STATISTICAL ANALYSIS PLAN

Project Number:	GLPG1690		
Study Number:	GLPG1690-CL-303 & 304		
Study Title:	Two Phase 3, randomized, double-blind, parallel-group, placebo-controlled multicenter studies to evaluate the efficacy and safety of two doses of GLPG1690 in addition to local standard of care for minimum 52 weeks in subjects with idiopathic pulmonary fibrosis		
Development Phase:	3		
Status:	Final		
Version:	2.0	Date:	25-Mar-2021
EudraCT:	2018-001405-87 & 2018-001406-29	IND:	130687
Sponsor:	Galapagos NV, Generaal De Wittelaan L11 A3, 2800 Mechelen, Belgium		
Biostatistician	[redacted], MSc		
	[redacted], MSc		
Medical director	[redacted], MD		

Confidentiality Statement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

VERSION HISTORY

SAP Amendment #	Date	Description of changes
SAP Version 1.0	13/02/2019	First version
SAP Version 2.0	25/25/2021	- Changes to planned analysis due to termination GLPG1690-CL-303/304 program (text in *Italic font* will not be performed) - Addition Section 7.7: exclusion of data in case of quality issues. - Section 9.2.7: addition more details on standard of care analyses - Section 9.4.2.5: Scoring based on CTCAE/Other Gradings: updated in agreement with clinicians - Addition COVID-19 analyses - Changes to CSP sections due to amendments CSP - General editorial changes to improve wording

LIST OF ABBREVIATIONS

%FVC	percent predicted forced vital capacity
6MWT	6 Minute Walk Test
AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
ATC	anatomical therapeutic classification
ATS	American Thoracic Society
BMI	body mass index
C_{Cr}	estimated creatinine clearance, calculated according to Cockcroft-Gault calculation
CEAC	clinical endpoint adjudication committee
CI	confidence interval
CSR	clinical study report
CTCAE	Common Terminology Criteria for Adverse Events
DBP	diastolic blood pressure
DLCO	diffusing capacity of the lung for carbon monoxide
ECG	electrocardiogram
eCRF	electronic case report form
EoSA	end of study assessments
EoST	end of study treatment
EQ-5D	5-Dimensions Questionnaire
ETD	early treatment discontinuation
EU	European Union
FAS	full analysis set
[redacted]	[redacted]
FEF_{25-75}	forced expiratory flow between 25% and 75% of exhaled volume
FEV_1	forced expiratory volume in 1 second
FVC	forced vital capacity
FSH	follicle-stimulating hormone
Hb	hemoglobin
HRCT	high-resolution computed tomography
ICF	informed consent form
ICH	International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use
IDMC	independent data monitoring committee
IP	investigational product
IPF	idiopathic pulmonary fibrosis

ISE	integrated summary of efficacy
IWRS	interactive web response system
K-BILD	King's brief interstitial lung disease
LB	lung biopsy
LCQ	Leicester Cough Questionnaire
LLN	lower limit of normal range
LS	least squares
MAR	missing at random
MMRM	mixed-effects model for repeated measures
MNAR	missing not at random
[redacted]	[redacted]
PK	pharmacokinetic(s)
PP	per protocol
QTcF	QT interval corrected for the heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SBP	systolic blood pressure
SD	standard deviation
SE	standard error
SGRQ	St. George's Respiratory Questionnaire
SI	international system of units
SOC	standard of care
TEAE	treatment-emergent adverse event
ULN	upper limit of the normal range
US	United States
VAS	Visual Analogue Scale
WBC	white blood cell
WHO	world health organization

1. INTRODUCTION

This statistical analysis plan (SAP) describes the analyses for each of the GLPG1690-CL-303 and GLPG1690-CL-304 final analyses for the clinical study reports (CSR). Since the studies are identically designed, only one SAP is covering both CSRs.

A decision was taken to halt the ISABELA Phase 3 clinical studies (study numbers GLPG1690-CL-303 and GLPG1690-CL-304) with the investigational autotaxin inhibitor ziritaxestat (GLPG1690) in patients with idiopathic pulmonary fibrosis. The decision was based on the recommendations of the Independent Data Monitoring Committee (IDMC) which, following a regular review of unblinded data on February 9, concluded that ziritaxestat's benefit-risk profile no longer supported continuing these studies.

In view of the above, no market authorisation application will be submitted for ziritaxtestat for the treatment of IPF. Therefore, it was decided to limit the analyses to the primary, key secondary and safety endpoints and any analyses important to better understand the reason for the results seen, or relevant for scientific knowledge for future compounds. Since no treatment effect was observed and the study was stopped for lack of efficacy, sensitivity analyses to investigate the robustness of results will also be limited.

The results will be presented in an abbreviated CSR.

All descriptions as documented in the original SAP were kept for reference, but endpoints and analyses deemed not applicable for the abbreviated CSR are in *Italic font*, and will not be included for the final analysis. A list of the analyses that will not be performed can be found in section 9.1.

Analyses on the pooled data from both studies for the integrated summaries of efficacy and safety were planned to be very similar (adjusted for study as appropriate) and to be detailed in a separate pooled-analysis SAP. As no market authorisation application will be submitted, these pooled analyses will not be performed and the pooled-analysis SAP will not be written.

Technical details on derivations and mock tables, listings and figures will be presented in a separate document.

A separate SAP will be written for the population pharmacokinetics [redacted] (PK[redacted]) analyses.

2. STUDY DESIGN AND OBJECTIVES

The studies are identically designed with the same objectives, endpoints, etc.

All objectives and endpoints as described in the CSP were kept as reference in the below sections, but endpoints that will not be described in the abbreviated CSR are in *Italic font*. Details on the analyses that will not be performed can be found in section 9.1.

2.1. STUDY OBJECTIVES

2.1.1. Primary Objectives

To evaluate the efficacy of two doses of GLPG1690 in addition to local standard of care compared to placebo in subjects with idiopathic pulmonary fibrosis (IPF) as evaluated by the rate of decline of forced vital capacity (FVC) over a period of 52 weeks.

2.1.2. Key Secondary Objectives

To evaluate the impact of two doses of GLPG1690 in addition to local standard of care compared to placebo in subjects with IPF on:

- disease progression defined as deterioration of FVC or all-cause mortality at 52 weeks
- respiratory-related hospitalization until the end of the study
- changes in quality of life (measured by St. George's Respiratory Questionnaire [SGRQ] total score) at 52 weeks

2.1.3. Other Secondary Objectives

- To evaluate the efficacy of two doses of GLPG1690 in addition to local standard of care compared to placebo in subjects with IPF as evaluated by the rate of decline of FVC until the end of the study
- To evaluate the impact of two doses of GLPG1690 in addition to local standard of care compared to placebo in subjects with IPF on:
    - disease progression defined as deterioration of FVC or all-cause mortality until the end of the study
    - changes in quality of life (measured by SGRQ total score) until the end of the study
    - all-cause non-elective hospitalization until the end of the study
    - respiratory-related mortality until the end of the study
    - lung transplant until the end of the study
    - acute IPF exacerbation until the end of the study
    - all-cause mortality or lung transplant until the end of the study
    - all-cause mortality, or lung transplant, or qualifying for lung transplant until the end of the study
    - all-cause mortality, deterioration of FVC, or respiratory-related hospitalization until the end of the study
    - all-cause mortality or respiratory-related hospitalizations until the end of the study
- To evaluate the effect of two doses of GLPG1690 in addition to local standard of care compared to placebo on the changes from baseline in FVC at 52 weeks and until the end of the study
- To evaluate the safety and tolerability of two doses of GLPG1690 in addition to local standard of care compared to placebo until the end of the study
- *To evaluate changes compared to placebo in subjects with IPF in:*
    - *cough-related quality of life (measured by the Leicester Cough Questionnaire [LCQ] and by the Visual Analogue Scale [VAS] Cough and Urge to Cough) at 52 weeks and until the end of the study*

- *quality of life (measured by EuroQOL 5-Dimensions Questionnaire [EQ-5D] and King's Brief Interstitial Lung Disease [K-BILD] total score and domains over time) at 52 weeks and until the end of the study*

– To evaluate the PK of GLPG1690, pirfenidone, and nintedanib (as appropriate) in subjects with IPF at 52 weeks and until the end of the study
– To evaluate the effect of two doses of GLPG1690 in addition to local standard of care compared to placebo in subjects with IPF on the changes in functional exercise capacity measured by the 6-Minute Walk Test (6MWT) at 52 weeks and until the end of the study
– To evaluate the effect of two doses of GLPG1690 in addition to local standard of care compared to placebo on the diffusing capacity of the lung for carbon monoxide (DLCO) at 52 weeks and until the end of the study

2.1.4. [redacted]

[redacted]

2.2. Study Endpoints

2.2.1. Primary Endpoint

– Rate of decline of FVC (in mL) over a period of 52 weeks.

