of "Other *fibrosing ILDs" shown in Table 9.13: 1 with the flag "other ILD"*)
- ILD history including:
  - Time since ILD diagnosis based on imaging [years]
  - Diagnosis of ILD confirmed by surgical biopsy (Yes, No, Unknown)
  - Diagnosis of ILD confirmed by transbronchial biopsy (Yes, No, Unknown)
  - For auto-immune ILD (only): Time since rheumatologic diagnosis [years]

- For auto-immune ILD (only): Diagnosis confirmed by a rheumatologist? (Yes, No, Unknown)
- For exposure related ILDs (only): exposure still present (Yes, No)

- Criteria for progressive ILD (all combinations of: Decline in FVC % pred ≥10%, Decline in FVC % pred ≥5-<10% combined with worsening of respiratory symptoms, Decline in FVC % pred ≥5-<10% combined with increasing extent of fibrotic changes on chest imaging, Worsening of respiratory symptoms as well as increasing extent of fibrotic changes on chest imaging)
- HRCT assessment results:
  - HRCT pattern ("HRCT with UIP-like fibrotic pattern only", "Other HRCT fibrotic patterns")

5.4.1.3 Baseline characteristics for lung function

- FVC [ml]
- FVC [% predicted]
- FEV1 / FVC
- Oxygen Saturation on Pulse Oximetry (SpO2) [%]
- DLCO (corrected for Hb) [% predicted]: refer to Section 5.3.1.4 for the derivation of DLCO % pred.

5.4.1.4 Baseline characteristics for questionnaires and derived outcomes

- K-BILD breathlessness and activities score, psychological score, chest symptoms score, and total score
- L-PF
  - Total score
  - Impact score
  - Symptoms dyspnea, cough, and fatigue domain scores and total score
- PF-IQOL summary score

5.4.2 Compliance

5.4.2.1 Over 52 weeks

Compliance will be calculated over 52 weeks as:

$$\text{Compliance [\%]} = \frac{\begin{array}{c}\textit{Number of capsules actually taken}\\ \textit{over the } 52-\textit{week treatment period}\end{array}}{\begin{array}{c}\textit{Number of capsules which should have been taken}\\ \textit{over the } 52-\textit{week treatment period}\end{array}} x100$$

The number of capsules which should have been taken over the 52-week treatment period will be calculated as:

Number of capsules which should have been taken over the 52-week treatment period [capsules] = (date of last administration [1] – date of first administration + 1) [days] × 2 [2]

[1] Or Visit 9, if earlier.

[2] The time of first/ last administration are taken into account, so only one capsule may be expected to be taken on some days (calculation is then adapted accordingly).

In case of dose reduction to 100 mg bid, the number of capsules taken per day is still 2, so the calculation of compliance remains the same. Only treatment interruptions not due to AE will be considered as a compliance issue and will be taken into account in the calculation (duration of interruptions due to AE will be subtracted from the duration of the treatment period, as defined in Section 6.1).

Compliance will also be categorised into classes: <50%, ≥50% - <80%, ≥80% - ≤ 120%, > 120%.

5.4.2.2 Over the whole trial

Similarly, compliance over the whole trial will be calculated and categorised into the same classes.

5.4.3 Exposure

5.4.3.1 Over 52 weeks

Duration of exposure [months] = (date of last administration[1] – date of first administration +1 day) / 30.5

[1] Or week 52 time-point (372 days after date of first trial drug intake), if earlier.

Duration of exposure in categories: ≤3 months (91 days); >3 months (91 days) to ≤ 6 months (182 days); > 6 months (182 days) to ≤12 months (365 days); > 12 months (365 days)

Treatment interruptions will not be subtracted from the duration of exposure.

Duration on actual dose (100 or 150 mg bid) (sum of durations on-treatment for each dose effectively taken): in weeks and in categories (≤ 8 weeks, > 8 weeks)

Total dose [g]: Duration of exposure [days] * actual dose [g]

Dose intensity [%]: amount of drug actually administered over the study (dose 100mg and 150mg) divided by the amount of drug that would have been administered had dose 150mg bid been administered over all the study (from date of first administration to date of last administration, whether or not trial drug was prematurely discontinued). Dose intensity will be summarized in percent and in categories (≤ 30 %, >30% - ≤50%, >50% - ≤90%, >90% - <100%, 100%).

Time to first dose reduction or time to first treatment interruption in classes : ≤ 31 days, > 31 − ≤ 91 days, > 91 - ≤182 days, > 182 days

Time to first dose reduction, time to first treatment interruption, as well as time to premature treatment discontinuation will also be evaluated via survival methods. Censoring rules can be found in Table 5.4.3.1: 1, Table 5.4.3.1: 2 and Table 5.4.3.1: 3 respectively.

Duration of exposure in patient years will also be shown per treatment group.

Table 5.4.3.1: 1 Censoring Rules for Time to first dose reduction over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first dose reduction within the first 52 weeks or before the treatment was prematurely[1] discontinued and the date of the event is known	Event	Date of event
2	Patient had a documented first dose reduction within the first 52 weeks or before the treatment was prematurely[1] discontinued and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first dose reduction within the first 52 weeks or before the treatment was prematurely[1] discontinued	Censored	Earliest of Day 373 (372 days after first drug intake) or date of permanent premature[1] drug discontinuation

[1] Prematurely: Discontinuation of the treatment before completion of Visit 9.

Table 5.4.3.1: 2 Censoring Rules for Time to first treatment interruption over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first treatment interruption within the first 52 weeks or before the treatment was prematurely[1] discontinued and the date of the event is known	Event	Date of event
2	Patient had a documented first treatment interruption within the first 52 weeks or before the treatment was prematurely[1] discontinued and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first treatment interruption within the first 52 weeks or before the treatment was prematurely[1] discontinued	Censored	Earliest of Day 373 (372 days after first drug intake) or date of permanent premature[1] drug discontinuation

[1] Prematurely: Discontinuation of the treatment before completion of Visit 9.

Table 5.4.3.1: 3 Censoring Rules for Time to premature treatment discontinuation over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented premature treatment discontinuation within the first 52 weeks and the date of the event is known	Event	Date of event
2	Patient had a documented premature treatment discontinuation within the first 52 weeks and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a premature treatment discontinuation within the first 52 weeks	Censored	Day 373 (372 days after first drug intake)

5.4.3.2 Over the whole trial

Similarly, duration of exposure, duration on actual dose, total dose and dose intensity will be summarized over the whole trial. Time to first dose reduction and time to first interruption will also be computed and shown.

Over this period, the categories for the duration of exposure will be: ≤3 months (91 days); >3 months (91 days) to ≤ 6 months (182 days); > 6 months (182 days) to ≤12 months (365 days); > 12 months (365 days) to ≤ 18 months (547 days); > 18 months (547 days) to ≤ 24 months (730 days); > 24 months (730 days)

Time to first dose reduction, time to first treatment interruption, as well as time to premature treatment discontinuation will also be evaluated via survival methods. Censoring rules can be found in Table 5.4.3.2: 1, Table 5.4.3.2: 2 and Table 5.4.3.2: 3 respectively.

Duration of exposure in patient years will also be shown per treatment group.

Table 5.4.3.2: 1 Censoring Rules for Time to first dose reduction over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first dose reduction before the treatment was permanently discontinued or the patient switched into the extension trial and the date of the event is known	Event	Date of event
2	Patient had a documented first dose reduction before the treatment was permanently discontinued or the patient switched into the extension trial and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first dose reduction before the treatment was permanently discontinued or the patient switched into the extension trial	Censored	Earliest date of permanent drug discontinuation, database lock or date of Informed Consent in extension trial[1]

[1] Date of Informed Consent in extension trial is not applicable for the primary analysis.

Table 5.4.3.2: 2 Censoring Rules for Time to first treatment interruption over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first treatment interruption before the treatment was permanently discontinued or the patient switched into the extension trial and the date of the event is known	Event	Date of event
2	Patient had a documented first treatment interruption before the treatment was permanently discontinued or the patient switched into the extension trial and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first treatment interruption before the treatment was permanently discontinued or the patient switched into the extension trial	Censored	Earliest date of permanent drug discontinuation, database lock or date of Informed Consent in extension trial[1]

[1] Date of Informed Consent in extension trial is not applicable for the primary analysis.

Table 5.4.3.2: 3 Censoring Rules for Time to premature treatment discontinuation over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented premature treatment discontinuation and the date of the event is known	Event	Date of event
2	Patient had a documented premature treatment discontinuation and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented premature treatment discontinuation and the patient switched into the extension trial	Censored	Earliest date of database lock or date of Informed Consent in extension trial[1]

[1] Date of Informed Consent in extension trial is not applicable for the primary analysis.

5.4.4 Liver enzyme and bilirubin elevations

Liver enzyme and bilirubin elevations will be reported using the three following definitions:

- (ALT and/or AST ≥ 3xULN) AND bilirubin ≥ 2xULN*
- ALT ≥ 5xULN and/or AST ≥ 5xULN
- ALT ≥ 3xULN and/or AST ≥ 3xULN

*These elevations of ALT/AST/bilirubin are defined within a time window of 30 days i.e. the elevation of bilirubin should appear within 30 days after the elevation of AST and/or ALT.

Depending on the number of patients experiencing liver enzyme elevations, other cut-offs may be considered.

Signs of hepatic injury will also be presented based on the following definition:

- ALT and/or AST ≥8xULN
- ALT and/or AST ≥3xULN and unexplained INR > 1.5 in the same sample
- ALT and/or AST ≥3xULN and unexplained eosinophilia (>5%) in the same sample
- ALT and/or AST ≥3xULN and presence of fatigue, nausea, vomiting, right upper abdominal quadrant pain or tenderness, fever and/or rash (see Section 9.9 for details about MedDRA terms used to select relevant AEs) within +/- 7 days of the detected ALT or AST increase.

In addition, maximum individual elevations based on worst value on treatment will be defined as:

- ≥3xULN; ≥5xULN; ≥8xULN for AST and ALT and AST and/or ALT
- ≥1.5xULN; ≥2xULN for Bilirubin
- ≥1.5xULN; ≥2xULN for alkaline phosphatase (ALK)

Time to first liver enzyme elevation will also be evaluated via survival methods. Censoring rules can be found in Table 5.4.4: 1 and Table 5.4.4: 2 below.

Table 5.4.4: 1 Censoring Rules for Time to first liver enzyme elevation over 52 weeks

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first liver enzyme elevation within the first 52 weeks or before the treatment was prematurely[1] discontinued and the end of the residual effect period (REP) has been reached, and the date of the event is known	Event	Date of event
2	Patient had a documented first liver enzyme elevation within the first 52 weeks or before the treatment was prematurely[1] discontinued and the end of the REP has been reached, and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first liver enzyme elevation within the first 52 weeks or before the treatment was prematurely[1] discontinued and the end of the REP has been reached.	Censored	If day of last drug intake < 372 days after first drug intake: Date of last drug intake +28 days or date of death (if earlier)[2] If day of last drug intake ≥ 373: Date of first drug intake + 372 days

[1] Prematurely: Discontinuation of the treatment before completion of Visit 9.

[2] Patients that die between date of permanent drug discontinuation and the end of the residual effect period will be censored at their date of death

Table 5.4.4: 2 Censoring Rules for Time to first liver enzyme elevation over the whole trial

Rule #	Situation	Outcome (event or censored)	Date of event or censoring
1	Patient had a documented first liver enzyme elevation before the treatment was permanently discontinued and the date of the event is known	Event	Date of event
2	Patient had a documented first liver enzyme elevation before the treatment was permanently discontinued and the date of the event is unknown	Event	Imputed date of event
3	Patient did not have a documented first liver enzyme elevation before the treatment was permanently discontinued	Censored	Earliest date of database lock, date of permanent drug discontinuation + 28 days or date of death[1][2][3]

[1] Date of permanent drug discontinuation may not have been reached when the primary analysis takes place. In this case, the date of database lock will be used.
[2] Patients who switch to the extension study will only be evaluated until date of last drug intake in this study. The residual effect period will not apply, as patients will receive continuous treatment
[3] Patients that die between date of permanent drug discontinuation and the end of the residual effect period will be censored at their date of death.

5.4.5 Marked changes in vital signs

A marked increase is defined as:

- Systolic Blood Pressure >150 mmHg and increase ≥ 25 mmHg above baseline
- Diastolic Blood Pressure >90 mmHg and increase ≥ 10 mmHg above baseline
- Pulse Rate >100 bpm and increase ≥ 10 bpm above baseline

A marked decrease is defined as:

- Systolic Blood Pressure <100 mmHg and decrease >10 mmHg below baseline
- Diastolic Blood Pressure <60 mmHg and decrease >10 mmHg below baseline
- Pulse Rate <60 bpm and decrease >10 bpm below baseline

6. GENERAL ANALYSIS DEFINITIONS

6.1 TREATMENTS

For the definition of treatment administered during the trial, see Section 4 of CTP.

Note: the last day of each of the periods is excluded from the respective period. It defines the first day of the subsequent period.

- Screening: From informed consent to randomisation
- Post-randomisation (optional[a]): From randomisation to first trial drug intake in treatment period.
- Treatment period: From first trial drug intake (or re-start of treatment if interruption) to last trial drug intake (or the day before start date of interruption if interruption) plus one day
- Off-treatment (optional[a]): From start date of interruption to re-start of treatment
- Residual effect period[b]: From the last trial drug intake plus one day to last trial drug intake plus 28 days plus one day or to date of first trial drug intake in extension trial, whichever occurs earliest
- Follow-up (optional[a][b]): From last trial drug intake plus 29 days up to the beginning of post-study period. This period is only created if last trial drug intake took place more than 28 days before trial completion, or for patients having prematurely discontinued the treatment and still continuing the trial
- Post-study[a]: from the latest between last trial drug intake plus 29 days, 'date of trial completion' (from the Trial Completion page of the eCRF) plus one day and 'date of Informed Consent in extension trial' plus one day (if applicable). This period is not created if date of first trial drug intake in extension trial is before last trial drug intake plus 28 days

[a] This period is optional insofar as it does not necessarily exist for all patients.
[b] In addition, a residual effect period of 7 days will be used for safety analyses to more closely reflect the period of time after the last trial drug intake when measureable drug levels or pharmacodynamic effects are still likely to be present.

All analyses will be based on the planned treatment group (Placebo or Nintedanib 150mg bid) as randomised by IRT.

6.1.1 Over 52 weeks

For efficacy and safety analyses over 52 weeks (Part A), data from first drug intake up to ≤ 372 days after first drug intake will be considered. See Table 6.1.1: 1 for details about analysis periods for Part A according to the type of endpoint.

Table 6.1.1: 1 Summary of analysis periods over 52 weeks (Part A) according to the type of endpoint or analysis

Type of analysis	Analyses / Endpoints	Studied period	
		Start date[2]	End date [1]
Efficacy analyses over 52 Weeks [2]	• Analyses on all endpoints listed in Sections 5.1.1, 5.1.2 and 5.1.3.1 of the CTP	Date of first drug intake	Date of last measurement before or at 52 Weeks (≤ 372 days after first drug intake)
On-treatment efficacy analyses over 52 Weeks	• Annual rate of decline in FVC in ml over 52 weeks, including only data until end date of studied period	Date of first drug intake	Date of last trial drug intake OR Week 52 time-point (≤ 372 days after first drug intake), whichever occurs first
Descriptive analyses over 52 Weeks	Absolute change from baseline in: • FVC (ml) • FVC (% of predicted) • DLCO (% of predicted) • Vital signs • K-BILD • L-PF • PF-IQOLS Relative change from baseline in : • FVC (% of predicted)	Date of first drug intake	Date of last measurement before or at 52 Weeks (≤ 372 days after first drug intake)
Extent of exposure analysis over 52 weeks	See Section 5.4.3	Date of first drug intake	If day of last drug intake < 372 days after first drug intake • Date of last drug intake Else, if day of last drug intake ≥ Day 373: • Date of first drug intake + 372 days
Safety survival analysis	Time to • first liver enzyme elevation (for all definitions mentioned in Section 5.4.4) • first onset of adverse event with additional information collection (diarrhoea, bleeding, ILD)	Date of first drug intake	If day of last drug intake < 372 days after first drug intake: • Date of last drug intake + 28 days If day of last drug intake ≥ 373: Date of first drug intake + 372 days
	Time to • first dose reduction • first treatment interruption • premature treatment discontinuation	Date of first drug intake	If day of last drug intake < 372 days after first drug intake: • Date of last drug intake If day of last drug intake ≥ 373: • Date of first drug intake + 372 days
On-treatment safety analysis over 52 weeks	• Adverse events • Laboratory data • Vital signs	Date of first drug intake	If day of last drug intake < 372 days after first drug intake: • Date of last drug intake + 28 days If day of last drug intake ≥ 373: • Date of first drug intake + 372 days

[1] End date is included.
[2] Date of randomisation and date of first drug intake should be identical. If there are discordances between these dates on patient level, date of first drug intake will be used.

6.1.2 Over the whole trial

For efficacy and safety analyses over the whole trial period (Part A and Part B), data from first drug intake up to the last visit in the trial will be considered. Please see Table 6.1.2: 1 for details about analysis periods over the whole trial (Part A + Part B) according to the type of endpoint. For the main analysis only data collected over the whole trial available at the time of data cut-off for DBL1 will be considered, whereas all data collected within the trial will be considered and reported for the final analysis.

Table 6.1.2: 1 Summary of analysis periods over the whole trial (Part A and Part B) according to the type of endpoint or analysis

Type of analysis	Analyses / Endpoints	Studied period	
		Start date[2]	End date[1][4]
Efficacy analyses	Analyses on all endpoints listed in Sections 5.1.3.2 of the CTP	Date of first drug intake	Date of last measurement up to the last follow-up visit (included) or last contact date (as defined in Section 5)
Descriptive analyses over the trial period	Absolute change from baseline in: • FVC (ml) • FVC (% of predicted) • DLCO (% of predicted) • Vital signs Relative change from baseline in : • FVC (% of predicted)	Date of first drug intake	Date of last measurement up to last follow-up visit (included)
Efficacy survival analysis	Time to all-cause mortality (over the trial period)	Date of first drug intake	Last contact date (as defined in Section 5)
Extent of exposure analysis overall	See Section 5.4.3	Date of first drug intake	Date of last drug intake
Safety survival analysis [3]	Time to • first liver enzyme elevation (for all definitions mentioned in Section 5.4.4) • first onset of adverse event with additional information collection (diarrhoea, bleeding, ILD)	Date of first drug intake	Date of last drug intake + 28 days
	Time to • first dose reduction • first treatment interruption • premature treatment discontinuation	Date of first drug intake	Date of last drug intake
On-treatment safety analysis overall [3]	• Adverse events • Laboratory data • Vital signs	Date of first drug intake	Date of last drug intake + 28 days

[1] End date is included.
[2] Date of randomisation and date of first drug intake should be identical. If there are discordances between these dates on patient level, date of first drug intake will be used.
[3] Patients who switch to the extension study will only be evaluated until date of last drug intake in this study. The residual effect period will not apply, as patients will receive continuous treatment.
[4] For the main analysis only data collected over the whole trial available at the time of data cut off for DBL1 will be considered, whereas all data collected within the trial will be considered and reported for the final analysis.

6.2 IMPORTANT PROTOCOL DEVIATIONS

No per protocol set analysis will be performed for this study; however, the following list in Table 6.2: 1 defines the different categories of potentially important protocol deviations (IPD), and it shows whether patients will potentially be excluded from evaluations performed in this trial. Potentially important protocol deviations will be handled according to BI standards (6), and the proportion of patients with IPDs will be presented for completeness purposes and to demonstrate the adherence to the CTP. Several IPDs are defined over 52 weeks because the primary assessment of benefit-risk of nintedanib in PF-ILD will be based on the analysis over 52 weeks (Part A). These IPDs will also be considered over the whole trial, and the respective summary will be presented as well in the CTR.

Table 6.2: 1 Important protocol deviations

Category/Code		Description	Requirements	Excluded from
A		**Entrance Criteria Not Met**		
	A1	**Inclusion criteria not met**		
	A1.1	Male or female patients aged ≥ 18 years at Visit 1	Inclusion criterion 2 not met (or other age restriction specified as per a local amendment) *Automatic IPD*	None
	A1.2	At least one criteria for PF-ILD fulfilled within 24 months of Visit 1 despite treatment with unapproved medications used in clinical practice to treat ILD	Inclusion criteria 3a, 3b, 3c or 3d not met *Automatic IPD* *Additional note:* *Requirement for treatment with unapproved medications is relative, and does not apply for all ILDs. Therefore lack of previous therapy is NOT to be considered for this IPD*	None
	A1.3	Extent of lung fibrosis > 10% on HRCT performed within 12 months of Visit 1	Inclusion criterion 4 not met; or extent of fibrosis ≤ 10% according to database (evaluated by central review) *Automatic IPD* *Additional note:* *If the HRCT scan is older than 12 months but it is read by the central review, the case is not an IPD. Similarly, if the image acquisition criteria are not met but the image was read, the case is not an IPD.*	None
	A1.4	For patients with underlying connective tissue disease (CTD): stable CTD as defined by no initiation of new therapy or withdrawal of therapy for CTD within 6 weeks prior to Visit 1.	Inclusion criterion 5 not met. *Automatic IPD* *Additional note:* *Any change in CTD therapy within 6 weeks prior to Visit 1 is to be considered as IPD.*	None

Table 6.2: 1 Important protocol deviations (continued)

Category/Code		Description	Requirements	Excluded from
	A2	**Exclusion criteria met**		
	A2.1	Patient has laboratory values that indicate additional risk at Visit 1: a) >1.5xULN for AST or ALT b) >1.5xULN for Bilirubin c) Creatinine clearance < 30 ml /min (Cockcroft-Gault formula)	Laboratory values out of tolerated ranges at visit 1 according to the database: a) 1.5xULN for AST or ALT (exclusion criterion 1 not met) b) 1.5xULN for Bilirubin (exclusion criterion 2 not met) c) Creatinine clearance < 30 ml /min (Cockcroft-Gault formula) (exclusion criterion 3 not met) *Automatic IPD* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category.*	None
	A2.2	Patient with other under-lying diseases, conditions or risks which are excluded as per clinical trial protocol a) chronic liver disease (Child Pugh A, B or C hepatic impairment) b) IPF c) Significant pulmonary Arterial Hypertension d) Major extrapulmonary physiological restriction e) Cardiovascular disease within 6 months of Visit 1 f) Bleeding risk g) History of thrombotic events within 12 months of Visit 1 h) Known hypersensitivity to the trial medication or its components i) Peanut allergy	Pathients who: a) suffer from chronic liver disease (Child Pugh A, B or C) (exclusion criterion 4 met) b) are diagnosed with IPF (exclusion criterion 8 met) c) show significant pulmonary arterial hypertension (exclusion criterion 9 met) d) show major extrapulmonary physiological restriction (exclusion crite-rion 12 met) or according to central review ("are there any co-existing HRCT features that would exclude this patient" is ticked with "yes") e) have a history of cardiovascular disease within 6 months of Visit 1 (exclusion crite-rion 13 met) f) have extended bleeding risk (exclusion criterion 14 met) g) have a history of thrombotic events within 12 months of Visit 1 (exclusion criterion 15 met) h) have a known and documented hypersensitivity to the trial medication or its components (exclusion criterion 16 met) i) have a known and documented peanut allergy (exclusion criterion 17 met) at visit 1. *Automatic IPD* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category.*	None

Table 6.2: 1 Important protocol deviations (continued)

Category/Code		Description	Requirements	Excluded from
	A2.3	Forbidden previous therapy as per clinical trial protocol a) Previous treatment with nintedanib or pirfenidone b) Other investigational therapy received witin 1 month or 6 half-lives (whichever greater) prior to Visit 1 c) Any of the following medications for the treatment of ILD: azathioprine (AZA), cyclosporine, mycophenolate mofetil (MMF), tacrolimus, oral cortico-steroids (OCS) >20mg/day and the combination of OCS + AZA + n-acetylcysteine (NAC) within 4 weeks of Visit 2, cyclophosphamide within 8 weeks of Visit 2, rituximab within 6 months of Visit 2	Patients who have: a) ever been treated with nintedanib or pirfenidone (exclusion criterion 5 met) b) received other investigational therapy witin 1 month or 6 half-lives (whichever greater) prior to Visit 1 (exclusion criterion 6 met) c) used azathioprine, cyclosporine, MMF, tacrolimus, OCS >20mg/day and the combination of OCS+AZA+NAC within 4 weeks of Visit 2, cyclophosphamide within 8 weeks of Visit 2, rituximab within 6 months of Visit 2 (exclusion criterion 7 met) *Manual IPD based on exclusion criteria and / or to be identified at the site level on the manual PD log* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category*	None
	A2.4	Potential risk related to fetotoxicity	Women who fulfil the following criteria at Visit 1 and: a) are pregnant, nursing, or who plan to become pregnant while in the trial (exclusion criterion 21 met) b) are of childbearing potential not willing or able to use highly effective methods of birth control per ICH M3 (exclusion criterion 22 met) *Automatic IPD* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category.*	None
B		**Informed Consent**		
	B1	Informed consent not given	Inclusion criterion 1 not met *Automatic IPD*	All

Table 6.2: 1 Important protocol deviations (continued)

Category/Code		Description	Requirements	Excluded from
	B2	Informed consent given too late	eCRF date of informed consent is after date of Visit 1 *Automatic IPD* *Additional note:* *Signature of the wrong IC version, and later signature of the correct one will also be part of this IPD category. This can however only be determined via a manual process.*	None
	B3	Informed consent not given or withdrawn for optional HRCT but optional HRCT done	According to HRCT database and eCRF page informed consent for optional HRCT *Manual IPD*	None
C		**Trial medication and randomisation**		
	C1	Incorrect trial medication taken between baseline and Week 52	Medication taken (between randomisation and Week 52) that is not matching the treatment a patient was randomised to. *This can only be determine after unblinding. Programming support for listing generation needed, but a manual allocation process.* *Additional note:* *In the CTR Table only the deviation itself will be shown, additional information (e.g. dose level, time point, duration of the wrong medication intake, etc) will be part of a listing.*	None
	C2	Randomisation not followed	The wrong medication kit is given, leading to the patient taking treatment different from the one randomised by IRT at time of randomisation (Visit 2). *This can only be determine after unblinding. Programming support for listing generation needed, but a manual allocation process* *Additional notes:* *This is only an IPD if the medication error leads to an actual treatment switch.*	None

Table 6.2: 1 Important protocol deviations (continued)

Category/Code		Description	Requirements	Excluded from
	C3	Overall compliance between randomisation and Week 52 not between 80% and 120% inclusive (or noncompliance based on investigator assessment)	In case calculated overall compliance is missing: If the answer to the question "Did the patient take trial medication as instructed?" of the eCRF is "No" at at least one of the visits where the compliance cannot be calculated, then an IPD will be flagged. Otherwise no IPD although calculated overall compliance is missing. *Automatic IPD with review of MQRM listings*	None
	C4	Trial medication not interrupted when ALT or AST ≥ 5 fold ULN	Based on eCRF dose interruption and adverse event pages between first drug intake and and Week 52 *Manual IPD*	None
	C5	Trial medication not permanently discontinued after signs of liver enzyme elevations were observed that a) are indicative of hepatic injury as defined in Section 5.3.6.1 of the CTP b) correspond to ALT or AST ≥ 3 fold ULN despite dose reduction or treatment interruption for 2 weeks or more	All patients where signs of liver enzyme elevations are not dealt with according to the requirements of the CTP. *Programmatic support will be required to flag patients who may fulfil the criteria, however a manual review of outputs is required* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category..*	None

Table 6.2: 1 Important protocol deviations (continued)

Category/Code		Description	Requirements	Excluded from
D		**Concomitant Medication**		
	D1	Patients who received either azathioprine, cyclosporine, tacrolimus, Rituximab, Cyclophosphamide, mycophenolate mofetil, OCS >20mg/day for the treatment of ILD or CTD: a) within the first 6 months of study treatment b) after 6 months of study treatment without significant deterioration of ILD or CTD	Patients should be flagged programmatically who take any of the aforementioned medications. It would have to be manually checked if: a) Any of these medications is used during the first 6 months of study AND NO documentation of exacerbation of ILD is provided; or if the medication is used for the treatment of the CTD. b) Any of these medications is used after 6 months of study medication AND NO documentation of worsening of the ILD or exacerbation of ILD is provided; or no documentation of worsening of the CTD is provided *Programmatic support will be required to flag patients who may fulfil the criteria, however a manual review of outputs is required* *Additional note:* *In the CTR Table it should show which criteria are violated; however, deviations of multiple sub-categories will only counted once towards the overall category*	None
	D2	Patients who received either pirfenidone, nintedanib (in another occurrence than as study medication) or any other investigation drug during the entire duration of their study participation.	Any occurrence of such concomitant therapy is to be flagged as IPD. *Automatic IPD* *Additional note:* *In the CTR Table it should show which medication was taken; however, deviations regarding multiple medications will only counted once towards the overall category.*	

Automatic PDs are those detected via an automated programming process using SAS®.

6.3 PATIENT SETS ANALYSED

- Screened set (SCS):
  - This patient set includes all patients having signed informed consent.
- Randomised set: (RS)
  - This patient set includes all randomised patients, whether treated or not.
- Treated set (TS):
  - This patient set includes all randomised patients who received at least one dose of trial medication. The TS also reflects the "the overall population" for the analyses and is considered as one of the two co-primary populations.
    - The subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" (actual HRCT pattern as documented in the eCRF) is considered as the second co-primary population.
    - The subpopulation of patients with "other HRCT fibrotic patterns" (actual HRCT pattern as documented in the eCRF) is referred to as complementary population

    Both subpopulations are mutually exclusive, and together they form the overall population. Evaluations will be done on both co-primary populations and the complementary population.
- PK set (PKS)
  - This patient set includes those patients in the TS with at least one valid plasma concentration available

Please note: Details on which set or population is used for which analysis and output can be found in this statistical analysis plan.

6.4 SUBGROUPS

The subgroups listed in Table 6.4: 1 will be investigated, provided that the:

- actual size of the individual subgroup expression level is large enough (>10% of the treated patients)
- distribution of patients within subgroup expression level across the two treatment arms

allow for meaningful analyses. Groups that are too small will be pooled in a meaningful manner. Selected variables as defined in Section 5.4 will be used to characterize the study population in all subgroups. In addition, evaluations will be performed to determine efficacy as well as safety in these pre-selected subgroups. For more details on the planned analysis, please refer to Section 7.4.3.

Table 6.4: 1 Subgroup analyses – List of subgroups

	Description of study population [1][2]	**Efficacy analyses on the primary endpoint in the overall population** [1][3]	**Safety analyses in the overall population** [1][3]
Gender (Male / Female)	X	X	X
Age (<65 / ≥65)	X	X	X
Race[4] (White / Asian / Black or African American)	X	X	X
Baseline FVC % predicted (≤70% / >70%)	X	X	X
Underlying clinical ILD diagnosis (Hypersensitivity pneumonitis / Idiopathic nonspecific interstitial pneumonia / Unclassifiable idiopathic interstitial pneumonia / Autoimmune ILD[5] / Other ILDs[6])	X	X	X
Baseline bodyweight (≤65kg / >65kg)	X		X
Patients receiving / not receiving DMARDs with known hepatotoxic effects at baseline	X		X
Patients receiving / not receiving DMARDs with known gastrointestinal effects at baseline	X		X
Methotrexate use at baseline (Yes / No)	X		X

[1] For more details regarding analyses, please refer to Table 7.4.3: 1.