2.2.2. Secondary Endpoints

2.2.2.1. Key Secondary Endpoints

– Disease progression defined as the composite endpoint of first occurrence of ≥10% absolute decline in percent predicted forced vital capacity (%FVC) or all-cause mortality at 52 weeks
– Time to first respiratory-related hospitalization until the end of the study
– Change from baseline in the SGRQ total score at 52 weeks

2.2.3. Other Secondary Endpoints

– Rate of decline of FVC (in mL) until the end of the study
– Disease progression defined as the composite endpoint of first occurrence of ≥10% absolute decline in %FVC or all-cause mortality until the end of the study
– Change from baseline in the SGRQ total score until the end of the study
– Time to first all-cause non-elective hospitalization until the end of the study
– Time to respiratory-related mortality until the end of the study
– Time to lung transplant until the end of the study
– Time to first acute IPF exacerbation until the end of the study

- Time to all-cause mortality or lung transplant until the end of the study
- Time to all-cause mortality, or lung transplant, or qualifying for lung transplant until the end of the study
- Time to all-cause mortality, ≥10% absolute decline in %FVC, or respiratory-related hospitalizations until the end of the study
- Time to all-cause mortality or respiratory-related hospitalizations until the end of the study
- FVC analyses at 52 weeks and until the end of the study:
  - absolute and relative change from baseline of FVC and %FVC
  - absolute categorical change of %FVC until the end of the study: decrease by >5, increase by >5, and change within ≤5
  - absolute categorical change of %FVC until the end of the study: decrease by >10, increase by >10, and change within ≤10
- Safety and tolerability over time until the end of the study
- *Changes from baseline in cough-related quality of life, assessed by the LCQ total score and domains over time, and the VAS Cough and Urge to Cough, at 52 weeks and until the end of the study*
- *Changes from baseline in quality of life, assessed by the EQ-5D, K-BILD total score and domains over time, at 52 weeks and until the end of the study*
- Plasma concentration of GLPG1690, pirfenidone, and nintedanib (as appropriate) at 52 weeks and until the end of the study
- Change from baseline in functional exercise capacity, assessed by the 6MWT distance, at 52 weeks and until the end of the study
- Change from baseline in DLCO (corrected for hemoglobin [Hb]) at 52 weeks and until the end of the study

2.2.4. [redacted]

2.2.5. Adjudication

These endpoints will be adjudicated by the clinical endpoint adjudication committee (CEAC):

- Mortality: all-cause and respiratory-related (2 separate endpoints)
- Hospitalization: all-cause and respiratory-related (2 separate endpoints)
- Hospitalization for non-elective lung transplant
- Acute IPF exacerbation
- All-cause mortality or hospitalization for non-elective lung transplant (composite endpoint)

– All-cause mortality, or hospitalization for non-elective lung transplant, or hospitalization for qualifying for lung transplant (composite endpoint)
– All-cause mortality, hospitalization for ≥10% absolute decline in %FVC, or respiratory-related hospitalizations (composite endpoint)
– All-cause mortality or respiratory-related hospitalizations (composite endpoint)

Adjudication endpoints will be used in the analyses and not the corresponding investigator-reported events. Consistency with the investigator-reported information will be investigated, however.

2.3. Study Design

This clinical Phase 3 study is a randomized, double-blind, parallel-group, placebo-controlled multicenter study designed to evaluate the efficacy and safety of two doses (200 mg q.d. and 600 mg q.d.) of orally administered GLPG1690 in addition to local standard of care for at least 52 weeks in adult subjects with a centrally confirmed diagnosis of IPF. Local standard of care for IPF is defined as receiving either pirfenidone or nintedanib at a stable dose for at least two months before screening, and during screening; or neither pirfenidone or nintedanib (for any reason). A stable dose is defined as the highest dose tolerated by the subject during those two months. A total of approximately 750 subjects with confirmed diagnosis of IPF will be randomized, 250 subjects in each treatment group (GLPG1690 600 mg q.d., GLPG1690 200 mg q.d., or matching placebo).

The diagnosis of IPF will be confirmed by central reading of the chest high-resolution computed tomography (HRCT) and by central review of the available lung biopsy (LB), if required, based on the Fleischner White Paper as described in Section 6.3 and Appendix 1 of the CSP.

A schematic diagram of the clinical study design, procedures, and stages is provided CSP

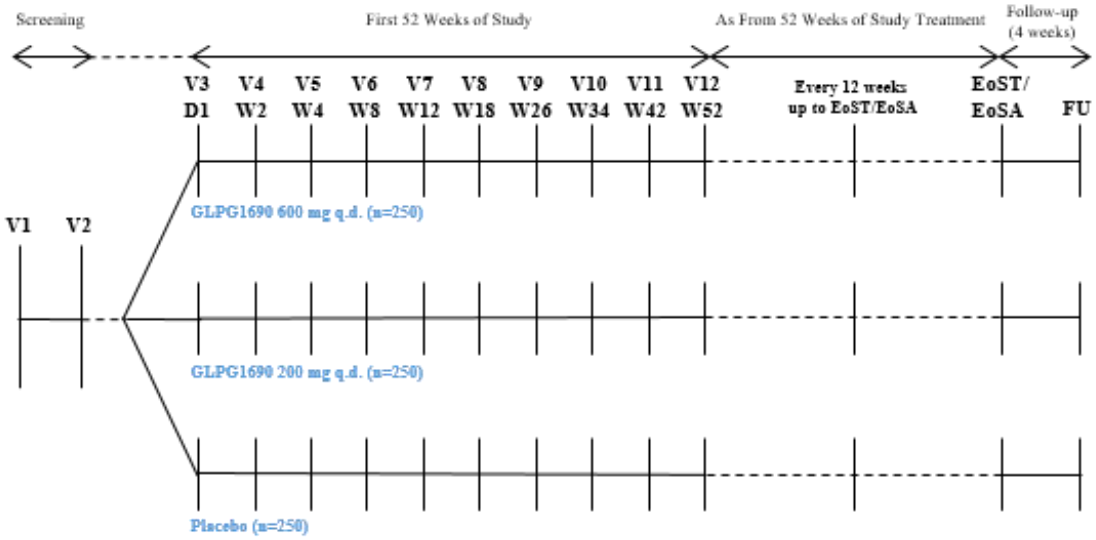

Figure 1: Schematic Study Overview

D=Day, EoST/EoSA=end of study treatment/end of study assessments, FU=follow-up, V=Visit, W=Week.

Enrolled subjects will come to the clinical study center at screening (two visits from Day -28 to Day -1). Visit 2 of screening can only take place after confirmed IPF diagnosis on HRCT and LB (if applicable).

At Visit 3, eligible subjects will be randomized in a 1:1:1 ratio to receive GLPG1690 600 mg q.d., GLPG1690 200 mg q.d., or matching placebo for at least 52 weeks. Randomization will be stratified for background local standard of care for treatment of IPF.

Subjects will come to the clinical study center on Day 1 (baseline), Weeks 2, 4, 8, 12, 18, 26, 34, 42, 52, and every 12 weeks thereafter. Additional unscheduled visits are allowed if, in the investigator's opinion, further evaluation (clinical, laboratory, or other) is needed.

When the last subject reaches 52 weeks (Visit 12), subjects still taking IP and who had a scheduled visit at the clinical study center within 6 weeks before this date will be contacted by their investigator (phone call) to discontinue IP within 7 days after this date. Then the last scheduled visit will be documented as the end of study treatment (EoST) visit. These subjects will be invited to their follow-up visit scheduled at the clinical study center 4 weeks after their last dose of IP (±7 days). For subjects still taking IP and who had their last scheduled visit more than 6 weeks before the date that the last subject reaches 52 weeks into the study (Visit 12), a scheduled visit will be planned within 2 weeks of this date for an EoST visit, followed by a follow-up visit 4 weeks later (visit at clinical center or phone call). Note that these subjects will continue IP intake until their EoST visit.

For subjects who, for any Coronavirus disease (COVID-19)-related reason, cannot perform study procedures, extended visit windows, the possibility to conduct phone/televisits and home (or other remote location) visits, and alternative assessment procedures are detailed in Section 6.1.1 of the CSP.

Subjects who discontinue IP early (early treatment discontinuation [ETD]), with the exception of patients lost to follow-up and patients who withdraw consent, will be encouraged to complete all following visits and evaluations as originally planned per protocol.

In particular, the subject will be requested if at all possible to attend:

- the first scheduled visit after their Early Treatment Discontinuation (ETD)
- The Week 26 visit (if not done before IP discontinuation)
- the Week 52 visit (if not done before IP discontinuation)
- after Week 52, visits every 24 weeks up to End of Study Assessments (EoSA), as described in Section 6.11.2 of the CSP.

For procedures in case of ETD, refer to Section 4.5.4 of the CSP. In the exceptional case that a clinical study center visit cannot be attended, a phone call by the investigator will be made to evaluate safety, in combination with scheduled clinical study center visits.

Each subject will have a screening period of maximum 28 days. The duration of study treatment for individual subjects will vary from 52 weeks for the last enrolled subject to e.g. 132 weeks for the first enrolled subject (assuming a recruitment period of 80 weeks, the actual duration will depend on the actual time to recruit subjects for the study). A follow-up visit (at the clinical study center or by phone call, decided by the investigator) is planned 4 weeks after the EoST/end of study assessment (EoSA) visit.

At the end of this study, treatment with IP in an optional extension study (under a separate study protocol) may be offered to all eligible subjects, provided Independent Ethics Committee (IEC)/Institutional Review Board (IRB) and Competent Authority approvals for such an extension are granted.

To protect the safety and integrity of the study data, an independent data monitoring committee (IDMC) will be implemented, as well as a clinical endpoint adjudication committee (CEAC). Charters are in place for all committees.

2.4. Clinical Study Protocol

This SAP was based on the version 7 dated 08 June 2020 of the two clinical study protocols.

2.5. Schedule of Activities

For detailed instructions on the clinical study procedures, please see referred Section 6.1 of the protocols.