[2] The description of study population for all described subgroups will be provided for the two co-primary populations as well as for the complementary population

[3] Please note that the primary endpoint analyses and safety analyses by subgroup will also be performed in the co-primary population of patients with "HRCT with UIP-like fibrotic pattern only" for gender, age and race

[4] Please note that only single race respondents are taken into account. "American Indian or Alaska Native" patients, or patients that classify themselves as "Native Hawaiian or other Pacific Islander", are pooled with "Asian" patients in the subgroup evaluations, due to the very low number of patients in those expression levels.

[5] For the purpose of the subgroup analysis based on grouped clinical ILD diagnoses, a new category of "Autoimmune ILDs" was created. This category includes all cases from the individual eCRF categories of "Rheumatoid Arthritis associated ILD", "Systemic sclerosis associated ILD", "Mixed connective tissue disease" as well as cases identified as "autoimmune ILD" from the eCRF category of "Other fibrosing ILD" (see Table 9.13: 1 for details). The classification of these cases in the eCRF category"Other fibrosing ILD was reviewed and confirmed by the Steering Committee.

[6] For the purpose of the subgroup analysis based on grouped clinical ILD diagnoses a new category of "Other ILDs" was created. This category includes all cases from the individual eCRF categories of "Exposure-related ILD", "Sarcoidosis" as well as cases itentified as other ILD ("other ILD" and "exposure-related ILD" – see Table 9.13: 1 for details) from the eCRF category of "Other fibrosing ILD". The classification of these cases in the eCRF category"Other fibrosing ILD" was reviewed and confirmed by the Steering Committee

6.5 POOLING OF CENTRES

This section is not applicable because centre/country is not included in the statistical model.

6.6 HANDLING OF MISSING DATA AND OUTLIERS

In general, missing data will not be imputed. Exceptions are detailed in the subsequent subsections.

Two patients with non-determined HRCT pattern by central readers were randomised by mistake, and those patients will generally be evaluated according to the information used

during their randomisation in the IRT system ("other HRCT fibrotic patterns" for both patients) in all analyses considering the actual HRCT pattern, including the primary analysis. They will be evaluated as "Patients with HRCT with UIP-like fibrotic pattern only" in the sensitivity analysis investigating the potential effect of mis-stratification.

6.6.1 Primary endpoint

The statistical model used for primary analysis (see Section 7.3.1 of CTP) allows for missing data, assuming they are missing at random (MAR). Even patients with only one post-baseline assessment can be included in the model and can therefore participate in variance estimation. The statistical model assumes that patients who prematurely discontinue study participation would have behaved similarly to those who remained in the study.

Sensitivity analyses using alternative assumptions will be conducted to investigate the potential effect of missing data on the results of the primary analysis (see Section 7.4.2 and Section 9.12.1).

6.6.2 Secondary and further endpoints

6.6.2.1 Change from baseline endpoints

The statistical Mixed effect Model for Repeated Measures (MMRM) used for the analysis of continuous secondary endpoints allows for form-level missing data, assuming they are missing at random. Item-level data for the PRO measures will be handled according to the instructions provided by the instrument developer.

6.6.2.1.1 K-BILD

Missing item scores are imputed based on the average of the non-missing item scores within the domain, rounded to the nearest integer. If the missing items are >50% per domain, then the domain score is set to missing. If item 15 is missing, it will be replaced by the average of all available items 1-14. If any of the domain scores are missing, the total score is set to missing.

6.6.2.1.2 L-PF

The L-PF scores represent means. Missing items are generally not counted in the denominator. Please note the following exception for the Dyspnea score items 1, 2.4-10 and 12

- If "No" is ticked and "B" is ticked, then the item is not scored and does not contribute to the denominator.
- if "No" is ticked and neither "A" nor "B" is ticked, the item is counted as missing BUT the item DOES contribute to the denominator

If the missing items are ≥50% within a score, then the corresponding score is set to missing.

6.6.2.1.3 PF-IQOLS

Missing item scores are imputed by using the average of all available individual dimension ratings (items 1-16) for the missing item. If the missing items are >50%, then the PF-IQOLS score is set to missing.

6.6.2.2 Time-to-event endpoints

In the analyses of the time-to-event endpoints, missing or incomplete data will be managed by standard survival analysis techniques (i.e. censoring).

For the time to first acute ILD exacerbation, or first non-elective hospitalization in case of partially missing dates, the following imputation will be done:

- If day is missing, then imputed day will be the 15th of the month
- If day and month are missing, then imputed date will be the 1st of July of the (non-missing) year
- If year is missing, date will not be imputed

A missing or incomplete date of death will be imputed/completed so that the derived date is the earliest possible date which is on or after date of onset of the fatal AE, and on or after treatment start (in case this AE is treatment emergent), and on or after date of last contact.

6.6.2.3 Categorical endpoints

In the analyses of the binary endpoints, patients with missing data will be considered as non-responders (worst case analysis).

6.6.3 Other variables

6.6.3.1 Concomitant therapies

In case of (partially) missing start and end dates of concomitant therapies, the dates will be imputed so that the extent of exposure to the concomitant therapy is maximal, i.e. the first day (month) of the month (year) for incomplete start dates and the last day (month) of the month (year) for incomplete end dates.

6.6.3.2 Safety endpoints

Missing or incomplete AE dates will be imputed according to BI standards (7).

6.6.3.3 Trial diagnosis date

In case of partially missing date of trial diagnosis, the following imputation will be done:

- If day is missing, then imputed day will be the 1st of the month
- If day and month are missing, then imputed date will be the 1st of January of the (non-missing) year
- If year is missing, date will not be imputed

6.6.3.4 Pharmacokinetic endpoints

Missing data and outliers of PK data are handled according to (4).

Handling of missing PK data will be performed according to the BI standard procedure (5).

6.7 BASELINE, TIME WINDOWS AND CALCULATED VISITS

As a general rule, the last assessment / measurement observed prior to start of trial medication will be assigned to baseline. Note that for some trial procedures (e.g. body weight, vital signs, laboratory tests) this may be the value measured on the same day trial medication was started. In these cases it will be assumed that the measurements were taken prior to the intake of any study medication (if the measurement time was not captured). If no further data available, this can also be the last screening assessment.

Visit windowing will be performed as described in Table 6.7: 1, Table 6.7: 2, Table 6.7: 3 and Table 6.7: 4, in order to assign data to the relevant study visit based on the actual day of the assessment. Data will be analysed using the re-calculated visits in the statistical tables. However, in the listings, all visits performed will be displayed (even if outside time-window), along with the re-calculated visit.

Table 6.7: 1 Time windowing rules for spirometry (FVC), physical exam, vital signs

Time window of actual day [1]			Allocated to		
Start (S_n) day	End (included) (E_n) day	Length of the time-window [days]	Visit number (n)	Visit name	Planned day of the visit (V_n)
-84	1	85	2	Baseline	1
2	22	21	3	2 weeks	15
23	36	14	4	4 weeks	29
37	64	28	5	6 weeks	43
65	127	63	6	12 weeks	85
128	211	84	7	24 weeks	169
212	309	98	8	36 weeks	253
310	373	64	9	52 weeks	365
374 (S_p)	$E_p = (V_p + V_{p+1})/2 = 533$	E_p-S_p +1 = 160	10	68 weeks	477 (V_p)
E_p + 1	$(V_{p+1} + V_{p+2})/2$	E_{p+1}-S_{p+1} +1	11	84 weeks	589 (V_{p+1})
…	…	…	…	Every 16 weeks thereafter	…

[1] First trial drug intake date is taken into account as a reference to calculate time windows

Table 6.7: 2 Time windowing rules for K-BILD, L-PF, PF-IQOLs, SpO2

Time window of actual day [1]			Allocated to		
Start day	End day (included)	Length of the time-window [days]	Visit number	Visit name	Planned day of the visit
-84	1	85	2	Baseline	1
2	127	126	6	12 weeks	85
128	211	84	7	24 weeks	169
212	309	98	8	36 weeks	253
310	373	64	9	52 weeks	365

[1] First trial drug intake date is taken into account as a reference to calculate time windows
Note: SpO_2 not measured at visits 6 and 8

Table 6.7: 3 Time windowing rules for DLCO

Time window of actual day [1]			Allocated to		
Start (S_n) day	End (included) (E_n) day	Length of the time-window [days]	Visit number (n)	Visit name	Planned day of the visit (V_n)
-84	1	85	2	Baseline	1
128	211	84	7	24 weeks	169
310	373	64	9	52 weeks	365
374 (S_p)	$E_p = (V_p + V_{p+1})/2 = 533$	E_p-S_p +1 = 160	10	68 weeks	477 (V_p)
E_p + 1	$(V_{p+1} + V_{p+2})/2$	E_{p+1}-S_{p+1} +1	11	84 weeks	589 (V_{p+1})
…	…	…	…	Every 16 weeks thereafter	…

[1] First trial drug intake date is taken into account as a reference to calculate time windows

Table 6.7: 4 Time windowing rules for laboratory measurements (and pregnancy test)

Time window of actual day [1]			Allocated to		
Start day	End day (included)	Length of the time-window [days]	Visit number	Visit name	Planned day of the visit
-42	1	43	2	Baseline	1
2	22	21	3	2 weeks	15
23	36	14	4	4 weeks	29
37	64	28	5	6 weeks	43
65	106	42	6	12 weeks	85
107	148	42	6a	18 weeks	127
149	190	42	7	24 weeks	169
191	232	42	7a	30 weeks	211
233	281	49	8	36 weeks	253
282	337	56	8a	44 weeks	309
338	373	36	9	52 weeks	365
374	449	76	9a	60 weeks	421
450	505	56	10	68 weeks	477
506 (S_p)	$E_p = (V_p + V_{p+1})/2 = 561$	E_p-S_p +1 = 56	10a	76 weeks	533 (V_p)
E_p + 1	$(V_{p+1} + V_{p+2})/2$	E_{p+1}-S_{p+1} +1	11	84 weeks	589 (V_{p+1})
…	…	…	…	Every 8 weeks thereafter	…

[1] First trial drug intake date is taken into account as a reference to calculate time windows

If after windowing of visits at baseline, two or more values fall within the same baseline interval, then the last value prior to first drug intake will be taken into account. If after windowing of post-baseline visits, two visits fall in the same interval, then the measurement

closest to the planned visit will be taken into account. In case two measurements are equidistant from the planned visit, then the last one will be picked.

7. PLANNED ANALYSIS

Planned analyses can be categorized based on:

- The study period for which they are performed:
  - Over 52 weeks (Part A) and over the whole trial (Part A and Part B together)
- The study population for which they are performed:
  - The co-primary populations ("overall" population or sub-population of patients with "HRCT with UIP-like fibrotic pattern only"), or the complementary patient population of patients with "other HRCT fibrotic patterns".

Unless specified otherwise, analyses over 52 weeks (Part A) and over the whole trial (Part A and Part B) will be produced on both co-primary populations. In addition, selected analyses over 52 weeks and over the whole trial will be produced on the complementary population of patients with "other HRCT fibrotic patterns". Details can be found in subsequent sections. Please note that analyses based on the complementary population will, by default with only few exceptions, be part of the CTR appendix.

The labelling and display format of statistical parameters will follow BI standards (9). For End-Of-Text tables, the set of summary statistics is: N / Mean / SD / Min / Median / Max. In descriptive statistics tables, mean, sd and median will be rounded to one additional digit than the raw individual value. In case extreme data outside of the expected range are observed, quartiles and percentiles will be presented additionally.

Tabulations of frequencies for categorical data will include all categories depicted in the eCRF and will display the number of observations in a category as well as the percentage (%) relative to the respective treatment group (unless otherwise specified, all patients in the respective patient set whether they have non-missing values or not). Percentages will be rounded to one decimal place. The category missing will be displayed only if there are actually missing values.

If not otherwise specified, two-sided p-values and confidence intervals will be displayed. Please note that the depiction of p-values in outputs will be limited to the following evaluations:

- Primary analysis performed in both co-primary populations
- Sensitivity analyses of the primary endpoint analysis in both co-primary populations
- Nominal p-values for the evaluation of the main secondary endpoints, evaluated in both co-primary populations
- Interaction p-values for the efficacy evaluations in the subgroup analyses

The primary analyses will include all treated patients.

7.1 DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS

A table in the CTR will present the number of patients screened, randomised and treated. The number of patients prematurely discontinuing their study treatment will be shown with the reasons for discontinuation. The number and percent of patients completing the end of treatment visit and the number of patients with treatment ongoing will also be presented. Please note there will be ongoing patients at the time of the data cut-off for the primary analysis; at the time point when the final analysis will take place there will not be any more ongoing patients in this trial. Where percentages are shown, the denominator will be the number of patients treated in each treatment group.

Descriptive statistics as well as frequency counts will be provided for all demographic and baseline characteristics depicted in Section 5.4.1. The CTR tables will show the relevant descriptive statistics (number and percent within categories; other descriptive statistics for continuous variables) by treatment arm.

7.2 CONCOMITANT DISEASES AND MEDICATION

Only descriptive statistics are planned for this section of the report.

7.2.1 Baseline conditions and cardiovascular medical history

The baseline conditions will be included as coded items using the current Medical Dictionary for Regulatory Activities (MedDRA) version in use at BI at the time of database lock. They will be summarised by MedDRA system organ class (SOC) and Preferred Term. The CTR table will show the counts of patients with a Baseline condition in each SOC present (SOC sorted by standard European Medicines Agency (EMA) order) and then the conditions (preferred terms) under that SOC in descending order of overall prevalence. This summary will be produced on all patients and by the patient's underlying HRCT patterns.

The relevant cardiovascular medical history will be summarised, with particular focus on myocardial infarction, transient ischemic attacks, stroke, cardiac arrhythmia and heart failure.

7.2.2 Concomitant therapies

The following categories of concomitant therapies have been created:

- Previous therapies
    - Defined as treatments with an end date before first trial drug intake.
- Baseline therapies
    - Defined as treatments with a start date before first trial drug intake and taken after or on the day of the first trial drug intake.
- On-treatment concomitant therapies
    - Defined as treatments with a start date after or on the day of first trial drug intake and before or on the day of last trial drug intake.

- Post-study drug discontinuation therapies
    - Defined as treatments with a start date after last trial drug intake and before trial completion (as defined in Section 6.2.3 of the CTP).

The analysis of concomitant therapies will be based on the following groupings of the aforementioned categories of concomitant therapies:

- Baseline therapies
- Baseline and on-treatment concomitant therapies
- On-treatment concomitant therapies

Concomitant therapies will be described over the first 52 weeks, as well as over the entire duration of the study.

Table 7.2.2: 1 and Table 7.2.2: 2 summarise the concomitant therapy outputs which will be provided over 52 weeks and over the whole trial respectively. Summaries by ATC and preferred name (PN) will use the ATC3 code, and will be sorted by alphabetical ATC class and decreasing frequency of PN in all patients within ATC class. Summaries by Customized drug groupings of interest (CDG) will be sorted by alphabetical CDG and decreasing frequency of PN in all patients within CDG. CDGs can be found in Section 9.7. Restricted or forbidden concomitant therapies including categorisations as well as World Health Organization Drug Dictionary (WHO-DD) code are listed in Section 9.8. Please note that the most recent WHO-DD version in use at BI at the time of database lock is used.

Table 7.2.2: 1 Concomitant therapy outputs over 52 weeks.

	By ATC and PN		**By CDG and PN**	
	Co-primary populations	**Complementary population**	**Co-primary populations**	**Complementary population**
Baseline therapies	15.1	16.1.13.1	15.1	16.1.13.1
All baseline and on-treatment concomitant therapies with a frequency >2% in at least one treatment arm	15.1	16.1.13.1	-	-
All baseline and on-treatment concomitant therapies	16.1.13.1	16.1.13.1	-	-
All on-treatment restricted concomitant therapies	-	-	15.1	16.1.13.1
All on-treatment prohibited concomitant therapies	-	-	16.1.13.1	16.1.13.1

Table 7.2.2: 2 Concomitant therapy outputs over the whole trial.