2.5.1. Schedule of Activities: Screening and First 52 Weeks of Study Treatment

For subjects who, for any COVID-19-related reason, cannot perform study procedures, extended visit windows, the possibility to conduct phone/televisits and home (or other remote location) visits, and alternative assessment procedures are detailed in the CSP

EVENT	**SCREENING PERIOD**		**FIRST 52 WEEKS OF STUDY TREATMENT**										
Study visit	**1**	**[2]**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**ETD**
Study days (D) or weeks (W) ± visit window	**D-28 to D-1**		**D1**	**W2 (D15) ±2 d**	**W4 (D29) ±2 d**	**W8 (D57) ±4 d**	**W12 (D85) ±4 d**	**W18 (D127) ±4 d**	**W26 (D183) ±4 d**	**W34 (D237) ±4 d**	**W42 (D295) ±4 d**	**W52 (D365) ±4 d**	
Informed consent	✓[1]												
FSH (if applicable) and serology	✓												
HRCT sent for central review	✓												
LB (if available) sent for central review	✓												
Inclusion/exclusion criteria	✓	✓											
Demographics	✓												
Medical history	✓												
Alcohol consumption and smoking habits	✓		✓						✓			✓	✓
Physical examination	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Vital signs	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

EVENT	SCREENING PERIOD		FIRST 52 WEEKS OF STUDY TREATMENT										
Study visit	**1**	**[2]**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**ETD**
Study days (D) or weeks (W) ± visit window	**D-28 to D-1**		**D1**	**W2 (D15) ±2 d**	**W4 (D29) ±2 d**	**W8 (D57) ±4 d**	**W12 (D85) ±4 d**	**W18 (D127) ±4 d**	**W26 (D183) ±4 d**	**W34 (D237) ±4 d**	**W42 (D295) ±4 d**	**W52 (D365) ±4 d**	
ECG triplicate recording	✓												
ECG triplicate recording before IP intake			✓	✓									
ECG triplicate recording 2-3 hours after IP intake			✓	✓									
ECG single recording before IP intake					✓	✓	✓	✓	✓	✓	✓	✓	
ECG single recording 2-3 hours after IP intake					✓		✓			✓			
ECG single recording													✓
Pregnancy test (serum)	✓												
Pregnancy test (urine)			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Clinical laboratory tests [2]	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Spirometry	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DLCO	✓		✓						✓			✓	
[redacted]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
6MWT [redacted]	✓	✓	✓						✓			✓	

EVENT	SCREENING PERIOD		FIRST 52 WEEKS OF STUDY TREATMENT										
Study visit	**1**	**[2]**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**ETD**
Study days (D) or weeks (W) ± visit window	**D-28 to D-1**		**D1**	**W2 (D15) ±2 d**	**W4 (D29) ±2 d**	**W8 (D57) ±4 d**	**W12 (D85) ±4 d**	**W18 (D127) ±4 d**	**W26 (D183) ±4 d**	**W34 (D237) ±4 d**	**W42 (D295) ±4 d**	**W52 (D365) ±4 d**	
Randomization by IWRS			✓										
Dispense IP			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Collect IP/perform drug accountability				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Diary card dispensing and collection (as applicable per visit) for drug accountability			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
EQ-5D and SGRQ (ePRO)	✓		✓						✓			✓	✓
LCQ and K-BILD questionnaire (ePRO)	✓		✓						✓			✓	✓
VAS Cough and Urge to Cough	✓		✓						✓			✓	✓
Intake P and N at clinical study center			✓						✓			✓	
PK blood samples 2-4 hours after N morning intake [3]	✓												
PK blood samples 5-10 hours after N morning intake [3]	✓												

EVENT	SCREENING PERIOD		FIRST 52 WEEKS OF STUDY TREATMENT										
Study visit	**1**	**[2]**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**ETD**
Study days (D) or weeks (W) ± visit window	**D-28 to D-1**		**D1**	**W2 (D15) ±2 d**	**W4 (D29) ±2 d**	**W8 (D57) ±4 d**	**W12 (D85) ±4 d**	**W18 (D127) ±4 d**	**W26 (D183) ±4 d**	**W34 (D237) ±4 d**	**W42 (D295) ±4 d**	**W52 (D365) ±4 d**	
PK blood samples before IP and P or N intake			✓						✓			✓	
PK blood samples before IP, but after P or N intake [4]							✓			✓			
PK blood samples 2-6 hours after N intake and before IP [3]				✓	✓	✓	✓	✓		✓	✓		
PK blood samples 2 to 3 hours after IP intake							✓			✓			
PK blood samples													✓
[redacted]	✓												✓
[redacted]			✓				✓		✓	✓		✓	
[redacted]							✓			✓			
[redacted]	✓												✓

EVENT	SCREENING PERIOD		FIRST 52 WEEKS OF STUDY TREATMENT										
Study visit	**1**	**[2]**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**ETD**
Study days (D) or weeks (W) ± visit window	**D-28 to D-1**		**D1**	**W2 (D15) ±2 d**	**W4 (D29) ±2 d**	**W8 (D57) ±4 d**	**W12 (D85) ±4 d**	**W18 (D127) ±4 d**	**W26 (D183) ±4 d**	**W34 (D237) ±4 d**	**W42 (D295) ±4 d**	**W52 (D365) ±4 d**	
[redacted]			✓				✓		✓			✓	
[redacted]				✓									
Vital status												✓	
Study medication intake			Throughout the treatment period										
AE assessment	Throughout the study												
Concomitant medication assessment and documentation	Throughout the study												

[1] The ICF signature is the start of the 28-day screening period and will be used as the date of Visit 1
[2] CK-MB to be measured if CK is elevated
[3] Only for subjects taking nintedanib at screening and randomization, until they stop taking nintedanib
[4] For subjects not taking nintedanib at screening and randomization and those stopping nintedanib during the study
d=days, ePRO=electronical patient-reported outcome, ETD=early treatment discontinuation, N=nintedanib, P=pirfenidone.
The preferred sequence of study assessments is described I the CSP.

2.5.2. Schedule of Activities: After 52 Weeks of Study Treatment

For subjects who, for any COVID-19-related reason, cannot perform study procedures, the possibility to conduct phone/televisits and home (or other remote location) visits, and alternative assessment procedures are detailed in the CSP.

EVENT	**AFTER 52 WEEKS OF STUDY TREATMENT**				
Study visit	**Every 12 weeks up to EoST/EoSA ±7 d**	**Every 24 weeks up to EoST/EoSA3 ±7 d**	**ETD**	**EoST/EoSA***	**FU***
Study weeks ± visit window					**(4 weeks after EoST/EoSA) ±7 d**
Alcohol consumption and smoking habits	✓		✓	✓	
Physical examination	✓		✓	✓	
Vital signs	✓		✓	✓	
ECG single recording			✓	✓	
ECG single recording before IP intake	✓				
Pregnancy test (urine)	✓		✓	✓	
Clinical laboratory tests	✓		✓	✓	
Spirometry	✓		✓	✓	
DLCO		✓			
[redacted]	✓		✓	✓	
6MWT [redacted]		✓			
Dispense IP	✓				
Collect IP/perform drug accountability	✓		✓	✓	

EVENT	AFTER 52 WEEKS OF STUDY TREATMENT				
Study visit	**Every 12 weeks up to EoST/EoSA**	**Every 24 weeks up to EoST/EoSA3**	**ETD**	**EoST/EoSA***	**FU***
Study weeks ± visit window	**±7 d**	**±7 d**			**(4 weeks after EoST/EoSA) ±7 d**
Diary card dispensing and collection (as applicable per visit) for drug accountability	✓		✓	✓	
EQ-5D and SGRQ (ePRO)		✓	✓	✓	
LCQ and K-BILD questionnaire (ePRO)		✓	✓	✓	
VAS Cough and Urge to Cough		✓	✓	✓	
PK blood samples			✓	✓	
PK blood samples before IP, but 2-6 hours after N intake[1]	✓				
PK blood samples before IP, but after P or N intake [2]		✓			
[redacted]			✓	✓	✓
[redacted]		✓			
[redacted]			✓	✓	
[redacted]		✓			
Vital status				✓	✓
Study medication intake	Throughout the treatment period				
AE assessment	Throughout the study				
Concomitant medication assessment and documentation	Throughout the study				

d=days, ETD=early treatment discontinuation (i.e. before the EoST/EoSA visit), ePRO=electronical patient-reported outcome, EoST/EoSA=end of study treatment/end of study assessments, N=nintedanib, P=pirfenidone.
The preferred sequence of study assessments is described in the CSP.
[1] Only for subjects taking nintedanib at screening and randomization, until they stop taking nintedanib
[2] For subjects not taking nintedanib at screening and randomization and those stopping nintedanib during the study
[3] At these visits, the assessments from both the Every 12 weeks up to EoST/EoSA and the Every 24 weeks up to EoST/EoSA columns are included
* This can be a scheduled clinical study center visit or a phone call.

2.6. Sample Size Justification

See Section 7.1 of the clinical study protocols.

2.7. Randomization and Blinding

Details on randomization and blinding are in Section 4.6 of the study protocols.

An independent data monitoring committee (IDMC) will monitor the safety of the subjects in the study on a regular basis and will also assess the effect of treatment with GLPG1690 on lung function as assessed by rate of decline in FVC to evaluate the risk-benefit. In addition, an interim analysis to assess futility will therefore be performed when a reasonable number of subjects have completed 52 weeks of treatment (e.g. at least 25% subjects from the two studies combined) (see Section 7.3.2 of the CSP). The results will be reviewed by the IDMC, who will then make a recommendation to the sponsor on the progress of the study. Specific details on timing of the analyses, futility criteria, and statistical analyses to be performed will be detailed in the IDMC charter or SAP, as appropriate. To allow collection of as much safety information as possible on a potential down-titration dose, a futile dose will continue if the other dose does not show futility. Either the study will continue as planned or both doses (and the study) will be terminated. Sponsor personnel will remain blinded and the study will not be stopped for a beneficial effect. The IDMC and statistical support group reporting to the IDMC will be unblinded, but they are independent of the study team. Sponsor personnel will remain blinded.

After it was decided to terminate the trial, part of the study team was unblinded in order to be able to investigate the findings.

3. ANALYSIS POPULATIONS

The analysis population will always be indicated in a subtitle in the table, listing or figure.

3.1. All Screened Subjects Set

All enrolled subjects who underwent screening assessments to check whether or not they are eligible to participate in the clinical study.

3.2. All Randomized Set

All enrolled subjects who underwent all screening assessments and were found to be eligible for the clinical study and who were randomized into the clinical study.

3.3. Full Analysis Set (FAS)

All randomized subjects who received at least one dose of IP.

All data will be included irrespective of IP interruption or termination and changes in standard of care. Only efficacy data after lung transplant will be excluded.

If a subject is unblinded (e.g. in case of a serious adverse event [SAE]), all data of this subject recorded in the database will be included in the analyses.

3.4. *Per Protocol (PP) Set*

Per protocol analysis set is kept as documented in the original SAP for reference but will not be performed; nor will protocol deviations impacting the analysis be identified.

All randomized subjects who received at least one dose of IP, excluding subjects/data points with a major protocol violation which impacts the efficacy results. These protocol deviations are determined prior to unblinding and database lock for the final analysis and are entered in the database.

If a subject is unblinded (e.g. in case of an SAE), efficacy data will be disqualified from the per protocol analysis. Such data will still be included in the listings.

3.5. Pharmacokinetic Analysis Set

All randomized subjects who received at least one dose of IP and for whom evaluable PK data were available (e.g. excluding protocol violations/deviations or AEs that may have an impact on the PK analysis).

3.6. [redacted]

[redacted]

4. PRIMARY ESTIMAND

International Council for Harmonisation (ICH) draft guidance E9(R1) "Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials" specifies that a central question for drug development and licensing is to quantify treatment effects: how the outcome of treatment compares to what would have happened to the same subjects under different treatment conditions (e.g. had they not received the treatment or had they received a different treatment). Intercurrent events need to be considered in the description of a treatment effect on a variable of interest because both the value of the variable and the occurrence of the event may depend on treatment. The definition of a treatment effect, specified through an estimand, should consider whether values of the variable after an intercurrent event are relevant, as well as how to account for the (possibly treatment-related) occurrence or non-occurrence of the event itself. More formally, an estimand defines in detail what needs to be estimated to

address a specific scientific question of interest. A description of an estimand includes four attributes. The attributes in the CL-303 and CL-304 studies are underlined:

A. the population, that is, the patients targeted by the scientific question:
   Patients with IPF, defined through appropriate inclusion/exclusion criteria to reflect the targeted patient population for approval;
B. the variable (or endpoint), to be obtained for each patient, that is required to address the scientific question:
   Rate of decline in FVC to Week 52;
C. the specification of how to account for intercurrent events to reflect the scientific question of interest:
   The primary approach will include all data before lung transplant, irrespective of protocol violations, changes in SOC or IP interruptions or discontinuations.
   Sensitivity analyses are planned to investigate the impact of the intercurrent events;
D. the population-level summary for the variable which provides, as required, a basis for a comparison between treatment conditions:
   Difference in estimated slopes for each of the GLPG1690 doses versus placebo.

Together these attributes describe the estimand, defining the treatment effect of interest.

The design of the CL-303 and CL-304 studies targets this estimand as they are randomized parallel-group studies where every attempt is made to collect all measurements throughout the study.

5. TREATMENT GROUPS

5.1. Randomized Versus Actual Treatment

For efficacy, [redacted] and safety analyses, the treatment group as assigned by the randomization will be used in the analysis (i.e. as-randomized analysis).

For PK analyses, the actual treatment will be used in the analyses.

Differences between as-treated and as-randomized will be flagged in the listing of subject randomization.

5.2. Totals over Groups

A total over all groups will be presented for the General and Baseline Characteristics part of the analysis, but not for efficacy, safety, PK nor [redacted] analyses.

6. ANALYSIS PERIODS AND ANALYSIS TIME POINTS

6.1. Relative Number of Days

The timing of an assessment or an event relative to a reference date will be calculated as follows:

$$Relative\ day = concerned\ date - reference\ date + 1\ day,$$

$$= \begin{cases} \text{when concerned date} \geq \text{reference date} \\ concerned\ date - reference\ date, \\ \text{when the concerned date} < \text{reference date} \end{cases}$$

- The reference date is the date of the first dose of IP administration, unless specified otherwise.
- The concerned date could be the measurement date of the assessment, or the start or end date of the event.
- *Date* implies a complete date having day, month and year available. Unless otherwise specified, the *relative day* will remain missing when it cannot be calculated due to absence or incompleteness of the concerned and/or reference dates.

6.2. Analysis Periods for Non-visit Data

These analysis periods are to be used for allocation of events into periods (e.g. adverse events).

Table 1 Analyses Periods for Non-Visit Data

Analysis period	Start period	End period
Screening	Date of signing the ICF.	1 day before the date of first dose of IP
Treatment	First dose of IP date.	Last dose of IP date + 30 days.
Follow-up	Last dose of IP date + 31 days.	Study termination date. Note: if the study termination date is before the last dose of IP + 30 days no follow-up period will be created.

Note:

- If times are collected or other information with respect to timing (e.g. before or IP intake) this information will be taken into account in addition to the dates.

- The last analysis period in case of early termination will always end by the study termination date (date of last contact in the study or last known alive).

6.3. Algorithm of Allocating Visits to Time Analysis Windows

For the FAS analysis for efficacy, [redacted] and safety endpoints, all data (including data obtained at unscheduled visits) will be placed into time analysis windows according to their relative day in the study, using to the following allocation tables (Table 2 to Table 4).

Table 2 Analysis Windows for Efficacy and Safety Endpoints Recorded at All Time points (FAS and PP Approach)

Time point label	Target day	Interval lower bound	Interval upper bound
Baseline	1	-INF	1

Time point label	Target day	Interval lower bound	Interval upper bound
Week 2	15	2	21
Week 4	29	22	43
Week 8	57	44	71
Week 12	85	72	106
Week 18	127	107	154
Week 26	183	155	211
Week 34	239	212	267
Week 42	295	268	330
Week 52	365	331	407
Week 64	449	408	491
Week 76	533	492	575
Week 88	617	576	659
Week 100	701	660	743
Week 112	785	744	827
Week 124	869	828	911
Week 136	953	912	995
Etc.			

Table 3 Analysis Windows for DLCO, 6MWT, and Questionnaires (FAS Approach)

Time point label	Target day	Interval lower bound	Interval upper bound
Baseline	1	-INF	1
Week 26	183	2	274
Week 52	365	275	449
Week 76	533	450	617
Week 100	701	618	785
Week 124	869	786	953
Etc.			

Table 4 [redacted]

[redacted]			

Tables, figures and listings will present the analysis windows, not the site visits.

If times are collected or other information with respect to timing (e.g. before or IP intake) this information will be taken into account in addition to the dates.

Per parameter and postbaseline analysis window, the value closest to the target date will be used in analysis tables and figures. Other values will be listed only. If more than one value is located at the same distance from the target day, then the latest in time will be selected. If there are two values on the same day and no time indicating which one is last, the average of the two will be calculated.

In addition to the predose measurements, electrocardiogram (ECGs) and [redacted] are also taken 2-3 hours postdose on selected visits (Day 1, Week 2, 4, 12, 34 for ECG and [redacted]). These postdose measurements will not be taken into account in the above windowing, but tables will be created separately describing the pre- and postdose values at these selected visits.

6.4. Definition of Baseline

For all endpoints, except spirometry, [redacted] and ECG parameters, baseline is defined as the last non-missing value before the first IP administration.

For spirometry endpoints and [redacted], the average of the values before the first IP administration will be calculated to reduce variability. For spirometry, for visits 1 and 2 only results graded as "acceptable" can be used, for visit 3 results graded as both "acceptable" and "borderline acceptable" can be used.

For ECGs, the mean of the triplicate ECG values will be calculated for each individual ECG parameter. The last available mean of a triplicate before the first IP administration (i.e. excluding the postdose value on Day 1) has priority over single ECGs to become the baseline, even if these single ECGs have a later date and time than the triplicate. If no triplicate is available before the first dose, the last value(s) before start of IP will be used.

6.5. Definition of Last Contact and End of Study (Censoring Rules)

Since vital status information is to be collected also after subject's last visit in the study, the last contact has to be defined differently for endpoints including vital status information than for those not including vital status.

For endpoints including vital status:

The last contact, including vital status information is defined as the latest date collected in the electronic case report form (eCRF) from the visit dates, AE dates, concomitant treatment dates, drug intake dates or laboratory test dates, study completion date or last contact or last known alive date from the vital status form.

For endpoints not including vital status:

The last contact, not including mortality information is defined as the latest date collected in the eCRF from the date of death, visit dates, AE dates, concomitant treatment dates, drug intake dates or laboratory test dates, study completion date. Information from the vital status form is not to be taken into account for this calculation.

7. HANDLING OF DATA

7.1. Calculation of Descriptive Statistics

For continuous parameters, descriptive statistics will be presented when N≥2 per group. When N=1, the observation will not be shown in the table/figure but only in the listing(s).

Descriptive statistics will include at least the following:

- the number of non-missing data points (n)
- the arithmetic mean
- the standard deviation (SD) and standard error (SE) of the arithmetic mean
- the median, minimum and maximum
- 95% confidence interval (CI) of the mean (only when requested).

7.2. Calculation of Percentages

For event-type data (e.g. adverse events), the denominator for calculation of percentages will be all subjects in the analysis set and analysis period. For other data (e.g. worst-case analysis of assessments, demographic data), the denominator will be all subjects with (postbaseline) data for the parameter, in the analysis set and analysis window/period. Missing values will not be included in the denominator count.

7.3. Handling of Values Below (or Above) a Threshold

Except for [redacted], values below (above) the detection limit will be imputed by the value one unit smaller or larger than the detection limit itself. In listings, the original value will be presented.

Example: if the database contains the value “<0.04”, then for the descriptive statistics the value “0.03” will be used. The value “>1000” will be imputed by “1001”.

For [redacted] data, values below the detection limit will be imputed by zero.

7.4. Handling of Missing Data

7.4.1. Primary Approach

As the primary approach, there will be no imputation of missing values and missing values will be excluded from analyses and data is assumed to be missing at random (MAR). Missing data resulting from stopping the study is assumed to be missing completely at random (MCAR).

Missing data/visits due to the study conduct being affected by the COVID-19 pandemic (e.g. lockdown rules) can be considered MAR.

7.4.2. Handling of Missing Date Time Data

No imputations will be done in case of missing date (time) fields, nor for the missing parts of partially known date (time) fields.

Assessments with missing date (time) will be omitted from the analysis.

Event-type data (e.g. adverse events, concomitant medications) with missing date (time) will be allocated to analysis periods using a worst-case approach as explained in the respective sections.

7.4.3. Sensitivity Analyses

Since no treatment effect was observed and the study was stopped for lack of efficacy, sensitivity analyses to investigate the robustness of results will be limited. Description of sensitivity analyses are kept as documented in the original SAP for reference but most of them will not be performed (details in sections 9.1 and 9.3.2.2)

Several sensitivity analyses will be performed to investigate the impact of missing data, including analyses assuming missing not at random (MNAR), *of protocol deviations*, *early study termination and treatment interruptions or discontinuation as well as changes in standard of care.* Details are in the relevant efficacy sections below.