	By ATC and PN		By CDG and PN	
	Co-primary populations	**Complementary population**	**Co-primary populations**	**Complementary population**
All baseline and on-treatment concomitant therapies with a frequency >2% in at least one treatment arm	16.1.13.1	16.1.13.1	-	-
All baseline and on-treatment concomitant therapies	16.1.13.1	16.1.13.1	-	-
All on-treatment restricted concomitant therapies	-	-	16.1.13.1	16.1.13.1
All on-treatment prohibited concomitant therapies	-	-	16.1.13.1	16.1.13.1

Definitions of restricted and prohibited medications are depicted in Section 4.2.2 of the CTP. Please note that disease-modifying anti-rheumatic drugs (DMARDS) with known hepatotoxic or gastrointestinal effects – which are part of the CDGs – taken at baseline or prescribed on-treatment will be shown additionally in a separate output for convenience. Section 9.7 shows which medications were considered to belong to the two categories of DMARDs.

7.3 TREATMENT COMPLIANCE

Treatment compliance will be computed for both treatment arms. Missed medication intakes, excluding those in accordance with the requirements of the CTP (e.g. due to the occurrence of an adverse event) will be assessed. The outputs will show the number of patients who have missed at least one dose, as well as a percentage value of medication taken compared to what the patients should have received. The percentage rate shown will be the average percentage rate over all patients.

Additional outputs will be created to investigate whether patients were treated as specified according to the definitions given in Section 4 of the CTP. These outputs show:

- The dose levels used
- The appropriate use of the dose reduction scheme

The evaluations will be done on the co-primary populations as well as on the complementary population, for the first 52 week period and over the whole trial, and only descriptive statistics for the raw values as well as the categorisations for compliance shown in Section 5.4.2 will be created for this part of the report.

7.4 PRIMARY ENDPOINT

The primary efficacy endpoint is the annual rate of decline in Forced Vital Capacity (FVC; expressed in ml over 52 weeks). Please see Sections 7.2, 7.3 and 7.3.1 of the CTP for additional information on the underlying hypotheses, multiplicity adjustments and assessment strategy.

7.4.1 Primary analysis

The primary analysis will be based on the TS (according to randomised treatment), using all available data from baseline (excluded) up to Week 52, including visits done after premature treatment discontinuation, EOT visits and follow-up visits done before Week 52 (i.e. including all measurements after first drug intake and before or on day 373). Patients will be analysed based on the fibrotic pattern data entered in the eCRF, which may not necessarily coincide with data used for randomisation. Details on sensitivity analyses can be found in Section 7.4.2.

The primary analysis uses a restricted maximum likelihood (REML) based approach with a random slope and intercept model. The analysis will include the fixed, categorical effects of treatment, HRCT pattern, fixed continuous effects of time and baseline FVC (ml) as well as the treatment-by-time and baseline-by-time interactions. Random effects will be included for patient response for both time and intercept.

The statistical model can be written as follows:

$$y_{ijkm} = (\alpha + a_i + \vartheta_m + \beta_0 S_i + \tau_k) + (\gamma + g_i + \beta_s S_i + \varphi_k)t_{ij} + e_{ij}$$
$$(a_i, g_i) \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$$
$$e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

The components of the model are as follows:

y_{ijkm}	=	response variable for patient i with HRCT pattern m at visit j receiving treatment k
α	=	mean patient intercept
a_i	=	random intercept effect for patient i, i=1,2,…
ϑ_m	=	intercept coefficient of HRCT pattern (“other HRCT fibrotic patterns” as the class reference)
β_0	=	intercept coefficient of baseline FVC (ml)
S_i	=	baseline FVC (ml) measurement for patient i
τ_k	=	intercept coefficient of the effect of treatment k, k=1,2
γ	=	mean patient slope
g_i	=	random slope effect for patient i
β_s	=	slope coefficient of baseline FVC (ml)

φ_k = slope coefficient of the effect of treatment k

t_{ij} = time of measurement j for patient i, j= 1,2,… J

e_{ij} = the random error associated with the j^{th} visit of patient i. Measurement errors are independent and normally distributed with mean 0 and variance σ^2, and uncorrelated with a_i and g_i.

$\boldsymbol{\Sigma}$ = a 2x2 unstructured covariance matrix

For the co-primary evaluation of FVC (in ml over 52 weeks) in patients with "HRCT with UIP-like fibrotic pattern only", the following model will be utilised:

$$y_{ijk} = (\alpha + a_i + \beta_0 S_i + \tau_k) + (\gamma + g_i + \beta_s S_i + \varphi_k) t_{ij} + e_{ij}$$
$$(a_i, g_i) \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$$
$$e_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

The components of the model are as follows:

y_{ijk} = response variable for patient i at visit j receiving treatment k

α = mean patient intercept

a_i = random intercept effect for patient i, i=1,2,…

β_0 = intercept coefficient of baseline FVC (ml)

S_i = baseline FVC (ml) measurement for patient i

τ_k = intercept coefficient of the effect of treatment k, k=1,2

γ = mean patient slope

g_i = random slope effect for patient i

β_s = slope coefficient of baseline FVC (ml)

φ_k = slope coefficient of the effect of treatment k

t_{ij} = time of measurement j for patient i, j= 1,2,… J

e_{ij} = the random error associated with the j^{th} visit of patient i. Measurement errors are independent and normally distributed with mean 0 and variance σ^2, and uncorrelated with a_i and g_i.

$\boldsymbol{\Sigma}$ = a 2x2 unstructured covariance matrix

For both models, the Kenward-Roger approximation will be used to estimate denominator degrees of freedom and adjust standard errors. Significance tests will be based on least-squares means using a two-sided $\alpha = 0.05$ (two-sided 95% confidence intervals).The primary treatment comparison of slopes will be assessed through the treatment-by-time interaction coefficient.

In the event of non-convergence, the following methods will be attempted (in order) to overcome it:

1. Add the 'singular=1e-10' option in the model statement – This raises the threshold at which columns are declared linearly dependent (from typically 1e-12).
2. Set 'maxiter=100' in the Proc Mixed statement – This increases the number of convergence iterations used from a default of 50.
3. Set 'scoring=4' to specify use of the Fisher scoring algorithm in the first 4 iterations.
4. Include the statement 'performance nothread' – this removes multi-threading from the calculations.
5. Provide starting values for covariance parameters using a 'parms' statement. Estimates will be obtained from using a simpler covariance matrix.

The first model to converge will be used as the primary analysis on the analysis of the FVC rate of decline at week 52.

Patients will be analysed according to the stratum to which they belong to, regardless of any miss-assignment to treatment based on identification of the wrong stratum.

Please refer to Section 9.5 for more details concerning the statistical model and SAS code specifications.

For the primary endpoint a Hochberg procedure will be used in order to maintain an overall type 1 error rate of 5%. Statistical significance will be declared if the primary endpoint analyses in both co-primary populations are significant at the two-sided 5% level, or if the analyses in either co-primary population are statistically significant at the two-sided 2.5% level. No other endpoints i.e. secondary or further endpoints will be evaluated in a confirmatory manner.

Please note that for completeness the same evaluation of FVC (in mL over 52 weeks) that is done for the co-primary patient population of patients with "HRCT with UIP-like fibrotic pattern only" will be run as well in patients with "other HRCT fibrotic patterns" utilising the same model, but no p-value will be shown.

7.4.2 Sensitivity analyses

Sensitivity analyses for FVC (in ml over 52 weeks) using different assumptions will be conducted to investigate the potential effect of data handling, missing data and the analysis model (e.g. assumption of linearity) on the results of the main analysis. In addition, it will be evaluated, whether potential discrepancies between IRT and eCRF of HRCT pattern have an impact if more than 3% of the randomised patients show such discordance.

For all sensitivity analyses to data handling assumptions the estimate, two-sided 95% Confidence Interval and p-value of the treatment effect will be represented on the same Forest plot. Tables presenting the results of the sensitivity analyses will be created in addition to graphical representations, and included in the CTR Appendix.

Interpretation of sensitivity analyses:

Separate forest-plots will be produced for each of the co-primary evaluations in Section 15 of the CTR. In case the results of some of the sensitivity analyses are inconsistent with those of the primary analysis (i.e. the treatment effect estimate of a sensitivity analysis lies outside the confidence interval of the treatment effect of the primary analysis), clinical input will be needed to assess the clinical relevance of the difference.

Sensitivity analyses described in the CTP will be discussed in this section, additional sensitivity analyses in Section 9.12.

7.4.2.1 Sensitivity to data handling assumptions

7.4.2.1.1 Statistical model using on-treatment measurements only

A sensitivity analysis including only on-treatment measurements of FVC (ml) will be presented. The same model as for the primary analysis will be used (see Section 7.4.1). This model implies that data are assumed to be missing at random (MAR) and it is implicitly supposed that patients who dropout would have behaved similarly to those who remained in the study.

This analysis is considered of principal importance amongst the planned sensitivity analyses since it most closely reflects the expected biologic effect of nintedanib in the treatment of patients with PF-ILD.

7.4.2.1.2 Sensitivity to missing data handling

To investigate the potential impact of missing data on the treatment effect, patients will be classified into different patterns depending on the availability of data:

- Patients with a 52 week FVC value (see Table 6.1.1: 1 for further information regarding the analysis period):

    1. Those who received trial drug until 52 weeks (defined as patients who did not prematurely discontinue the trial medication before 52 weeks according to the "end of trial medication" page of the eCRF) (pattern 1)
    2. Those who prematurely discontinued trial drug before 52 weeks (as per information given on the "end of trial medication" page of the eCRF) but who were followed up until week 52 (pattern 2)

- Patients without a 52 week FVC value:

    3. those who were alive at 52 weeks (based on "vital status" page of the eCRF, and no fatal AE recorded on the "adverse event" page of the eCRF) (pattern 3)
    4. those who died before 52 weeks (based on "vital status" page of the eCRF, or fatal AEs recorded on the "adverse event" page of the eCRF) (pattern 4)

These four patterns will be used in sensitivity analyses to estimate the treatment effect under differing assumptions regarding the persistence of efficacy post withdrawal of randomised

treatment. As described hereafter, three resulting alternative analyses will be defined. See an overview in Table 7.4.2.1.2: 1.

Multiple imputation will be used to handle missing data at week 52. Non-monotone missing data and/or missing data at visits before week 52 will not be imputed. The imputation model will be similar to the statistical model of the primary analysis (see Section 7.4.1).

The number of imputations will be set to 1000 in order to ensure adequate efficiency for the estimation of missing data. For each imputed dataset, the same statistical model as defined for the primary analysis will be used for the analysis as depicted in Section 7.4.1. The results will be pooled following the standard multiple imputation procedure (16). See also Section 9.5.3 for further technical information on the implementation of the multiple imputation approach.

Table 7.4.2.1.2: 1 Primary and sensitivity analyses for handling of missing data

	Pattern 3: Missing week 52 data in patients still alive at 52 weeks		**Pattern 4: Missing week 52 data in patients who died before 52 weeks**	
Analysis	**Handling of missing week 52 data**	**Underlying assumption regarding persistence of efficacy post-withdrawal**	**Handling of missing week 52 data**	**Underlying assumption regarding persistence of efficacy after death**
Primary	no imputation	assumes MAR	no imputation	assumes MAR
Sensitivity 1[1]	based upon the slope (SE) estimates in Drug and Placebo in patients of pattern 2, multiple imputation of missing week 52 data in the respective treatment group	rate of decline in patients with missing week 52 data is similar to rate of decline in patients of pattern 2 in the respective treatment group (e.g. treatment effect persists in same manner as for pattern 2 patients after trial drug discontinuation)	multiple imputation of missing 52 week data due to death based on the same slope (SE) estimates in Placebo patients of pattern 2, but truncated[3] to force the slope in patients who died to be more severe than in those who survived	Assuming that deaths observed in the trial will likely be related to worsening of the underlying disease, it seems reasonable to assume that the unobserved FVC values should on average be lower than those in patients who did not die prior to week 52 Rate of decline in patients who died before week 52 is similar to rate of decline in the Placebo patients of pattern 2 with most severe slopes
Sensitivity 2[1]	based upon the slope (SE) estimates in Placebo patients of pattern 2: multiple imputation of missing week 52 data in all patients regardless of treatment group	rate of decline in all patients with missing week 52 data is similar to rate of decline in Placebo patients of pattern 2 (e.g. treatment effect does not persist after trial drug discontinuation)		
Sensitivity 3[2]	based upon the slope (SE) estimates in Placebo patients from the primary analysis model (MMRM), i.e. in patients from pattern 1 or 2: multiple imputation of missing week 52 data in all patients regardless of treatment group	rate of decline in all patients with missing week 52 data is similar to rate of decline estimated in all Placebo patients (e.g. treatment effect does not persist after trial drug discontinuation)	multiple imputation of missing 52 week data due to death based on the same slope (SE) estimates in all Placebo patients (i.e. in patients from pattern 1 or 2), but truncated[3] to force the slope in patients who died to be more severe than in those who survived	Assuming that deaths observed in the trial will likely be related to worsening of the underlying disease, it seems reasonable to assume that the unobserved FVC values should on average be lower than those in patients who did not die prior to week 52 Rate of decline in patients who died before week 52 is similar to rate of decline in the Placebo patients (i.e. in patients from pattern 1 or 2) with most severe slopes

[1] Sensitivity analyses 1 and 2 will only be performed if the number of patients in pattern 2 is greater than 10 (from both Nintedanib and Placebo group for sensitivity analysis 1 and from placebo group for sensitivity analysis 2)

[2] Patients falling into pattern 2 are used as the basis for multiple imputations but since the number of patients in that pattern may be small, a third sensitivity analysis will be performed to confirm the robustness of the primary analysis results.

[3] If β represents the true slope with $f(\beta) \sim N(\hat{\beta}, \hat{\sigma}^2)$ where $\hat{\beta}$ and $\hat{\sigma}$ are the placebo slope and SE estimates from either patients in pattern 2 or all placebo patients, then sampling for patients who died prior to 52 weeks is restricted to the interval $(-\infty, \hat{\beta}]$ of the truncated distribution f(β)/2. In this way, it is guaranteed that, on average, the imputed FVC slope for patients who died is steeper than the average slope in patients who survived to week 52.

In addition to the patterns defined in the scope of the multiple imputation, the number (and percentage) of patients in each possible pattern of FVC data will be explored. Graphs showing FVC data over time in each pattern of monotonic missing data will also be provided. These outputs will be shown in the CTR Appendix.

7.4.2.2 Sensitivity to the analysis model

7.4.2.2.1 Sensitivity to linearity assumption

The linearity assumption for the decline in FVC [ml/yr] will be explored graphically. The following graphical displays will be provided:

- The mean (± Standard Error of the Mean (SEM)) observed FVC [ml] for each treatment group over time
- The mean (± SEM) observed FVC change from baseline [ml] for each treatment group over time
- The mean (± SEM) estimated FVC [ml] for each treatment group over time (as estimated in the primary analysis)

The following alternative models will also be performed and a graphical display only will be provided:

- A polynomial time model (quadratic form in t) will be explored
- An exponential model (of the form: a + b exp(-c*t)) will be explored as well

For each model, a plot will be provided representing both the mean (± SEM) estimated FVC [ml] over time and the mean (± SEM) estimated FVC [ml] over time using the linear form (by treatment group).

In these models, data are assumed to be missing at random, which implies that patients who dropout would have behaved similarly to those who remained in the study.

7.4.2.2.2 Sensitivity to mis-stratification

If potential discrepancies between IRT and eCRF with regard to HRCT pattern exceed 3% of the randomised patients a statistical model similar to the one depicted in Section 7.4.1 analysis will be used, with baseline FVC (ml) and HRCT pattern (“HRCT with UIP-like fibrotic pattern only” / “other HRCT fibrotic patterns”) as covariates. The values used for

HRCT pattern will however not come from the eCRF, but reflect the values used during the randomisation process and stored with the IRT provider.