7.5. Handling of Outliers

All measured values will be included in the analyses. Extreme outliers will be evaluated before unblinding and if needed, additional sensitivity analyses will be added.

7.6. Stratification Factors

The analyses will be stratified by the type of standard of care taken at randomization: nintedanib, pirfenidone or neither. In all analyses, except if indicated otherwise, the stratum will be the SOC at randomization and will not change if the subject switches SOC during the study.

In case there is a discrepancy in the stratification reported in interactive web response system (IWRS) versus in the eCRF, then the stratum reported in the eCRF will be used in analyses, and the discrepancy will be flagged in the appropriate listing.

7.7. Exclusion of Data in Case of Quality Issues

In case of serious breach or other site-related quality concerns the study team in collaboration with the quality department will assess the impact on the data and may decide to exclude the data from analyses, this will be documented in the eTMF.

8. INTERIM ANALYSES AND DATA MONITORING COMMITTEE REVIEW

To protect the safety of the subjects in the study and integrity of the study data, an IDMC will be implemented.

In parallel to assessing potential safety risks on a regular basis, the IDMC will also examine the effect of treatment with GLPG1690 on lung function as assessed by rate of decline in FVC using the data from the identically designed studies GLPG1690-CL-303 and GLPG1690-CL-304. An interim analysis to assess futility will be performed when a reasonable number of subjects have completed 52 weeks of treatment (at least 25% subjects completing Week 52 or 70% information available, whichever comes latest, the exact timing will be determined based on the actual recruitment). The results will be reviewed by the IDMC, who will then make a recommendation to the sponsor on the progress of the study. The Sponsor will remain blinded to treatment allocations and results of the analyses. To allow collection of as much safety information as possible on a potential down-titration dose, a futile dose will continue if the other dose does not show futility. Either the study will continue as planned or both doses (and the study) will be terminated. Specific details on timing of the analyses, futility criteria, and statistical analyses to be performed will be detailed in the IDMC charter and SAP, as appropriate. A summary of the futility approach is also provided in Appendix 5.

Given there is no plan to terminate the study early for overwhelming efficacy results, no adjustment of the alpha level is considered needed.

No other interim analyses are planned for these studies.

9. STATISTICAL ANALYSES

9.1. Changes to the Planned Analyses, Not Covered by Protocol Amendments

In view of the termination of the ISABELA studies and the decision to write an abbreviated CSR, it was decided to limit the analyses to the primary, key secondary and safety endpoints and any analyses important to better understand the reason for the results seen, or relevant for scientific knowledge for future compounds. Since no treatment effect was observed and the study was stopped for lack of efficacy, sensitivity analyses to investigate the robustness of results will also be limited.

A detailed list of the analyses that will not be performed can be found in the table below. In the remainder of the document the descriptions as documented in the original SAP are kept for reference; but analyses that will not be performed are in *Italic font*.

Section SAP	Analyses to be omitted
9.2.1	- Baseline EQ_5D - Baseline LCQ total score and domains - Baseline K-BILD total score and domains - Baseline VAS Cough and Urge to Cough - [redacted] - Baseline SpO2 during 6MWT (%) - Frequency tabulation baseline Physical examinations
9.2.2	- The number (percent) of subjects in the different analysis populations - The number (percent) of subjects per analysis window as defined in Section 6.3 for analysis. - Descriptive statistics of the duration of the analysis periods - Number of subjects with a deviation of more than the number of days allowed per protocol from the target of the analysis windows, per window. - Descriptive statistics of the number of days distance from the target of the window, as specified in the protocol, as well as of the relative day per window. These will allow to assess if the later visits (especially the close-out visits) are done close to scheduled visits.
9.2.3	All major protocol deviations will be tabulated, not only the ones impacting efficacy, as this assessment will not be done. A list of all major protocol deviations will be provided without the flag for impacting efficacy.
9.2.5.3	Analyses restricted to medications taken up to 30 days after the last dose of IP.
9.2.6	Analyses on exposure and compliance parameters on data up to 52 weeks
9.2.7	KM analyses for time to first switch in standard of care
9.3.1	Pooled analyses will not be performed.
9.3.2.2	9.3.2.2.2 Per protocol analysis 9.3.2.2.3 On-treatment 9.3.2.2.4 After Lung Transplant 9.3.2.2.6 Completers vs Non-completers 9.3.2.2.7 Impact of missing data – Pattern mixture 9.3.2.2.8 Impact of SOC
9.3.3	-On-treatment analysis -Negative binomial model
9.3.5	-Per protocol analysis -Random coefficient regression model

9.3.6	9.3.6.1 EQ_5D 9.3.6.2 LCQ 9.3.6.3 K-BILD 9.3.6.3 VAS Cough and Urge to Cough
9.3.7	Functional Exercise Capacity and [redacted]: all analyses omitted, except for total distance walked in 6 minutes, this parameter will be analyzed.
9.3.9	[redacted] - No subject profiles will be created.
9.3.10	No analyses per age, sex, race and [redacted]. Analyses per stratum and per geographical region will be done.
9.3.11	[redacted]
9.4.5	Physical examinations frequency table
9.4.3.6	- Tabulation of ECG interpretations - Table summarizing the pre- and postdose values at these selected visits.

Tables and listings to investigate the impact of COVID-19 on the study have been added.

If changes are needed after unblinding, these will be documented in the CSR with an explanation for the reason for the change.

9.2. Subject Information

Subject information will be tabulated using the FAS population. No inferential testing will be performed nor will p-values be provided.

9.2.1. Demographic and Baseline Disease Characteristics

Demographic and baseline disease characteristics will be tabulated with descriptive statistics per treatment group and for all subjects pooled. The same summaries will also be provided by stratum. The following parameters will be summarized:

- Stratum: standard of care (pirfenidone, nintedanib or neither)
- Sex
- Age at the moment of signing the ICF (years) as recorded in the database
- Age categories:
  - ≥18 - <65 years
  - ≥65 -< 85 years
  - ≥ 85 years
- Race

- Ethnicity
- Height at baseline (cm)
- Weight at baseline (kg)
- Body mass index (BMI) at baseline (kg/m²) = $\frac{weight\ (kg)}{height^2\ (m^2)}$
- BMI at baseline, categorized:
  - ≤ 18.5 kg/m²
  - >18.5- ≤ 25.0 kg/m²
  - >25.0- ≤ 30.0 kg/m²
  - > 30.0 kg/m²
- Systolic and diastolic blood pressure (mmHg), heart rate (beats/min) , respiratory rate (breaths/min) and oxygen saturation (%) at baseline
- Alcohol status: never, current, former
- Smoking status: never, current, former
- Duration of IPF (years) = (date of first intake of IP - date of initial diagnosis)/365.25. If the date of initial diagnosis is incomplete, then the following rules will be applied: Missing day: use the first day of the month. Missing month: use January.
- Duration of IPF, categorized as:
  - <0.5 year
  - ≥0.5- <1 year
  - ≥1- <2 year
  - ≥ 2 year
- Baseline spirometry values:
  - forced vital capacity (FVC) (mL),
  - percent predicted FVC (%),
  - forced expiratory volume in 1 second (FEV_1) (mL),
  - percent predicted FEV_1 (%),
  - FEV_1/FVC ratio,
  - forced expiratory flow between 25% and 75% of exhaled volume (FEF_{25-75}) (mL/s).
- Other baseline efficacy measures:
  - SGRQ total score and domains:
    - symptoms (assessing the frequency and severity of respiratory symptoms),
    - activity (assessing the effects of breathlessness on mobility and physical activity),
    - impact
  - *EQ-5D,*
  - *LCQ total score and domains,*
  - *K-BILD total score and domains,*
  - *VAS Cough and Urge to Cough,*
  - 6MWT distance,

- [redacted]
- *SpO2 during 6MWT (%)*
- DLCO (corrected for Hb)

A frequency tabulation per system organ class and treatment group (and overall) of the baseline physical exam results, categorized as normal/abnormal will also be provided.

9.2.2. Disposition Information

The following tabulations will be provided by treatment group and overall:

- The number of subjects screened, randomized and not-randomized, with the reason for not being randomized. Number randomized and not treated.
- Number of subjects randomized per country and investigator.
- *The number (percent) of subjects in the different analysis populations as defined in Section 3.*
- *Descriptive statistics of the duration of the analysis periods will be provided.*
- *The number (percent) of subjects per analysis window as defined in Section 6.3 for analysis.*
- *Number of subjects with a deviation of more than the number of days allowed per protocol from the target of the analysis windows, per window.*
- *In addition, descriptive statistics will be provided of the number of days distance from the target of the window, as specified in the protocol, as well as of the relative day per window. These will allow to assess if the later visits (especially the close-out visits) are done close to scheduled visits.*
- Overall summary of number of subjects (percent) per number of missed and virtual visits due to COVID-19 per treatment group
- Summary of number of subjects (percent) with missed, virtual visits due to COVID-19, missed visits not due to COVID-19 and complete visits per study visit and treatment group
- The number (percent) of subjects with completion/discontinuations and the reasons for discontinuation from study and from IP. This table will also be provided by stratum.
- The number (percent) of subjects with discontinuations due to COVID-19 and the reasons for discontinuation from study and from IP.
- Kaplan-Meier estimates (graphical and tabular) of the time in the study, defined as time from the date of the first intake of IP to the last contact in the study, excluding data from the vital status eCRF. See Section 6.5 for the censoring rules.
- Kaplan-Meier estimates (graphical and tabular) of the time on treatment, defined as time from the date of the first to last intake of IP.

If a high number of discontinuations is seen in a blinded fashion during the course of the study, more descriptions may be added to further investigate the reasons for discontinuations over time. Details will then be added to this SAP before unblinding for the final analysis.

In addition, the following listings will also be provided:

- Subjects receiving investigational product from specific batches, where more than one batch was used (showing the batch numbers of the investigational products actually used, including the expiration dates),
- Randomization scheme and codes (subject identification, treatment assigned, with a flag if the code was broken and the reason).
- Date of the earliest ICF signed for this study and date of last scheduled visit performed in this study.
- Listing of missed and virtual visits due to COVID-19.
- Listing of Study and/or study drug discontinuations due to COVID-19
- Listing of comments containing information related to COVID-19

9.2.3. Protocol Deviations and Eligibility

Listings will be provided for the subjects excluded from the different analyses sets with the reason for exclusion.

Tabulation per treatment group (and overall) of the major protocol deviations *impacting the efficacy analyses* will be provided. The same table will also be provided by stratum. A listing will be provided for all major protocol deviations *with a flag indicating which impact efficacy analyses and are thus excluded from the PP analysis*.

Tabulation per treatment group (and overall) of the number (percent) of subjects with major protocol deviations related to COVID-19, overall and per COVID-19 related protocol deviation will be provided. Listings of major and minor COVID-19 related protocol deviations will be created.

9.2.4. Medical History and Concomitant Diseases

Frequency tabulations per treatment group (and overall) per system organ classes and preferred terms will be provided for the medical history findings (i.e. condition no longer present at the start of the study) as well as for the concurrent diseases (i.e. conditions present at the start of the study).

9.2.5. Prior and Concomitant Therapies

9.2.5.1. Classification of Therapies

All prior and concomitant therapy records will be categorized as follows, considering their date and flags indicating the relative timing versus study (drug) start or end (before, after, ongoing…):

- Prior only: when the record ended before the first IP administration date.
- Concomitant only: when the record started on the same date or after the first IP administration date.

– Prior and concomitant: when the record started before the date of first IP administration, and ended on a date after first IP administration (or on the same date), or continued.

Records without a start date are assumed to have started before the date of first IP administration. Records without an end date are assumed to be ongoing.

When the start or end date of the prior and concomitant therapy records are incomplete, the date of first IP administration will be considered to the same level of information provided by these incomplete dates in order to categorize the timing of these records. This means a record only having month and year will be categorized comparing only to the month and the year of the date of first IP administration. If the start date is unknown but the medication is flagged to have started prior to the first dose of IP, this information will be taken into account.

For both the start and the end dates of the concomitant therapy records, their day relative to the day of first IP administration will be calculated as described in Section 6.1.

9.2.5.2. Coding of Therapies

All therapies are coded using WHO Drug coding. In the table(s), the generic term will be used. The anatomical therapeutic classification (ATC) classes level 2 and 4 will be used for analysis. Multiple records of the same generic term for the same subject with the same categorization will be counted only once. The table will therefore present subjects, not occurrences.

9.2.5.3. Analyses

A frequency tabulation per treatment group (and overall) of the ATC classes level 2 and 4 and generic terms of the prior medications (defined as ‘prior only’ or ‘prior and concomitant’) will be provided and as well as of the concomitant medications (defined as ‘concomitant only’ or ‘prior and concomitant’).

Additional tables will be created for respiratory-related concomitant and prior medications. The definition of respiratory-related medications will be provided in a separate document before unblinding for the final study report.

Similar tables will be provided restricted to medications taken up to 30 days after the last dose of IP.

9.2.6. Exposure to IP and Compliance

The following parameters will be calculated for the compliance and extent of exposure to IP:

– Overall compliance (%) = 100 x (number of tablets actually used) / (number of tablets that should have been used, according to the randomization).
– Total treatment duration (days) = last dose of IP intake date – first dose of IP intake date + 1 day.
– Total treatment duration, excluding days off drug (days): sum of all durations (last – first +1) in the drug log pages where there is at least one dose taken, i.e. the sum of the number of days with at least one dose of IP.

- Percentage days with an intake = 100% x $\left(\frac{\text{total treatment duration,excluding days off drug}}{\text{total treatment duration}}\right)$
- Total compliant treatment duration (days): sum of all durations (last – first +1) in the drug log pages where there is a dose with exactly 3 tablets.
- Percentage compliance of treatment duration = 100% x $\left(\frac{\text{total compliant treatment duration}}{\text{total treatment duration}}\right)$)

Similar calculations will be done using only the data up to Week 52.

Summary tables will be provided for the different compliance and extent of exposure parameters by treatment and overall. In addition, frequency tables will also be provided for the compliance parameters (overall compliance, percentages days with an intake and percentage compliance of treatment duration), using the following categories: <80%, ≥80%-<100%, 100%, >100%-≤120%, >120%.

Frequency tables with the number of subjects who had at least one down-titration and with the number of subjects who had at least one dose interruption, and the reasons for down-titration and interruption by treatment group and overall, will also be provided.

Patient-years of exposure to IP by treatment group and overall.

The same tables will be created by stratum and treatment.

A listing will be provided of the subjects who took incorrect IP for at least part of the treatment period, all prime therapy records are to be included to provide an overview of the amount of incorrect treatment relative to the total treatment period.

In addition, listings of the subjects who had their IP down-titrated and interrupted for at least part of the treatment period will be provided, all prime therapy records are to be included to provide an overview of the amount of down-titration and interruption relative to the total treatment period.

9.2.7. Exposure to Standard of Care (SOC)

The following parameters will be derived per type of SOC (nintedanib, pirfenidone and neither):

- Duration of SOC= last dose of SOC – first dose of SOC + 1 day

- Compliance to SOC (%) = 100 x (number of SOC tablets actually used) / (number of SOC tablets that should have been used in the study).
  Where the number of SOC tablets is to be calculated based on the total study duration (from first dose of IP to last contact).

Summary tables by treatment, type of SOC will be provided for these two parameters.

A frequency table per treatment group, summarizing the SOC intake during the study, of the number and percentage of subjects with the following:

- Baseline SOC pirfenidone:
  - Stable dose until study end date
  - Reduced dose at least once
  - Temporarily stopped at least once
  - Permanently stopped and started nintedanib
  - Permanently stopped (without starting nintedanib)
- Baseline SOC nintedanib:
  - Stable dose until study end date
  - Reduced dose at least once
  - Temporarily stopped at least once
  - Permanently stopped and started pirfenidone
  - Permanently stopped (without starting pirfenidone)
- Baseline SOC neither:
  - Stable until study end date
  - Started nintedanib or pirfenidone and permanently stopped (without restarting other SOC)
  - Started nintedanib or pirfenidone and temporarily stopped (i.e. period without any SOC) at least once
  - Started nintedanib or pirfenidone and continued until study end date (includes subjects that switch SOC without temporary stop)

Note: these categories should be counted in a worst-case manner i.e. a subject appears only once in one of the categories (in the order as specified above with “permanent stop” or “starting SOC” as the worst cases).

Additionally, a frequency table of the type of switch with the reasons for switching SOC will be provided per treatment group.

A frequency table describing the doses taken of Pirfenidone and Nintedanib per treatment group will also be created.

Patient-years exposure to each SOC by treatment group and stratum and overall will be provided.

A frequency table, summarizing SOC intake prior (e.g. Currently taking Pirfenidone; Never took Nintedanib, …) to the study by treatment group and stratum and overall will be created.

In addition, a listing will be provided of the subjects who had any change in SOC medication (pirfenidone, nintedanib or neither) during the treatment period compared to baseline e.g. switching between nintedanib and pirfenidone, starting pirfenidone/nintedanib, stopping pirfenidone/nintedanib, down-titration or up-titration of pirfenidone/nintedanib. All SOC therapy records are to be included.

In addition, Kaplan-Meier tables and graphs will be created for the time from first dose of IP to switch of SOC, where a switch in SOC is defined as:

- *switching from nintedanib to pirfenidone or vice versa for those in the 'nintedanib' or 'pirfenidone' strata,*

- *starting nintedanib or pirfenidone, for those in the 'neither' stratum.*

Furthermore, for those in the 'pirfenidone' or 'nintedanib' stratum, a Kaplan-Meier curve and table will be provided for the time to discontinuation of the SOC they started on.

For the neither stratum, a tabulation of the number of subjects (%) by reason for not taking Pirfenidone and Nintedanib.

A listing of COVID-19 related SOC intake changes will be provided.

Heterogeneity between countries (and/or regions) will be investigated.

If a high number of subjects switching SOC are observed in a blinded fashion during the course of the study, more descriptions may be added to further investigate the patterns over time. Details will then be added to this SAP before unblinding for the final analysis.

9.3. Efficacy Analyses

This section provides an overview of the efficacy analyses performed for this study.

9.3.1. Level of Significance

Given no market authoristation applicaton for ziritaxestat for the treatment of IPF will be submitted anymore, no pooled analyses combining data from both studies will be performed and the below multiplicity adjustment strategy for countries outside US, does not apply. Only abbreviated CSRs per study will be written as explained earlier in this document.

Statistical tests will be done at a 2-sided level. For the analysis of the primary endpoint, to account for the multiple testing due to two doses being compared to placebo, a Bonferroni approach (Hochberg, 1988) will be used with higher priority given to the high-dose group. The primary endpoint will be tested at a 4% level when comparing GLPG1690 600 mg to placebo and a 1% level for GLPG1690 200 mg group versus placebo.

With respect to the multiplicity adjustment for the key secondary endpoints an additional approach was requested by the [redacted] compared to European Union (EU) countries. For countries outside US and EU, the EU approach will be applied.

For [redacted]

The [redacted] requested a multiplicity adjustment for the key secondary endpoints within each study. A closed-testing hierarchical approach will therefore be used for each dose group separately. The

key secondary endpoints will be ordered as listed in Section 2.2.2.1 and they will be tested at a 4% level when comparing GLPG1690 600 mg to placebo and a 1% level for GLPG1690 200 mg group versus placebo. In this approach, the first key secondary endpoint in a dose group will only be tested if the primary endpoint for that dose group showed significance. The next one will only be tested if the previous one showed significance. These analyses will be presented in the study CSR. Figure 2 details the testing order within each of the identically designed studies.