7.4.3 Subgroup analyses

For each subgroup analysis (refer to Section 6.4 for an overview of subgroups) of the primary endpoint, the heterogeneity of the subgroup treatment effect on the slope will be estimated: A random slope and intercept mixed model will be fitted based upon the statistical model for the primary analysis but parametrised using the treatment-by-subgroup and the treatment-by-subgroup-by-time interaction terms. A contrast statement, with appropriate contrasts, will be used to conduct an F-test of heterogeneity across all expression levels of the subgrouping. The results will be discussed and may be explored further.

It can be shown that the following simplified, saturated model is equivalent to the fully specified model and will therefore be used for the analysis of subgroups:

$$y_{ijkmu} = (\alpha + a_i + \vartheta_m + \beta_0 S_i + \delta_{ku}) + (\gamma + g_i + \beta_s S_i + \theta_{ku}) t_{ij} + e_{ij}$$
$$(a_i, g_i) \sim N_2\,(\mathbf{0}, \boldsymbol{\Sigma})$$
$$e_{ij} \overset{iid}{\sim} N\,(0, \sigma^2)$$

The components of the model are identical to the primary model (see Section 7.4.1 for explanation of notions) only adding the subgroup covariate:

δ_{ku} = intercept coefficient for the effect of treatment k in subgroup category u for patient i

θ_{ku} = slope coefficient for the effect of treatment k in subgroup category u for patient i

Please note: For the subgroups where the co-primary population is also evaluated, the term for HRCT pattern will be excluded from the model. In addition, evaluations are also only reasonable if enough subjects are present within the subgroup – for this evaluation there need to be at least 5 observations per treatment per subgroup expression level per visit. If this is not given, it is possible to pool similar subgroup expression levels together. Whether subgroup expression levels can and should be pooled will be decided at the final Blinded Report Planning Meeting (BRPM) and documented in the minutes. Section 9.8 of the CTR will then also discuss the pooling of subgroup expression levels and provide a medical rationale.

In addition to the subgroup evaluations described above, one output will be computed treating the HRCT pattern and its expression levels ("HRCT with UIP-like fibrotic pattern only" / "other HRCT fibrotic patterns") as subgroups. The model would exclude the term for the HRCT pattern, and the obtained interaction p-value would serve as objective measurement whether the results of the primary endpoint are consistent between the subpopulation of patients with "HRCT with UIP-like fibrotic pattern only" and the subpopulation of patients with "other HRCT fibrotic patterns".

Table 7.4.3: 1 Subgroup analyses – List of analyses

Description of study population	• Disposition of patients • Demographic data • Trial indication characteristics • Baseline pulmonary efficacy variables • Baseline conditions • Exposure to study drug over 52 weeks • Exposure to actual treatment dose received over 52 weeks
Efficacy analyses on the primary endpoint in the overall population [1]	• Forest-plot for rate of decline in FVC (ml/yr) over 52 weeks in all subgroups: representation on the same graph of the estimate, 95% Confidence Interval and p-value of F-test for heterogeneity across all levels of the subgrouping. • Rate of decline in FVC (ml/yr) over 52 weeks by subgroup • Graphical representation of the mean (SEM) observed FVC change from baseline (ml) over time by subgroup over 52 weeks • Graphical representation of the mean (SEM) estimated FVC change from baseline (ml) over time by subgroup over 52 weeks
Safety analyses in the overall population [1][2]	• Adverse event overall summary over 52 weeks • Frequency of patients with adverse events by SOC and PT over 52 weeks • Frequency of patients with adverse events occurring with incidence in preferred term > 5% in at least one treatment arm by SOC and PT over 52 weeks • Frequency of patients with serious adverse events by SOC and PT over 52 weeks • Frequency of patients with investigator defined related adverse events by SOC and PT over 52 weeks • Frequency of patients with adverse events leading to treatment discontinuation events by SOC and PT over 52 weeks • Frequency of patients with adverse events leading to death by SOC and PT over 52 weeks • Frequency of patients with adverse events by safety topic over 52 weeks • Frequency of patients with serious adverse events by safety topic over 52 weeks

[1] Please note that the primary endpoint analyses and safety analyses by subgroup will also be performed in the co-primary population of patients with "HRCT with UIP-like fibrotic pattern only" for gender, age and race

[2] The depicted evaluations over 52 wees will also be repeated over the whole trial. The outputs over the whole trial will also contain incidence rates per 100 patient years.

7.5 SECONDARY ENDPOINTS

Please note that any p-values presented for secondary endpoints will be considered nominal in nature and no adjustment for multiplicity will be made.

7.5.1 Key secondary endpoints

This section is not applicable as no key secondary endpoint has been specified in the protocol.

7.5.2 Secondary endpoints

All secondary endpoints are defined over the 52 weeks period only.

7.5.2.1 Main secondary efficacy endpoints

7.5.2.1.1 King's Brief Interstitial Lung Disease Questionnaire (K-BILD) Total Score

The absolute change from baseline in K-BILD total score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status. The descriptive analyses of the absolute change from baseline in K-BILD total score will also be shown for the complementary population.

All sub-domain scores (breathlessness and activities, psychological, chest symptoms) will also be presented for all time points separately, as well as change from baseline.

In addition to the descriptive evaluation, absolute change from baseline in K-BILD total score between week 52 and baseline will also be analysed in the TS using a restricted maximum likelihood (REML) based repeated measures approach. An unstructured variance-covariance structure will be used to model the within patient measurements. In the event of non-convergence, the same methods used in the primary analyses (Section 7.4.1) will be implemented to overcome the issue. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co-primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the baseline K-BILD total score.

A graph of the mean K-BILD total score absolute change from baseline (± SEM) over time for each treatment group will be displayed. The same graph will be performed on adjusted mean changes from baseline.

7.5.2.1.2 Time to death

A stratified log-rank test will be used to evaluate the effect of Nintedanib on time to death compared to Placebo. The test will be stratified by the randomisation stratification variable HRCT pattern ("HRCT with UIP-like fibrotic pattern only" vs. "Other HRCT fibrotic patterns", as documented in the CRF). A Cox proportional-hazards model stratified by the same factors as the log-rank test will be used to derive the hazard ratio and 95 % confidence interval (CI) between the two randomised treatment regimens (Hazard ratios <1 will favour nintedanib). In the co-primary population of patients with "HRCT with UIP-like fibrotic pattern only", the respective analyses will be performed without stratification.

Breslow's method for handling ties will be used. Kaplan-Meier plots by treatment group will also be presented. The assumption of proportional hazards and the homogeneity of the hazard ratio will be checked descriptively.

If the proportion of patients experiencing an event is less than 5%, only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.5.2.1.3 Time to first acute ILD exacerbation or death

The same type of evaluation as described for time to death (see Section 7.5.2.1.2) will be implemented. If, however, the proportion of patients experiencing at least 1 acute ILD exacerbation over 52 weeks is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.5.2.2 Other secondary efficacy endpoints

7.5.2.2.1 Time to death due to respiratory cause

In general, the same type of evaluation for time to death due to respiratory cause will be implemented as described for time to death (see Section 7.5.2.1.2). However; only patients whose cause of death was attributed to respiratory causes (as determined by an independent adjudication committee) will be included. If the proportion of such patients within the first 52 weeks is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.5.2.2.2 Time to progression or death

The same type of evaluation as described for time to death (see Section 7.5.2.1.2) will be implemented for this endpoint. If, however, the proportion of patients experiencing a progression (see Section 5.2.2.2.2 for details about the definition of progression) within the 52 week period is less than 5%, then only frequencies of events by treatment group will be provided. In this case no other analyses will be performed.

7.5.2.2.3 Relative decline from baseline in FVC % pred

Patients will be categorised based on the relative decline since baseline in FVC % pred greater than 5% or 10%. Comparisons between treatment groups will be performed using a logistic regression model adjusting for the continuous covariate baseline FVC % pred and the binary covariate HRCT pattern. The likelihood-ratio test will be used to test for differences between treatments. Adjusted odds ratios together with 95% confidence intervals will be used to quantify the effect of treatment, comparing nintedanib 150 mg bid to placebo as the reference. Additionally, descriptive analyses of the relative change from baseline in FVC % pred will be shown, for the co-primary populations as well as for the complementary population.

7.5.2.2.4 Living with Pulmonary Fibrosis Symptoms and Impact Questionnaire (L-PF)

The absolute change from baseline in L-PF symptoms dyspnoea domain score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The absolute change from baseline in L-PF symptoms cough domain score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The descriptive analyses of the absolute change from baseline in L-PF symptoms dyspnoea domain score as well as L-PF symptoms cough domain score will also be shown for the complementary population.

In addition to the descriptive evaluation, absolute change from baseline in L-PF symptoms dyspnoea domain score between week 52 and baseline, as well as in L-PF symptoms cough domain score between week 52 and baseline will also be analysed using a restricted maximum likelihood (REML) based repeated measures approach. An unstructured variance-covariance structure will be used to model the within patient measurements. In the event of non-convergence, the same methods used in the primary analyses (Section 7.4.1) will be implemented to overcome the issue. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the respective baseline L-PF symptoms score.

7.6 FURTHER ENDPOINTS

All further endpoints will be considered exploratory in nature, and no p-values will be presented.

7.6.1 Over 52 weeks

7.6.1.1 Time to first non-elective hospitalisation or death

The same type of evaluation as described in Section 7.5.2.1.2 will be implemented. If, however, the proportion of patients experiencing a non-elective hospitalisation over 52 weeks is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.6.1.2 Absolute change from baseline in FVC

The absolute change from baseline in FVC – in ml, but also % pred – at week 52 will be analysed using a restricted maximum likelihood (REML) based repeated measures approach. Analyses will include the fixed, categorical effects of HRCT pattern, treatment, visit, and treatment-by-visit interaction, as well as the continuous, fixed covariates of baseline and baseline-by-visit interaction. An unstructured (co)variance structure will be used to model the within-patient measurements. In the event of non-convergence, the same methods that are used for the primary endpoint analysis as described in Section 7.4.1 will be used to overcome the issue. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the baseline FVC value. Additionally, descriptive analyses of the absolute change from baseline will be shown, for the co-primary populations as well as for the complementary population.

7.6.1.3 Absolute decline from baseline in FVC % pred

The same type of evaluation as described in Section 7.5.2.2.3 will be implemented for this endpoint.

7.6.1.4 Carbon Monoxide Diffusion Capacity (DLCO) % pred

A similar analysis based on a restricted maximum likelihood (REML) based repeated measures approach as depicted in Section 7.6.1.2 will be performed for the absolute change in DLCO % pred. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the baseline DLCO value

7.6.1.5 Living with Pulmonary Fibrosis Symptoms and Impact Questionnaire (L-PF)

The absolute change from baseline in L-PF total score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The absolute change from baseline in L-PF impact score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The absolute change from baseline in L-PF symptoms total score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The absolute change from baseline in L-PF symptoms fatigue domain score between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status.

The descriptive analyses of the absolute change from baseline in L-PF total score, L-PF impact score, L-PF symptoms total score and L-PF symptoms fatigue domain score will also be shown for the complementary population.

In addition to the descriptive evaluation, absolute change from baseline in L-PF total score between week 52 and baseline, in L-PF impact score between week 52 and baseline, in L-PF symptoms total score between week 52 and baseline and in L-PF symptoms fatigue domain score between week 52 and baseline will also be analysed using a restricted maximum likelihood (REML) based repeated measures approach. An unstructured variance-covariance structure will be used to model the within patient measurements. In the event of non-convergence, the same methods used in the primary analyses (Section 7.4.1) will be implemented to overcome the issue. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the respective baseline L-PF score.

7.6.1.6 Pulmonary Fibrosis Impact on Quality of Life Scale (PF-IQOLS)

The absolute change from baseline in PF-IQOLS summary score (which is the average of the individual dimension ratings) between week 52 and baseline will be presented graphically as well as numerically to illustrate the impact of the underlying condition on the patient's health status. The descriptive analyses of the absolute change from baseline will also be shown for the complementary population.

In addition to the descriptive evaluation, absolute change from baseline in PF-IQOLS summary score between week 52 and baseline will also be analysed using a restricted maximum likelihood (REML) based repeated measures approach. An unstructured variance-covariance structure will be used to model the within patient measurements. In the event of non-convergence, the same methods used in the primary analyses (Section 7.4.1) will be implemented to overcome the issue. The HRCT pattern will be included as a covariate (with exception when the analysis is run in the co primary population of patients with "HRCT with UIP-like fibrotic pattern only") as well as the respective baseline PF-IQOLS summary score.

7.6.2 Over the whole trial

7.6.2.1 Time to first acute ILD exacerbation or death

The same type of evaluation as described in Section 7.5.2.1.3 will be implemented, but data over the whole trial duration will be taken into consideration. Similarly, if the proportion of patients experiencing at least 1 acute ILD exacerbation is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.6.2.2 Time to death

The same type of evaluation as described in Section 7.5.2.1.2 will be implemented, but data over the whole trial duration will be taken into consideration. If the proportion of patients experiencing an event is less than 5%, only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.6.2.3 Time to death due to respiratory cause

The same type of evaluation as described in Section 7.5.2.2.1 will be implemented, but data over the whole trial duration will be taken into consideration. Similarly, if the proportion of such patients is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.6.2.4 Time to progression or death

The same type of evaluation as described in Section 7.5.2.2.2 will be implemented, but data over the whole trial duration will be taken into consideration. Similarly, if the proportion of patients experiencing a progression is less than 5%, then only frequencies of events by treatment group will be provided. In this case no other analyses will be performed.

7.6.2.5 Time to first non-elective hospitalisation or death

The same type of evaluation as described in Section 7.6.1.1 will be implemented, but data over the whole trial duration will be taken into consideration. Similarly, if the proportion of patients experiencing a non-elective hospitalisation is less than 5%, then only frequencies of events by treatment group will be provided for this endpoint. In this case no other analyses will be performed.

7.6.2.6 Descriptive evaluations of lung function parameters

Descriptive analyses of the absolute change from baseline of FVC in ml, FVC (% of predicted) and DLCO (% of predicted), as well as the relative change from baseline in FVC (% of predicted) will be shown, for the co-primary populations as well as for the complementary population.

7.6.3 Pharmacokinetics

Descriptive analysis

Trough concentrations (i.e. pre-dose concentrations) of nintedanib at Visits 4 and 7 will be used. If both concentrations are available, the geometric mean will be used. If one trough concentration value is missing or invalid, the other one will be used without adjustment. To account for potential dose reductions, all trough concentrations will be dose normalised by the actual dose taken by the patient at the time of PK sampling. The dose normalised trough concentrations for nintedanib (Cpre,ss_D) will be summarised descriptively. These analyses will be performed using the PKS.

7.6.4 Biomarkers

All biomarker evaluations will be specified and reported separately.

7.6.5 Longitudinal HRCT assessments

All evaluations of the optional longitudinal HRCT assessments will be specified and reported separately.

7.7 EXTENT OF EXPOSURE

Exposure will be presented based on dose received and time to first dose reduction or interruption. Reasons for dose reduction or interruption will be presented. Exposure will be summarised and presented separately over the first 52 weeks and over the whole trial. In addition, exposure will be presented for patients taking DMARDs with known hepatotoxic or gastrointestinal effects at baseline and on-treatment. Please see Section 5.4.3 for details.

7.7.1 Over 52 weeks

In addition to both co-primary populations, exposure data over 52 weeks will also be summarised on the complementary patient population i.e. patients with “other HRCT fibrotic patterns”.

The duration on treatment and dose intensity will be computed overall, as well as on the individual dose strengths. A summary of treatment interruptions will be created including number of patients with at least one interruption, number and reason of interruptions, as well as time to first interruption. A similar summary will be performed for dose changes.

For the co-primary populations only, a Kaplan-Meier plot of time to premature treatment discontinuation will be produced. Similarly, Kaplan-Meier plots will be performed for time to first dose reduction and for time to first treatment interruption. No statistical tests will be performed.

7.7.2 Over the whole trial

Similar to the 52 week period, exposure data over the whole trial will be summarised on both co-primary populations, as well as the complementary patient population.

7.8 SAFETY ANALYSIS

All safety analyses will be performed on the treated set. Particular focus will be given to evaluations over the 52 week period (Part A of the study).