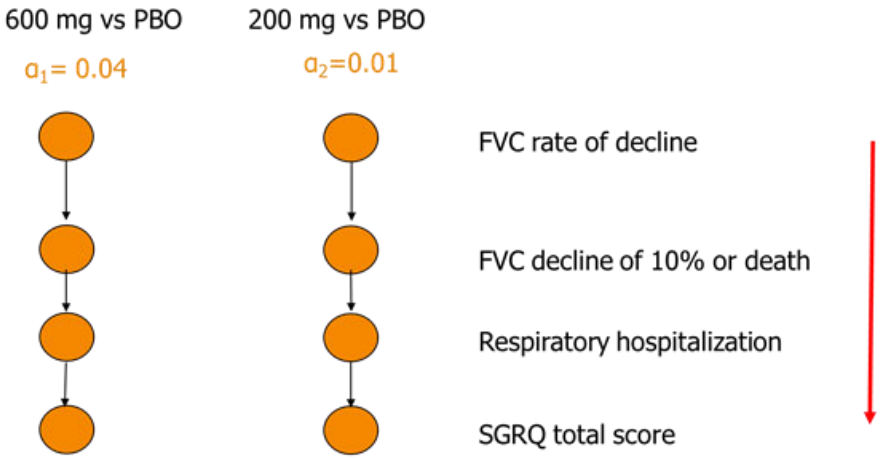

Figure 2 Multiplicity adjustment within study for [redacted] approach

For Countries outside the US:

To increase the probability to detect a potential treatment effect on the key secondary endpoints, some of which are rare but clinically important events, they will be analyzed using the pooled data from the two studies (GLPG1690-CL-303 and GLPG1690-CL-304). The confirmatory testing on these secondary endpoints will be done only if the primary endpoints in both studies show significance and a step-down closed-testing multiplicity adjustment will be applied, as shown in Figure 3. A key secondary endpoint in the pooled analysis will only be tested for a dose if the higher order endpoint also showed significant for that dose. This analysis will be presented in the Section 2.7.3 Summary of Clinical Efficacy (and Module 5 Integrated Summary of Efficacy). Details of the pooled analyses will be provided in the pooled-analysis SAP.

Results for the key secondary endpoints for each individual study will be summarized within its CSR and will enable the evaluation of consistency of the results between the studies. An overview of the multiplicity adjustment for the primary and key secondary endpoints is provided in Figure 3.

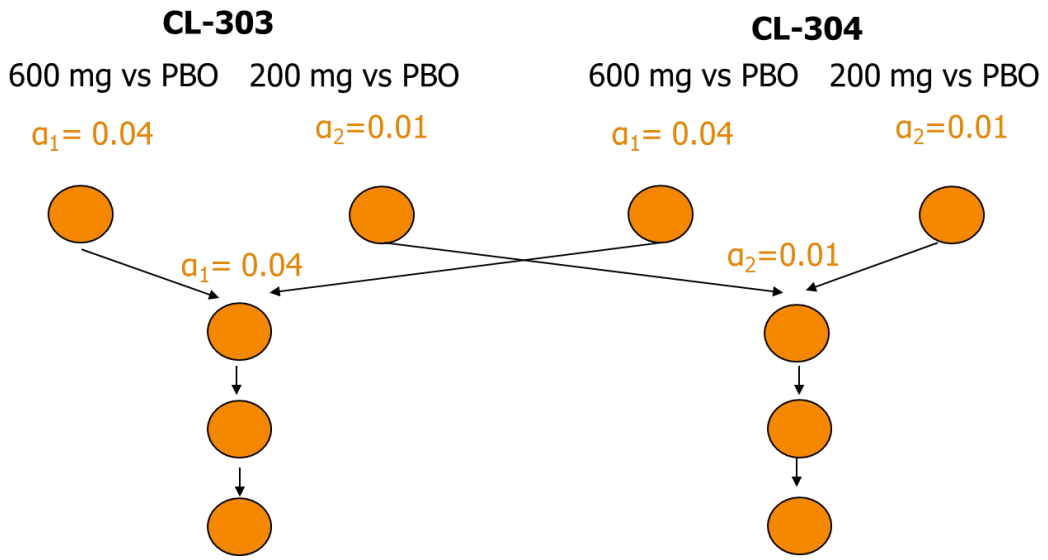

Figure 3 Multiplicity adjustment within study for Countries outside the US

9.3.2. Spirometry Parameters

All spirometry evaluations should be performed pre-bronchodilator. Spirometry data taken post-bronchodilator will be excluded from the analysis.

The clinical study center-based spirometry must meet the criteria for acceptability and repeatability as defined in the American Thoracic Society / European Respiratory Society / Japanese Respiratory Society / Latin American Thoracic Association (ATS/ERS/JRS/ALAT) guidelines (Miller, et al., 2005) for screening, and will be based on ATS/ERS/JRS/ALAT guidelines after randomization.

Pulmonary function will be measured in a standardized manner and results should be transmitted electronically during the visit immediately after performing the spirometry and evaluated by a central reader. In case the acceptability and repeatability criteria as specified by ATS/ERS/JRS/ALAT guidelines are not met, a repeat spirometry should be performed during the same visit of the screening period. After randomization, a repeat spirometry should be performed during the same visit, based on the ATS/ERS/JRS/ALAT guidelines.

The data will be coded as “acceptable”, “borderline acceptable” and “unacceptable”, “unacceptable” values will be excluded from analyses.

The following parameters will be measured or calculated as part of the spirometry assessment:

- FVC (mL)
- Percent predicted FVC (%), as available in the database
- FEV_1 (L) and percent predicted FEV_1(%)
- FEV_1/FVC ratio
- FEF_{25-75} (mL/s).

The '2012 Global Lung Function Initiative Equations' will be used to calculate the predicted values (Ratitch B, 2013) .

The spirometry endpoints will be analyzed using the FAS, including all available data before lung transplant for subjects who took at least one dose of IP, and excluding values post-bronchodilator.

For each of the spirometry parameters, descriptive statistics will be provided for the values, changes and percent change from baseline per time point up to the end of the study and per treatment group. Results will also be graphically displayed by means of line plots of the mean (SE) over time-by-treatment group.

A similar summary table and graph will also be provided by stratification factor (standard of care) and reason for not being on SOC.

9.3.2.1. Primary Endpoint: Rate of Decline in FVC over a Period of 52 Weeks

The primary endpoint of annual rate of decline in FVC at Week 52 (in mL) will be analyzed, including data up to and including the 52-week analysis window defined in Section 6.3. A random coefficient regression model (linear slope model) will be used including sex, age, height and stratification factor as covariates and a random intercept and slope. The treatment effect is determined by using estimated slopes for each treatment group on the basis of the time-by-treatment interaction term from the mixed model. All available FVC values (as defined by the analysis windows) from baseline to Week 52 will be used including FVC measurements at the follow-up visit for subjects who discontinue IP prematurely and do not complete the study visits through Week 52. The model will be fit using the SAS PROC MIXED procedure, details are in Appendix 1.

The assumptions of the model will be investigated.

The same model will be used for the pooled analysis for the integrated summary of efficacy adjusted for study. More detail will be in a separate SAP.

A similar analysis will also be provided by stratification factor (standard of care).

The percent predicted FVC will be analyzed using a similar random coefficient regression model (linear slope model) as FVC expressed in mL.

9.3.2.2. Sensitivity Analyses to the Primary Analyses

Since no treatment effect was observed and the study was stopped for lack of efficacy, sensitivity analyses to investigate the robustness of results will be limited. Description of sensitivity analyses are kept as documented in the original SAP for reference, analyses that will not be performed are in italic font.

Several sensitivity analyses are planned to investigate the impact of protocol violators, missing data, treatment discontinuation, and changes in standard of care. If during the blinded review of

the data more sensitivity analyses are deemed necessary these will be added to this SAP before unblinding for the final analysis of the CSR.

For the below sensitivity analyses and per dose group (600 mg, 200 mg) the estimate, 95% CI of the treatment effect compared to placebo will be presented on the same forest plot together with those of the primary analysis.

9.3.2.2.1. Sensitivity Analysis not Assuming a Linear Trend over Time

A mixed-effects model for repeated measures (MMRM) with treatment, time (as a categorical factor defined by the analysis windows in Section 6.3), treatment*time, stratum and baseline value as factors in the model will be applied to the change from baseline in FVC value in mL. Least squares (LS) means (with 95% CI) will be estimated from the model at each time point for each treatment group as well as the differences from placebo in LS means (and 95% CIs). Details on the model are in Appendix 3.

This approach provides another estimate (LS mean difference) for the primary estimand.

9.3.2.2.2. Sensitivity Analyses – Per Protocol to Investigate the Impact of Protocol Violators

To investigate the impact of non-compliance and protocol violators a Per Protocol analysis will also be performed for the primary endpoint of decline in FVC using the random coefficient regression model. Protocol violations potentially impacting the results will be excluded and identified before database lock and unblinding for the final analysis.

This approach provides another estimand than the primary one, namely the estimate in the population of subjects who take treatment according to protocol.

9.3.2.2.3. Sensitivity On-treatment Analysis

The same random coefficient regression model will be run for the rate of decline in FVC, using only the data up to the last intake of IP. Data after lung transplant and taken post-bronchodilator will be excluded.

This approach provides another estimand than the primary one, namely the estimate in the population of subjects who are on treatment.

9.3.2.2.4. Sensitivity Analysis Including Data After Lung Transplant

The same random coefficient regression model will be run including all data, also after lung transplant.

9.3.2.2.5. Sensitivity Analysis on All Study Data (also beyond Week 52)

The same random coefficient regression model will be run using all data up to the end of the study for FVC in mL and percent predicted FVC. Data after lung transplant and taken post-bronchodilator will be excluded.

This approach provides another estimate of the primary estimand.

9.3.2.2.6. Sensitivity Analysis for Completers versus those Who Discontinued Early

The same random coefficient regression model will be run up to Week 52, for those who completed the study up to Week 52 separately from those who discontinued the study before Week 52.

9.3.2.2.7. Sensitivity Analyses to Investigate the Impact of Missing Data – Control-Based Pattern Mixture Approach

Data after lung transplant and taken post-bronchodilator will be excluded.

Non-monotone missing data and missing data at visits before Week 52 will not be imputed. This assumes that the number of non-monotone missing data is either negligible or MAR, which is probably a reasonable assumption.