7.8.1 Adverse events

Unless otherwise specified, the analyses of adverse events will be descriptive in nature. All analyses of AEs will be based on the number of patients with AEs and NOT on the number of AEs.

For analysis multiple AE occurrence data on the eCRF will be collapsed into an AE provided that all of the following applies:

- All AE attributes are identical (LLT, intensity, action taken, therapy required, seriousness, reason for seriousness, relationship, outcome, AESI; CTCAE if applicable)
- The occurrences were time-overlapping or time-adjacent (time-adjacency of 2 occurrences is given if the second occurrence started on the same day or on the day after the end of the first occurrence).

For further details on summarization of AE data, please refer to (7, 8).

The analysis of adverse events will be based on the concept of treatment emergent adverse events. That means that all adverse events occurring between first drug intake until last drug intake + 28 days (and + 7 days as sensitivity analysis) will be assigned to the randomised treatment.

All adverse events occurring before first drug intake will be assigned either to ‘screening’ or ‘post-randomisation’ (for listings only).

All adverse events occurring after last drug intake + 28 days (and + 7 days as sensitivity analysis) will be assigned to ‘post-treatment’ or ‘follow-up’ or ‘post-study’ (for listings only). All adverse events occurring between the start of an interruption of trial medication and the

end of interruption of trial medication will be assigned to 'off-treatment' period in the listings. For details on the treatment definition, see Section 6.1.

According to ICH E3 (10), AEs classified as 'other significant' need to be reported and will include those non-serious and non-significant adverse events with (i) 'action taken = discontinuation' or 'action taken = reduced', or (ii) marked haematological and other lab abnormalities or lead to significant concomitant therapy as identified by the Clinical Monitor/Investigator at a Medical Quality Review Meeting.

Adverse events related to gastrointestinal perforation and hepatic injury will be considered as protocol-specified AEs of special interest (AESIs), and ticked as such in the eCRF.

Table 7.8.1: 1 Protocol-specified AESI

Protocol-specified AESI	**AE type**
Gastrointestinal perforation	AEs established for other tyrosine kinase inhibitors as possible adverse reactions
Hepatic injury	Other AEs of interest

Further adverse event groupings by safety topic have been defined outside the trial protocol, which will be continuously updated at project level (18). These safety topics are deemed of particular importance, and these definitions can be based on selection of coded terms based on MedDRA. The latest approved version of the project level overview archived prior to either DBL will be used in the corresponding CTR.

7.8.1.1 Over 52 weeks

An overall summary of adverse events having occurred over the first 52 weeks will be presented on the co-primary populations, as well as on the complementary population,, including all AEs occurring between first drug intake until:

- last drug intake + 28 days (and + 7 days as sensitivity analysis) for patients who prematurely discontinued the study medication before 52 weeks (i.e. 372 days after first trial drug intake)
- week 52 for patients who did not prematurely discontinue the study medication before 52 weeks

The frequency of patients with adverse events in the co-primary populations, as well as in the complementary population, will be summarised by treatment, primary system organ class and preferred term. Separate tables will be provided for patients with other significant adverse events according to ICH E3 (10) and for patients with protocol-specified adverse events of special interest (as ticked in the AE page of the eCRF [gastrointestinal perforation and hepatic injury]). Displays will also be provided for patients with any adverse events, severe adverse events and also serious adverse events, for patients with adverse events occurring with an incidence in preferred term greater than 5% (in at least one treatment arm), for patients with adverse events leading to permanent dose reduction, for patients with adverse events leading to treatment discontinuation, for patients with investigator defined drug-related adverse events, for patients with drug-related serious adverse events (in co-primary populations only),

and for patients with adverse events leading to death. Further, the frequency of patients with adverse event by system will be provided.

The frequency of patients with adverse events will also be provided by system, safety topic, subcategory (if applicable) and preferred term, with groupings as defined in (18). These displays will focus on patients with any adverse events, patients with serious adverse events and patients with related adverse events.

Additional outputs will be created for the subgroups depicted in Section 6.4 focusing on the overall summary of adverse events, patients with adverse events occurring with an incidence in preferred term greater than 5% (in at least one treatment arm), patients with serious adverse events and patients with related adverse events. An exact overview can be found in Table 7.4.3: 1.

The system organ classes (SOC) will be sorted according to the standard sort order specified by EMA, preferred terms will be sorted by frequency in Nintedanib treatment arm (within SOC).

7.8.1.2 Over the whole trial

Time at risk analyses of AEs will be presented and incidence rates per 100 patient years will be calculated based on the first onset of an AE (see Section 7.8.1.5 for details). An overall summary of adverse events over the whole trial will be presented on the co-primary populations.

The frequency of patients with adverse events in the co-primary populations, as well as in the complementary population, will be summarised by treatment, primary system organ class and preferred term. Separate tables will be provided for patients with other significant adverse events according to ICH E3 (10) and for patients with protocol-specified adverse events of special interest (as ticked in the AE page of the eCRF [gastrointestinal perforation and hepatic injury]). Displays will also be provided for patients with any adverse events, severe adverse events and also serious adverse events, for patients with adverse events occurring with an incidence in preferred term greater than 5% (in at least one treatment arm), for patients with adverse events leading to permanent dose reduction, for patients with adverse events leading to treatment discontinuation, for patients with investigator defined drug-related adverse events, for patients with drug-related serious adverse events (in co-primary populations only), and for patients with adverse events leading to death. Further, the frequency of patients with adverse event by system will be provided.

The overall summary of adverse events and the frequency of patients with adverse events (by treatment, primary system organ class and preferred term) will also be presented on patients with “other HRCT fibrotic patterns”. Separate tables will be provided for patients with protocol-specified adverse events of special interest, for patients with serious adverse events, for patients with adverse events leading to permanent dose reduction, for patients with adverse events leading to treatment discontinuation, and for patients with adverse events leading to death.

Displays by safety topic will also be created over the whole trial, as well as outputs for the subgroups.

7.8.1.3 Adjudicated adverse events

An independent adjudication committee will review all fatal cases and adjudicate cause of death to respiratory, cardiovascular or other. The adjudication committee will also review all AEs categorised as MACE according to the definition in the adjudication charter.

In addition to standard safety analyses, the frequency of patients with AEs leading to death will be summarised by treatment, adjudicated cause of death (Cardiovascular, Respiratory or Other), and PT.

The frequency of patients with AEs categorised as MACE (that is all AEs categorised as MACE and therefore sent for adjudication) will be summarised by treatment and outcome of adjudication (adjudicated as MACE or adjudicated as not MACE). The frequency of patients with AEs adjudicated as MACE will also be summarised by treatment and PT.

7.8.1.4 Adverse events with additional information collection

Diarrhoea, bleeding and ILD are AEs with additional AE-specific information collected on the eCRF. These are investigator reported on the eCRF and will be identified using this information for this analysis. That is if the diarrhoea information has been completed for an adverse event then the adverse event will be considered as diarrhoea for this analysis regardless of subsequent MedDRA coding of the verbatim term. Likewise, if the bleeding information has been completed for an adverse event then the adverse event will be considered as bleeding for this analysis regardless of subsequent MedDRA coding of the verbatim term. The same applies to ILD.

The frequency of patients with AEs with additional information collection will be summarised by treatment, primary SOC and PT separately for diarrhoea, bleeding and ILD. The additional information collected will also be summarised at the AE level rather than at the patient level separately for diarrhoea, bleeding and ILD. For the time to first onset of diarrhoea, bleeding or ILD, respectively, Kaplan-Meier plots by treatment will be created. The same censoring rules as for time to first liver enzyme elevation will be used. See Table 5.4.4: 1 and Table 5.4.4: 2 for details. Please note that these evaluations are prepared for the co-primary populations, as well as in the complementary population, and covering both analysis periods “over 52 weeks” and “over the whole trial”.

7.8.1.5 Exposure adjusted analysis of adverse events

For a specific AE, the total AE time at risk [years] is defined as the sum of time at risk [days] across all contributing patients / 365.25, with for each patient the time at risk [days] defined as follows:

- Date of first start of AE – date of first study medication administration +1 day for patients with the specific AE
- End of time at risk – date of first study medication administration + 1 day for patients without the specific AE

For the AE analyses over 52 weeks, the end of time at risk is the end of the REP for patients who discontinue prior to week 52 or day 373 for patients who do not discontinue prior to

week 52. For the AE analyses over the whole trial, the end of time at risk is the minimum of either "+28 days after termination of trial mediation" or the date of the corresponding database lock.

The AE incidence rate [1/100 Patient years (pt-yrs)] = 100 * number of patients with specific AE / total specific AE time at risk [years]. The 95% confidence intervals for incidence rates are derived using the method described by Rothman and Greenland (2008) (19).

The incidence rate ratio for a given AE is defined as the incidence rate of that event in patients in the Nintedanib 150mg bid treatment group divided by the incidence rate of that event in the Placebo group. The estimates and 95% confidence intervals for incidence rate ratios are based on a Cochran-Mantel-Haenszel test.

The incidence rate difference for a given AE is defined as the incidence rate of that event in patients in the Nintedanib 150mg bid treatment group minus the incidence rate of that event in the Placebo group. The estimates and 95% confidence intervals for incidence rate differences are derived based on the method described in Greenland and Robins (1985) (20).

For this trial, separate frequency tables based on SOC and PT, but also on safety topic level, will be created for the week 52 analyses as well as for the analyses over the whole trial, for the overall population, showing the incidence rate, incidence rate ratio and incidence rate difference (each with 95% confidence interval) per the aforementioned specifications, and added to Section 16.1.13.1.

7.8.2 Laboratory data

The analyses of laboratory data will be descriptive in nature and based on SI units according to BI standards (11). Please refer to Section 7.3.4 of the CTP for further details. All analyses will be performed on the co-primary populations as well as on the complementary patient population. The main focus of the analysis of laboratory parameters will be on the first 52 weeks; however, all analyses of laboratory data will be repeated with data collected over the whole trial for all analyses populations.

Descriptive statistics for laboratory tests based on US conventional units will be provided in addition for the co-primary populations as well as on the complementary patient population, covering both analysis periods "over 52 weeks" and "over the whole trial". All of these outputs will be part of the CTR appendix.

7.8.2.1 Over 52 weeks

A thorough description of liver enzymes and bilirubin elevations over 52 weeks will be given, as defined in Section 5.4.4, including a display of maximum individual elevation. A summary table of liver enzymes and bilirubin elevations over 52 weeks will also be provided (please refer to Table 6.7: 4 for time windowing definition). The time to onset of first liver enzyme and bilirubin elevation [days] will be summarized by categories (≥1-≤30; ≥31-≤60; ≥61-≤90; ≥91). For each patient having experienced a liver enzyme and bilirubin elevation, a graphical representation of AST, ALT and bilirubin over time will be provided. In addition, a Kaplan-Meier plot of time to first liver enzyme elevation (if at least 5% of patients show liver enzyme elevations) will be created and evaluated

7.8.2.2 Over the whole trial

The analysis of liver enzyme elevations over the whole trial will be performed as depicted in Section 7.8.2.1 above.

7.8.3 Vital signs

7.8.3.1 Over 52 weeks

Summary statistics will be presented for observed values and change from baseline by treatment and visit. The frequency of patients with marked changes in vital signs over 52 weeks will also be summarised by treatment according to the endpoints defined in Section 5.4.5 of this document. Please note that these evaluations are prepared for the co-primary populations, as well as in the complementary population.

7.8.3.2 Over the whole trial

The analysis of vital signs over the whole trial will be performed as depicted in Section 7.8.3.1 above.

7.8.4 ECG

Not applicable.

7.8.5 Others

Not applicable.

8. REFERENCES

1	*CPMP/ICH/363/96*: "Statistical Principles for Clinical Trials", ICH Guideline Topic E9, Note For Guidance on Statistical Principles for Clinical Trials, current version.
2	*001-MCS-40-415*: “Statistical Analysis Plans (SAPs)”, current version; IDEA for CON.
3	*001-MCG-159_RD-06*: “Standard table shells for inferential and descriptive Company Standard Displays (CSD-Catalogue)”, current version; IDEA for CON.
4	*001-MCS-36-472*: "Standards and processes for analyses performed within Clinical Pharmacokinetics/Pharmacodynamics", current version; IDEA for CON.
5	*001-MCS-36-472_RD-01*: "Noncompartmental Pharmacokinetic / Pharmacodynamic Analyses of Clinical Studies", current version; IDEA for CON.
6	*001-MCS-40-413*: "Identify and Manage Important Protocol Deviations (iPD)", current version; IDEA for CON.
7	*001-MCG-156_RD-01*: "Handling of missing and incomplete AE dates", current version; IDEA for CON.
8	*001-MCG-156*: "Analysis and Presentation of Adverse Event Data from Clinical Trials", current version; IDEA for CON.
9	*001-MCG-159*: "Reporting of Clinical Trials and Project Summaries", current version; IDEA for CON.
10	*CPMP/ICH/137/95*: “Structure and Content of Clinical Study Reports”, ICH Guideline Topic E3; Note For Guidance on Structure and Content of Clinical Study Reports, current version
11	*001-MCG-157*: "Handling, Display and Analysis of Laboratory Data", current version; IDEA for CON.
12	Patel AS, Siegert RJ, Brignall K, Gordon P, Steer S, Desai SR, Maher TM, Renzoni EA, Wells AU, Higginson IJ, Birring SS. The development and validation of the King’s Brief Interstitial Lung Disease (K-BILD) health status questionnaire. Thorax 2012; 67 (9), 804-810 [R12-4171]

13	Macintyre N, Crapo RO, Viegi G, Johnson DC, Grinten CPM van der, Brusasco V, Burgos F, Casaburi R, Coates A, Enright P, Gustafsson P, Hankinson J, Jensen R, McKay R, Miller MR, Navajas D, Pedersen OF, Pellegrino R, Wanger J. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J 2005; 26 (4), 720 – 735 [R06-2002]
14	Wilson SR, Knowles SB, Mulligan M, Chausow A, Ayala E, Wise RA, et al. Relative and absolute reliability and other properties of the Asthma Impact on Quality of Life Scale (A-IQOLS) and Flanagan QOLS: results of the AQOLIS test-retest study. ATS 2016, 112th Int Conf of the American Thoracic Society (ATS), San Francisco, 13 - 18 May 2016. Am J Respir Crit Care Med 2016; 193, A2679, Abstr [R16-2387]
15	Wilson S, Wise RA, Knowles SB, Huang Q, Wei CY, Castro M. Psychometric properties of the Asthma Impact on Quality of Life Scale (A-IQOLS) and the Flanagan QOLS in adults with well-controlled asthma:baseline results in the LASST trial. ATS 2015, 111th Int Conf of the American Thoracic Society (ATS), Denver, 15 - 20 May 2015. Am J Respir Crit Care Med 2015; 191, A5189, Abstr [R16-2388]
16	Rubin, DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons. 1987. [R12-2378]
17	Japanese Ministry of Health, Labour and Welfare, Format for Preparing the Common Technical Document for Submission of New Drug Applications to Reduce Total Review Time, 2011 [R18-1356]
18	Specifications for adverse event groupings by safety topic for Nintedanib: Nintedanib / Clinical / interstitial lung disease / Project Data Management and Statistics / Section 8 PSAP and Programming / 8-07-other-safety-topic-definition, current version; BIRDS.
19	Rothman KJ, Greenland S, Lash TL. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins, 2008. [R10-1239]
20	Greenland S and Robins JM. Estimation of a common effect parameter from sparse follow-up data, Biometrics 1985, 41: 55-68 [R09-1299]
21	*BI-KMED-COPS-BUP-0007*: Clinical Bridging Study Waiver (BSW) and Subgroup Analysis (SGA) Reports for Asian countries, current version; IDEA for CON.

9. ADDITIONAL SECTIONS

9.1 K-BILD SCORING INSTRUCTIONS

9.1.1 Overview

The K-BILD is comprised of 15 items in three distinct health status domains, including breathlessness and activities (items 1, 4, 11, and 13), chest symptoms (items 2, 7, and 9), and psychological (items 3, 5, 6, 8, 10, 12, and 14) domains. A final item is used to assess the impact of the respondent's lung condition on their financial state. Rasch analysis confirmed the 15 items could be combined into a total score (12). The domain and total scores each range from 0–100, and higher scores indicate a better health status.

Response options for all 15 items are on a 7-point Likert scale, ranging from 1 to 7. For each of the items, the response options vary, including: item 1 (1 = "Every time," 7 = "Never"); items 2–4, and 6–14 (1 = "All of the time," 7 = "None of the time"); item 5 (1 = "None of the time," 7 = "All of the time"); and item 15 ("A significant amount," 7 = "Not at all").

9.1.2 Scoring Algorithm

To score the K-BILD, the Likert response scale weightings for individual items are combined, to ensure they detect progressive change in health status. Table 9.1.2: 1 shows the recoded scale weightings associated with each raw, item-level score value. Raw domain and total scores are then calculated by summing the recoded scale values provided in that table.

Table 9.1.2: 1 K-BILD Response Option Weightings

Item	Response						
	1	**2**	**3**	**4**	**5**	**6**	**7**
1	0	1	2	3	4	5	6
2	0	0	1	1	2	2	3
3	0	1	2	3	4	5	6
4	0	0	1	1	2	2	3
5	0	0	1	1	1	2	2
6	0	1	2	3	4	5	6
7	0	0	1	1	2	2	3
8	0	0	1	1	2	2	3
9	0	0	1	1	1	2	2
10	0	1	2	3	4	5	6
11	0	1	2	3	4	5	6
12	0	1	2	3	4	5	6
13	0	1	2	3	4	5	6
14	0	1	2	3	4	4	5
15	0	0	1	1	1	2	2

Finally, the raw domain and total scores are transformed to a range of 0–100 by using logit values and a simple look-up table to yield interval-level scores, where higher scores indicate better health status. Thus, for each raw summed domain and total score, look up the corresponding transformed domain and total score from Table 9.1.2: 2 below.

Table 9.1.2: 2 K-BILD Score Transformations

Breathlessness and Activities		Chest Symptoms		Psychological		Total	
Raw Score	Transformed Score	Raw Score	Transformed Score	Raw Score	Transformed Score	Raw Score	Transformed Score
0	0	0	0	0	0	0	0
1	10.4	1	17.3	1	10.6	1	9.2
2	17.7	2	32.1	2	17.5	2	15.3
3	22.9	3	44	3	21.9	3	19.4
4	27	4	54.3	4	25.3	4	22.6
5	30.3	5	63.7	5	28	5	25.1
6	33.1	6	73.4	6	30.2	6	27.2
7	35.6	7	85.2	7	32.2	7	29
8	37.8	8	100	8	33.9	8	30.5
9	39.9			9	35.5	9	32
10	41.9			10	37	10	33.3
11	43.9			11	38.5	11	34.4
12	45.9			12	39.8	12	35.5
13	48			13	41.2	13	36.5
14	50.2			14	42.5	14	37.5
15	52.5			15	43.8	15	38.4
16	55.2			16	45.1	16	39.3
17	58.5			17	46.4	17	40.1
18	62.7			18	47.7	18	40.9
19	68.8			19	49.1	19	41.7
20	79.9			20	50.5	20	42.4
21	100			21	52	21	43.2
				22	53.5	22	43.9
				23	55.2	23	44.6
				24	56.9	24	45.2
				25	58.8	25	45.9
				26	60.8	26	46.5
				27	63	27	47.2
				28	65.5	28	47.8
				29	68.3	29	48.5

Table 9.1.2: 2 K-BILD Score Transformations (continued)

Breathlessness and Activities		Chest Symptoms		Psychological		Total	
				30	71.6	30	49.1
				31	75.6	31	49.7
				32	80.9	32	50.4
				33	88.6	33	51
				34	100	34	51.6
						35	52.2
						36	52.9
						37	53.5
						38	54.1
						39	54.8
						40	55.4
						41	56.1
						42	56.7
						43	57.4
						44	58.1
						45	58.8
						46	59.5
						47	60.2
						48	61
						49	61.8
						50	62.6
						51	63.5
						52	64.4
						53	65.3
						54	66.4
						55	67.5
						56	68.7
						57	70
						58	71.5
						59	73.2
						60	75.2
						61	77.6
						62	80.6
						63	84.6
						64	90.8
						65	100

9.2 L-PF SCORING INSTRUCTIONS

9.2.1 Symptoms

9.2.1.1 Dyspnea (items 1-12)

For items 1, 2, 4-10, and 12:

- if "No" is ticked and "A" is ticked, the score is 5
- if "No" is ticked and "B" is ticked, the item is not scored and does not contribute to the denominator
  if "No" is ticked and neither "A" nor "B" is ticked, the item is counted as missing and handled per "Instructions for Missing Items". Items for which this occurs DO contribute to the denominator

For items 3 and 11:

- scores equal whichever box is ticked (0-4)

$$\text{Dyspnea Symptom Score} = \frac{\text{sum of responses from items } 1-12}{\text{total score possible for items without "No"/"B" response}} \times 100$$

9.2.1.2 Cough (items 13-18)

$$\text{Cough Symptom Score} = \frac{\text{sum of responses from items } 13-18}{24} \times 100$$

9.2.1.3 Energy (items 19-23)

For items 19-22:

- scores equal whichever box is ticked (0-4)

For item 23:

- if "Yes" is ticked, score equals whichever box is ticked (0-4)
- if "No" is ticked, the item is not scored and does NOT contribute to denominator

$$\text{Energy Symptom Score} = \frac{\text{sum of responses from items } 19-23}{\text{total score possible (20 or 16 if response for 23 is "No")}} \times 100$$

9.2.1.4 Symptoms Total Score

$$\text{Symptoms Total Score} = \frac{(\text{Dyspnea Symptom} + \text{Cough Symptom} + \text{Energy Symptom})}{3}$$

9.2.2 Impacts

9.2.2.1 Dyspnea (items 1-6)

$$\text{Dyspnea Impact Score} = \frac{\text{sum of responses from items } 1-6}{24} \times 100$$

9.2.2.2 Cough (items 7-11)

$$\text{Cough Impact Score} = \frac{\text{sum of responses from items } 7-11}{20} \times 100$$

9.2.2.3 Energy (items 13-14)

$$\text{Energy Impact Score} = \frac{\text{sum of responses from items } 13-14}{8} \times 100$$

9.2.2.4 Global (items 12, 15-21)

$$\text{Global Impact Score} = \frac{\text{sum of responses from items } 12, 15-21}{32} \times 100$$

9.2.2.5 Impacts Total Score

$$\text{Impacts Total Score} = \frac{(\text{Dyspnea Impact} + \text{Cough Impact} + \text{Energy Impact} + \text{Global Impact})}{4}$$

9.3 PF-IQOLS SCORING INSTRUCTIONS

9.3.1 Summary score

The scoring of the PF-IQOLS follows Flanagan's practice in that a very straightforward 5-point Likert-type rating scale is used, and the scoring is analogous to that of the QOLS – i.e., the summary score is the average of the individual dimension ratings 1-16. The response option for all items ranges from 1 (no negative effect at all) to 5 (extremely negative effect). Therefore the score range for the summary score is 1.0-5.0. The individual item scores can be directly taken from the answered questions in the questionnaire. Please see more details in the questionnaire (Figure 9.3.1: 1 and Figure 9.3.1: 2)

Over the past four weeks, how much did your pulmonary fibrosis negatively affect your life in each of the following areas?

Consider the effects of the pulmonary fibrosis itself, the medications you take for pulmonary fibrosis (including supplemental oxygen if you use it), and anything you did to avoid, treat, or get medical care for symptoms of pulmonary fibrosis.

	No Negative Effect At All 1	Slightly Negative Effect 2	Moderately Negative Effect 3	Very Negative Effect 4	Extremely Negative Effect 5
1. *Material comforts* — things like a desirable home, good food, possessions, conveniences, an increasing income, and security for the future.	1	2	3	4	5
2. *Health and personal safety* — being physically fit and vigorous, free from anxiety and distress, and avoiding bodily harm.	1	2	3	4	5
3. *Relationships with your parents, brothers, sisters, and other relatives* — communicating, visiting, and doing things with, understanding, and helping and being helped by your relatives.	1	2	3	4	5
4. *Having and raising children* — being a parent and helping, teaching, and caring for your children.	1	2	3	4	5
5. *Close relationship with a husband, wife, or partner*	1	2	3	4	5
6. *Close friends* — sharing activities, interests, and views; being accepted, visiting, giving and receiving help, love, trust, support, guidance.	1	2	3	4	5
7. *Helping and encouraging others* — adults or children other than relatives or close friends. These can be your own efforts or efforts as a member of a church, club, or volunteer group.	1	2	3	4	5
8. *Participation in activities relating to local and national government and public affairs.*	1	2	3	4	5

IQOLS scale developed by Sandra R. Wilson, PhD. Dimensions from Flanagan, JC. *American Psychologist*. 1978;33:138-147, with addition by Burckhardt, CS, *et al*. *Research in Nursing & Health* 1989; 12:347-35.

Figure 9.3.1: 1 Part I of the PF-IQOLS Questionnaire.

Over the past four weeks, how much did your pulmonary fibrosis negatively affect your life in each of the following areas?

	No Negative Effect At All 1	Slightly Negative Effect 2	Moderately Negative Effect 3	Very Negative Effect 4	Extremely Negative Effect 5
9. *Learning*, attending school, improving your understanding, or gaining additional knowledge.	1	2	3	4	5
10. *Understanding yourself* — knowing your assets and limitations, knowing what life is all about and making decisions on major life activities. For some people, this includes religious or spiritual experiences; for others, it is developing an attitude toward life or a philosophy.	1	2	3	4	5
11. *Independence* — doing for yourself: being able to take care of and make decisions about your daily needs, personal care, where you live, and your financial affairs.	1	2	3	4	5
12. *Work* in a job or at home that is interesting, rewarding, and worthwhile.	1	2	3	4	5
13. *Expressing yourself* in a creative manner in music, art, writing, photography, practical activities, or in leisure time activities	1	2	3	4	5
14. *Socializing* — meeting other people, doing things with them, and hosting or attending parties or other social gatherings.	1	2	3	4	5
15. *Reading, listening to music, or observing* sporting events or entertainment.	1	2	3	4	5
16. *Participation in active recreation* such as playing sports, traveling and sightseeing, playing games or cards, singing, dancing, playing an instrument, acting, and other such activities.	1	2	3	4	5

IQOLS scale developed by Sandra R. Wilson, PhD. Dimensions from Flanagan, JC. *American Psychologist*. 1978;33:138-147, with addition by Burckhardt, CS, *et al*. *Research in Nursing & Health* 1989; 12:347-35.

Figure 9.3.1: 2 Part II of the PF-IQOLS Questionnaire

9.4 ADDITIONAL ANALYSES FOR REGIONAL SUBMISSIONS

For the Japanese submission, the analyses will be conducted according to "Format for Preparing the Common Technical Document for Submission of New Drug Applications to Reduce Total Review Time" (17).

For the analysis due to local regulatory submission (21), the following patient subsets are defined.

Table 9.4: 1 Patient subset definitions for local regulatory submission in Asia

Category	Patient Subset
Local	Japanese patients Chinese patients The patients in each local site and Asian race will be included (e.g. Japanese patients: Site in Japan and single race is Asian). If Asian race will be "Multiple" and multiple race responder will be "Asian & xxx", the patient will be handled as "Asian". (e.g. Japan: Site in Japan and multiple race is Asian & White.)
East Asia	East Asian patients The patients from East Asian countries or region (i.e. Japan, Korea and China) with Asian race. Asian race definition is same definition as for local category.

Some outputs from CTR Section 15 or section 16 will be selected for the patient subsets defined in Table 9.4: 1. Due to small sample size of some of the patient subsets, some tables might be skipped. The same analysis models as defined in Section 7 will be used. A detailed plan will be described in the CSAP. In the event of non-convergence, the same methods as described in Section 7.4.1 will be used to overcome the issue. But in case of very limited number of patients, the more simplified model or descriptive statistics may be used if convergence is not achieved via the described methods in Section 7. These patient subset analyses will not be produced as part of the CTR appendix and will not be described in the text of the CTR.

Other subset analysis than described in Table 9.4: 1 may be conducted in response to regulatory requests and the detailed plan for that case will be described in the CSAP.

9.5 PROGRAMMING CODE FOR INFERENTIAL ANALYSIS

9.5.1 Primary endpoint model

The key code is given below. Time is defined as duration since first trial drug intake divided by 365.25. Visit is the visit number. For the co-primary population, please remove HRCT as variable from the model.

```
ODS OUTPUT estimates=estimateout;
PROC MIXED data=FVC cl method=reml  order= formatted covtest;
        CLASS       patient treatment hrct;
        MODEL       endpoint= treatment treatment*time hrct fvcbaseline fvcbaseline*time /solution CL
                    ddfm=KR;
        RANDOM      intercept time/ type=un subject=patient;

        ESTIMATE    'Nintedanib' fvcbaseline*time meanbase treatment*time  0 1 / e CL;
        ESTIMATE    'Placebo' fvcbaseline*time meanbase treatment*time 1 0 / e CL;
        ESTIMATE    'Nintedanib 150mg bid – Placebo' treatment*time -1 1 / e CL;
RUN;
ODS OUTPUT CLOSE;
```

9.5.2 Primary endpoint subgroup analysis model:

Example subgroup used: Gender (Male / Female).

```
ODS OUTPUT   estimates=estimatesout Contrasts=contrastsout ;
PROC MIXED data=FVC cl method=reml  order= formatted covtest;
        CLASS  patient treatment hrct subgroup;
        MODELendpoint= treatment*subgroup treatment*subgroup*time
                    hrct fvcbaseline fvcbaseline*time
                    /solution CL ddfm=KR;
        RANDOM      intercept time/ type=un subject=patient;

        ESTIMATE    'Nintedanib - Female' fvcbaseline*time meanbase treatment * subgroup *time 0 0 1 0
                    / cl e;
        ESTIMATE    'Placebo - Female' fvcbaseline*time meanbase treatment * subgroup *time 1 0 0 0
                    / cl e;

        ESTIMATE    'Nintedanib - Male' fvcbaseline*time meanbase treatment * subgroup *time 0 0 0 1 /
                    cl e;
        ESTIMATE    'Placebo - Male' fvcbaseline*time meanbase treatment * subgroup *time 0 1 0 0
                    / cl e;

        ESTIMATE    'Nintedanib – Placebo Female'  subgroup * treatment *time
                     -1 0 1 0 / e CL;
        ESTIMATE    'Nintedanib – Placebo Male'   subgroup * treatment *time
                    0 -1 0 1 / e CL;

        CONTRAST    'F-test of heterogenety at 52 weeks'
                    subgroup * treatment *time -1 1  1 -1 / e;
RUN;
ODS OUTPUT CLOSE ;
```

9.5.3 Implementation of multiple imputation:

The following steps provide a guidance of how missing FVC data at week 52 will be implemented in patients of pattern 3 and 4 using multiple imputation. Refer to Table 7.4.2.1.2: 1 for the description of patterns and the different sensitivity analyses regarding handling of missing data.

1. Run the MMRM of the primary analysis (either including patients of pattern 2 only for sensitivity analyses 1 and 2 or on all patients for sensitivity analysis 3) to get the slope

estimates (SE) by treatment group. Let $\hat{\beta}_D$ and $\hat{\sigma}_D$ denote the slope (SE) estimates in Nintedanib 150 mg and $\hat{\beta}_P$ and $\hat{\sigma}_P$ the slope (SE) estimates in placebo.

2. Let β_D and β_P represent the true slopes in drug and placebo respectively with $f(\beta_D) \sim N(\hat{\beta}_D, \hat{\sigma}^2{}_D)$ and with $f(\beta_P) \sim N(\hat{\beta}_P, \hat{\sigma}^2{}_P)$, using the slope (SE) estimates obtained in step 1.

3. Imputation of missing week 52 data in patients of pattern 3:

   Multiple impute missing values (1000 imputations per patient). To do so, draw random slopes from the distributions defined in step 2. Considering that the withdrawal of a patient leading to missing data can occur at any time during the study, the timepoint of the last available FVC value has to be taken into account to impute the missing FVC value at 52 weeks.

   FVC week 52 imputed_{ij} = last FVC value available_i + $\hat{\beta}_{ij}$(time between date of last FVC value available_i and planned 52 week timepoint [days])

   where i denotes the indicator of the i^{th} patient, j the indicator of the j^{th} imputation and $\hat{\beta}_{ij}$ denotes a random slope sampled from the distribution mentioned in step 3a for the i^{th} patient in the j^{th} imputation. Depending on the sensitivity analysis, use either random slopes drawn from $f(\beta_D)$ for patients randomised to Nintedanib 150 mg in sensitivity analysis 1 or $f(\beta_P)$ for all patients in sensitivity analyses 2 and 3.

4. Imputation of missing week 52 data in patients of pattern 4:

   Multiple impute missing values (1000 imputations per patient). To do so, draw random slopes from the truncated distribution $f(\beta_P)/2$ restricted to the interval $(-\infty, \hat{\beta}_P]$ defined in step 2.

   See step 3 for further details on how to use these slopes to impute the missing values at week 52.

5. Run the MMRM of the primary analysis on each imputed dataset, using a "by _IMPUTATION_ " statement.

6. Combine the estimates obtained in step 5 using PROC MIANALYZE.

9.6 SPECIAL ADAM COMPOSITION INSTRUCTIONS

9.6.1 ADEX

The analysis of exposure will be based on the drug administration as planned and documented in the eCRF, i.e. the derivation of ADEX will be mainly based on EC. In case an administration entry for a visit is missing, it will be checked whether the patient was treated using the compliance data in XA; if this is the case, the missing drug administration data will

be replaced using the drug dispensation data in DA for the same visit if available. Treatment start will be the first treatment date/time as in ADEX.

9.7 LIST OF BI CUSTOMISED DRUG GROUPINGS OF INTEREST

Drug groupings of interest for this trial are shown in the following Table.

Table 9.7: 1 Concomitant Therapy groupings

CT grouping name	Type of grouping
Biologic DMARDs	SDG
Non-biologic DMARDs	SDG
Antithrombotic drugs	SDG

The table below shows the overview of medications that are considered to fall into the respective category of Disease-Modifying Anti-Rheumatic Drugs (DMARDS) with known hepatotoxic or gastrointestinal effects. Please note that an individual substance can occur in both categories.

Table 9.7: 2 DMARDS with known hepatotoxic or gastrointestinal effects

DMARDS with known hepatotoxic effects		DMARDS with known gastrointestinal effects	
WHO-DD decode	**WHO-DD code**	**WHO-DD decode**	**WHO-DD code**
LEFLUNOMIDE	01414801001	ABATACEPT	05514401001
METHOTREXATE	00113801001	LEFLUNOMIDE	01414801001
METHOTREXATE SODIUM	00113802001	METHOTREXATE	00113801001
TOCILIZUMAB	01759101001	METHOTREXATE SODIUM	00113802001
TOFACITINIB	07918601001	SULFASALAZINE	00001601001
TOFACITINIB CITRATE	07918602001	TOFACITINIB	07918601001
		TOFACITINIB CITRATE	07918602001

9.8 SPECIFICATION OF RESTRICTED AND FORBIDDEN CONCOMITANT THERAPIES

Categorisations and detailed information on WHO-DD code can be found in the following Table.

Table 9.8: 1 Prohibited or restricted concomitant therapies

Type of medication	Category of drugs	WHO-DD code
Prohibited	Full dose therapeutic anticoagulation or high dose antiplatelet therapy	0000000126
Prohibited	Pirfenidone	06225101001
Prohibited	Nintedanib	08403301001, 08403302001
Restricted	Oral corticosteroids (OCS)	0000000010
Restricted	Immunomodulatory medications for ILD	00001501001, 00001502001, 00021101001, 00021102001, 00549701001, 01275102001, 01275104001, 01402501001 01219901001, 01219902001

9.9 LIST OF POTENTIAL TERMS FOR HEPATIC INJURY DERIVATION

The table below shows the list of potentially relevant MedDRA preferred terms to support the derivation of a potential hepatic injury.

Table 9.9: 1 List of potentially relevant MedDRA preferred terms

Symptom	MedDRA
Vomiting	PT Vomiting
Fatigue	PT Fatigue
Nausea	PT Nausea
Right upper abdominal quadrant pain or tenderness	HLT “Gastrointestinal and abdominal pains (excl oral and throat)”
Fever	PT Pyrexia
Rash	BIcMQ skin rash (narrow)

9.10 LIST OF PREFERRED TERMS INCLUDED IN STROKE PHARMACOVIGILANCE ENDPOINT

Amaurosis fugax
Amyloid related imaging abnormalities
Basal ganglia haemorrhage
Basilar artery occlusion
Basilar artery thrombosis
Brachiocephalic artery occlusion
Brain stem haematoma
Brain stem haemorrhage
Brain stem infarction
Brain stem ischaemia
Brain stem stroke
Brain stem thrombosis
Carotid aneurysm rupture
Carotid arterial embolus
Carotid artery occlusion
Carotid artery thrombosis
Central nervous system haemorrhage
Cerebellar artery occlusion
Cerebellar artery thrombosis
Cerebellar embolism
Cerebellar haematoma
Cerebellar haemorrhage
Cerebellar infarction
Cerebellar ischaemia
Cerebral arteriovenous malformation haemorrhagic
Cerebral artery embolism
Cerebral artery occlusion
Cerebral artery thrombosis
Cerebral haematoma
Cerebral haemorrhage
Cerebral haemorrhage foetal
Cerebral haemorrhage neonatal
Cerebral infarction
Cerebral infarction foetal
Cerebral ischaemia
Cerebral thrombosis
Cerebrovascular accident
Embolic cerebral infarction
Embolic stroke
Haemorrhage intracranial
Haemorrhagic cerebral infarction
Haemorrhagic stroke
Haemorrhagic transformation stroke
Intracranial haematoma
Intracranial tumour haemorrhage

Intraoperative cerebral artery occlusion
Intraventricular haemorrhage
Intraventricular haemorrhage neonatal
Ischaemic cerebral infarction
Ischaemic stroke
Lacunar infarction
Lateral medullary syndrome
Pituitary haemorrhage
Pituitary infarction
Post procedural stroke
Precerebral artery occlusion
Putamen haemorrhage
Reversible ischaemic neurological deficit
Ruptured cerebral aneurysm
Stroke in evolution
Subarachnoid haemorrhage
Subarachnoid haemorrhage neonatal
Subdural haemorrhage neonatal
Thalamic infarction
Thalamus haemorrhage
Thrombotic cerebral infarction
Thrombotic stroke
Transient ischaemic attack
Vertebral artery occlusion
Vertebral artery thrombosis

9.11 LIST OF TERMS FOR THE CATEGORISATION OF BLEEDING EVENTS

SMQ Haemorrhage terms (excl laboratory terms) (narrow)

9.12 ADDITIONAL SENSITIVITY ANALYSES

9.12.1 Tipping point analysis

An additional approach compared to the already discussed methods in Section 7.4.2 is to evaluate the robustness of the primary analysis to the missing at random (MAR) assumption by performing a tipping point sensitivity analysis. The aim of the tipping point approach is to assess how severe departures from MAR could be in order to reverse conclusions from the primary analysis under different assumptions regarding a declining persistence of efficacy post withdrawal of the randomised treatment. The robustness of the results will be discussed based on the magnitude of deviations from MAR required to change the results.

As a first step non-monotone missing data will be imputed m = 100 times using MCMC (Markov Chain Monte Carlo) to generate m data sets of longitudinal spirometry data (FVC [mL]) with monotone missingness pattern. Such a missingness pattern is the pre-requisite for subsequently applying sequential imputation and means that once a patient has a missing FVC value at a particular time point, FVC values at all subsequent time points also have missing

values. The seemingly large number of imputations (m=100) is chosen to minimize the standard error of those FVC estimates that are produced to fill the missing values. The seed number will be set to 1199247.

Once the monotone missing pattern has been created, the tipping point analysis for the longitudinal FVC data can commence and will be based on the Multiple Delta Adjustment Method. Data in each of the 100 generated datasets, now exhibiting a monotone missingness pattern, will be imputed once by using sequential regression. A delta adjustment will be added to each imputed value. The value of each delta adjustment will be given by:

$$\delta_{ij} = S_i * \varphi_j$$

δ_{ij} = adjustment for a patient in treatment $_i$ at visit $_j$

S_i = shift parameter for treatment $_i$ (i=1,2)

φ_j = number of weeks between visit $_{j-1}$ and visit $_j$

For patients in both treatment arms, the delta adjustment ($\boldsymbol{\delta}$) will be proportional to the time between the visits. For patients with more than one monotone missing visit, multiple adjustments must be applied and since the imputation method is sequential, the effect of the adjustments is cumulative. S_1 and S_2 therefore represent the slope of a linear adjustment over time. This step will be repeated for a variety of combinations of S_1 and S_2. The reason for selection of ranges of S_1 and S_2 is described below.

For each combination of shift values S_1 and S_2 the m complete data sets will be analysed using the primary analysis model. The estimate of treatment difference at week 52 will be derived. Rubin's rules (16) will be used to combine the results from the analyses for each pair of S_1 and S_2 to provide the required inferences for all combinations of shift parameters S_1 and S_2. Point estimates for treatment differences and respective p-values will be reported for all combinations of S_1 and S_2 in a cross-table.

The selection of shift values S_1 and S_2 is based on considerations of plausible and implausible cumulative changes of FVC (ml) values at week 52 (the end of Part A) of the patient population included into the trial, as well as observations made based on INPULSIS-1 and INPULSIS-2, given the possible range of missing data patterns.

In a first step the following combinations of S_1 and S_2 will be tested to cover a clinically reasonable space:

- S_1, S_2, = -20, -16, -13, -10, -6, -3, 0, +3, +6, +10 (resulting in 10x10 = 100 combinations matrix).

These shift parameters (-20 to +10) translate into

- Covering a median adjustment in imputed FVC (mL) at week 52 on patient level of -453 mL at the value -20
    - (-320 ml with only 1 missing value at week 52 to -749 ml with 7 missing values up to week 52)

- To a median adjustment in imputed FVC (mL) at week 52 on patient level of +227 mL at the value +10
    - (+160 ml with 1 missing value at week 52 to +374 ml with 7 missing values up to week 52)

See Table 9.12.1: 1 for additional details.

Table 9.12.1: 1 Overview of the median adjustment in imputed FVC (mL) at week 52

Shift parameter	Median adjustment in imputed FVC (mL) at week 52 on patient level							
	Overall	1 missing value	2 missing values	3 missing values	4 missing values	5 missing values	6 missing values	7 missing values
-20	-453	-320	-467	-626	-677	-695	-718	-749
-18	-408	-288	-420	-563	-609	-625	-646	-674
-16	-363	-256	-374	-500	-541	-556	-575	-599
-14	-317	-224	-327	-438	-474	-486	-503	-524
-12	-272	-192	-280	-375	-406	-417	-431	-449
-10	-227	-160	-234	-313	-338	-347	-359	-374
-8	-181	-128	-187	-250	-271	-278	-287	-299
-6	-136	-96	-140	-188	-203	-208	-215	-225
-4	-91	-64	-93	-125	-135	-139	-144	-150
-2	-45	-32	-47	-63	-68	-69	-72	-75
0	0	0	0	0	0	0	0	0
2	45	32	47	63	68	69	72	75
4	91	64	93	125	135	139	144	150
6	136	96	140	188	203	208	215	225
8	181	128	187	250	271	278	287	299
10	227	160	234	313	338	347	359	374

After investigation of estimates and p-values from these analyses further combinations with refined ranges of S_1 and S_2 may be used for the tipping point analysis.

9.13 LIST OF UNDERLYING ILD GROUPINGS FOR "OTHER FIBROSING ILD" ENTRIES

Patients, who have been entered in the eCRF as "other fibrosing ILD" were re-grouped for the analysis based on the following table (Table 9.13: 1).

Table 9.13: 1 Other Fibrosing ILD grouping

Other Fibrosing ILD	Grouping
ILD ASSOCIATED WITH LUPUS	autoimmune ILD
POLYMYOSITIS CTD ILD	autoimmune ILD
DESQUAMATIVE INTERSTITIAL PNEUMONIA	other ILD
PULMONARY FIBROSIS WITH AUTO-IMMUNE FEATURES	autoimmune ILD
CHRONIC EOSINOPHILIC PNEUMONIA	other ILD
FIBROSIS AND EMPHYSEMA	other ILD
PLEURO-PULMONARY FIBROELASTOSIS	other ILD
RESPIRATORY BRONCHIOLITIS-ILD WITH FIBROSIS	other ILD
PPFE (PLEUROPARENCHYMAL FIBROELASTOSIS)	other ILD
FIBROELASTOSIS	other ILD
LIPOIDIC FIBROSIS	other ILD
AFTER CHEMOTHERAPY	exposure related ILD
UCTD-ILD WITH POSITIVE ANTI-PL7 AUTO-AB	autoimmune ILD
OVERLAP NSIP - OP. NO UNDERLYING CTD	other ILD
OUTCOME OF ACUTE INTERSTITIAL PNEUMONIA	other ILD
SJOGREN DISEASE ILD	autoimmune ILD
IDIOPATHIC PLEUROPARENCHYMAL FIBROELASTOSIS (PPFE)	other ILD
SJOGREN'S SYNDROME	autoimmune ILD
IDIOPATHIC PLEUROPARENCHYMAL FIBROELASTOSIS	other ILD
IGG4 RELATED LUNG DISEASE	other ILD
SJOGREN SYNDROME ASSOCIATED ILD	autoimmune ILD
COMBINED PULMONARY FIBROSIS AND EMPHYSEMA	other ILD
SJOGREN'S SYNDROME ILD	autoimmune ILD
PULMONARY ALVEOLAR PROTEINOSIS	other ILD
MPA ASSOCIATED ILD	autoimmune ILD
ANCA ASSOCIATED ILD	autoimmune ILD
PLUEROPARENCHYMAL FIBROELASTOSIS	other ILD
CRYPTOGENICORGANIZING PNEUMONIA,IDIOPATHICDISEASE	other ILD
CRYPTOGENICORGANIZINGPNEUMONIA,IDIOPATHIC DISEASE	other ILD
UNDIFFERENTIATED CONNECTIVE TISSUE DISEASE-ILD	autoimmune ILD
CTD-OP	autoimmune ILD
UNCLASSIFIED CONNECTIVE TISSUE DISEASE	autoimmune ILD
SJOGREN SYNDROME	autoimmune ILD
CTD-ILD	autoimmune ILD

Table 9.13: 1 Other Fibrosing ILD grouping (continued)

Other Fibrosing ILD	Grouping
LUNG FIBROSIS IN ANTISYNTHETASE SYNDROME	autoimmune ILD
INTERSTITIAL PNEUMONIA WITH AUTOIMMUNE FEATURES	autoimmune ILD
IPAF	autoimmune ILD
CRYPTOGENIC ORGANIZING PNEUMONIA	other ILD
SJÃ–GREN SYNDROME	autoimmune ILD
SYSTEMIC LUPUS ERYTHEMATOSUS-ILD	autoimmune ILD
CT-ILD	autoimmune ILD
PLEUROPARENCHYMAL FIBRIOELASTOSIS	other ILD

10. HISTORY TABLE

Table 10: 1 History table

Version	Date (DD-MMM-YY)	Author	Sections changed	Brief description of change
Initial	**04-JAN-17**	[redacted]	None	This is the initial TSAP with necessary information for trial conduct
Final	**14-DEC-18**	[redacted]	1-9	This is the final TSAP