Multiple imputations will be used to handle missing data at Week 52. The imputation model is similar to the random coefficient regression model of the primary analysis.

In the pattern mixture approach, patterns of missing data are defined by subjects being classified in the following five categories:

1. *Those with Week 52 data who received treatment till Week 52*
2. *Those with Week 52 data who discontinued treatment before Week 52*
3. *Those without Week 52 data who discontinued treatment before Week 52 and are still alive*
4. *Those without Week 52 data and continued treatment till Week 52 or later*
5. *Those who died before Week 52.*

The number of imputations will be set to 1000 in order to ensure adequate efficiency for the estimation of the missing data. For each imputed dataset, the same model will be run as for the primary analysis and results will be combined using the standard multiple imputations approach of (Rubin, 1987)*. Imputations of the missing Week 52 value will be based on the estimates of the slope (SE), using data from prespecified groups depending on the assumptions. The assumptions behind are that the rate of decline in FVC in the subset of subjects imputed is similar to the rate of decline in the subset used for the imputation. A frequency table will be provided of the number (%) of subjects in each pattern.*

Given the groups may become small, the strata with SOC (either nintedanib or pirfenidone) will be combined and the combined stratum with 2 levels (nintedanib or pirfenidone vs neither) will be used. If the groups are still very small and less than 10% of subjects discontinue this approach will not be performed and will only be performed on the pooled data.

Appendix 2 provides more information on the multiple imputations approach.

Pattern Mixture Approach 1:

Pattern 3 will be imputed based on the data from Pattern 2 for the same treatment group and combined stratum (with 2 levels nintedanib or pirfenidone vs neither).

Pattern 4 will be imputed based on data from Pattern 1 for the same treatment group and combined stratum

Pattern 5 will be imputed based on data from the placebo groups of Pattern 2 for the same combined stratum.

This approach assumes that subjects who discontinue and are alive have the same rate of decline as the other subjects in the same treatment group and combined stratum, while subjects who died have a rate of decline similar to the discontinued placebo subjects in the same stratum after their last available FVC value.

Pattern Mixture Approach 2:

Patterns 3, 4, and 5 will be imputed based on the data from Pattern 1, 2 placebo subjects in the same combined stratum. This is a conservative approach which assumes that subjects who discontinued IP have a similar rate of decline in FVC as the placebo subjects after discontinuation.

Note:

The missing data will be investigated during the study in a blinded manner, especially if the monotone missingness assumption is reasonable, especially in the context of the bronchodilators use. If a high percentage of data is missing in a non-monotone pattern, appropriate measures will be taken (e.g. using an Markov chain Monte Carlo imputation of this data).

9.3.2.2.8. Sensitivity Analyses to Investigate the Impact of Changes in Standard of Care (Stratum)

Data after lung transplant and taken post-bronchodilator will be excluded. The below sensitivity analyses will be performed for the primary endpoint of annual rate of decline in FVC using the random coefficient regression model.

Sensitivity Analysis 1:

Subjects who were on the same standard of care less than 75% of the time will be excluded, i.e. in the 'neither' stratum, subjects will be excluded if they took either nintedanib or pirfenidone for more than 25% of the time, and in the 'nintedanib' and 'pirfenidone' strata they will be excluded if they did not take the SOC from their stratum for at least 75% of the time up to Week 52. Results will be presented overall and by stratum.

Sensitivity Analysis 2:

A similar approach as sensitivity analysis 1 will be performed combining the 'nintedanib' and 'pirfenidone' strata. In other words, if a patient starts on nintedanib treatment and

switches to pirfenidone or vice versa, they will be considered as still being on SOC for the calculation of the 75% of time on SOC.

The pattern of changes in SOC will be evaluated during the course of the study in a blinded (to randomization) manner and additional sensitivity analyses may be added as deemed necessary.

9.3.2.3. First Key Secondary Endpoint – Disease Progression

Disease progression is defined as the composite endpoint of first occurrence of ≥10% absolute decline in percent predicted forced vital capacity (%FVC) or all-cause mortality at 52 weeks. The number (%) of subjects who experience disease progression up to Week 52 will be presented and analyzed using logistic regression with similar factors in the model as for FVC rate of decline. The odds ratios (with 95% CI) versus placebo will be estimated from the model. This analysis corresponds to the key secondary endpoint and will be used in the multiplicity approach.

A similar summary analysis will also be provided by stratification factor (standard of care).

9.3.2.4. Other Spirometry Endpoints

The same will be used to describe the disease progression up to the end of the study.

Additionally, the two components of this composite endpoint of disease progression will also be summarized separately (number of subjects who died and number of subjects have either an absolute decline of 10 percentage point or more in percent predicted FVC).

Furthermore, to account for the differential periods of observation in the study a time to event analysis similar as described below for time to respiratory hospitalization, will be used to analyze the composite endpoint of percent predicted FVC decline and mortality up to the end of the study. See below for details on analyses of time to event endpoints.

In addition, a cumulative incidence table and graph will be presented for absolute change from baseline in percent predicted FVC per time point and worst-case up to the end of the study for the categories ≤-10, ≤-9, ≤-8,…, ≤0, …, ≤8, ≤9, ≤10 , by treatment group.

9.3.3. Second Key Secondary Endpoint - Time to First Respiratory-related Hospitalization until the end of the study

Respiratory hospitalizations will be adjudicated by the CEAC. The analysis will be based on the results from the adjudication, not on the investigator-reported information. Differences will be investigated and listed, if any.

The number (%) of subjects with an event during the study will be summarized as well as the number of events per 100 patients-years of follow-up. In addition, time to first event will be graphically displayed and tabulated (at each planned time point as described in Section 6.3, Week 2, Week 4, etc, and last available time point) using Kaplan-Meier estimates.

A Cox proportional hazards model with terms for age, sex, height, and stratum will be used to estimate the hazard ratios (and 95% CI) for each dose compared to placebo and to compare the active doses to placebo.

Censoring

Time to event endpoints are defined from the date of first dose of IP.

If the subject did not have an event, the time to event will be censored at the last contact date, as defined in Section 6.5.

Sensitivity Analyses

An on-treatment approach will be performed for the key secondary endpoint of time to first respiratory hospitalization.

Additional Analysis

The number of subjects with 0, 1, 2, 3, >3 respiratory-related hospitalizations will be presented by treatment group.

Since it is expected that a relatively high number of subjects will have more than one hospitalization, the rate of hospitalizations will be compared between each of the dose groups and placebo using a negative binomial model (generalized linear model) as explained in Keene et al. (2007).

9.3.3.1. Other Adjudicated Events (Secondary) Endpoints

A time to (first) event analysis will be conducted for each of the following secondary endpoints:

- Mortality: all-cause and respiratory-related
- Hospitalization: all-cause
- Hospitalization for non-elective Lung transplant
- Acute IPF exacerbation
- All-cause mortality or hospitalization for non-elective lung transplant
- All-cause mortality, or hospitalization for non-elective lung transplant, or hospitalization for qualifying for lung transplant
- All-cause mortality, hospitalization for ≥10% absolute decline in %FVC, or respiratory-related hospitalizations
- All-cause mortality or respiratory-related hospitalizations

Results will be based on the results coming from adjudication. Time to first event will be presented for each type of adjudicated event, similarly as time to first respiratory-related hospitalizations, using Kaplan-Meier tables, graphs and Cox PH models based.

In addition, for all-cause hospitalization and all-cause mortality a frequency table will be provided for the primary cause of death/hospitalization as defined by the CEAC.

Respiratory related events analyses will also be provided by stratification factor (standard of care).

In addition to the analysis of all-cause mortality based on adjudication, a time to all-cause mortality analysis will also be performed on all-cause mortality as reported by the investigator.

Analyses to assess the impact of COVID-19 cases (infection and suspected) on mortality and disease progression will be provided.

9.3.4. Other Time-to-Event Endpoints

The number (%) of subjects with an event during the study will be summarized as well as the number of events per 100 patients-years of follow-up. In addition, time to first event will be graphically displayed and tabulated (at each planned time point as described in Section 6.3, Week 2, Week 4, etc, and last available time point) using Kaplan-Meier estimates.

A Cox proportional hazards model with terms for age, sex, height, and stratum will be used to estimate the hazard ratios (and 95% CI) for each dose compared to placebo.

Censoring

Time to event endpoints are defined from the date of first dose of IP.

If the subject did not have an event, the time to event will be censored at the last contact date, as defined in Section 6.4.

9.3.5. Third Key Secondary Endpoint: SGRQ Scores

This questionnaire will be completed electronically, using a tablet device at the clinical study center. The SGRQ is a 50 item questionnaire split into three domains:

- symptoms (assessing the frequency and severity of respiratory symptoms),
- activity (assessing the effects of breathlessness on mobility and physical activity),
- impact (assessing the psychosocial impact of the disease).

Scores are weighted such that every domain score and the total score range from 0 to 100, with higher scores indicating a poorer health related quality of life.

Note that before completing the questionnaire the subject will have to answer a question on his/her present health. This info will be tabulated.

Table 5 SGRQ Questions/Items and Their Weights (from 0 to 100)

General			
Q#	**Question**	**Answer**	**Weight**
	Check how you describe your current health	Very good	NA
		Good	NA
		Fair	NA
		Poor	NA
		Very poor	NA
Part 1 **Describe how often your respiratory problems have affected you over the past 3 months.**			
Q#	**Question**	**Answer**	**Weight**
1	Over the past 3 months, I have coughed:	Almost every day	80.6
		Several days a week	63.2
		A few days a month	29.3
		Only with respiratory infections	28.1
		Not at all	0.0
2	Over the past 3 months, I have brought up phlegm (sputum):	Almost every day	76.8
		Several days a week	60.0
		A few days a month	34.0
		Only with respiratory infections	30.2
		Not at all	0.0
3	Over the past 3 months, I have had shortness of breath:	Almost every day	87.2
		Several days a week	71.4
		A few days a month	43.7
		Only with respiratory infections	35.7
		Not at all	0.0
4	Over the past 3 months, I have had wheezing attacks:	Almost every day	86.2
		Several days a week	71.0
		A few days a month	45.6
		Only with respiratory infections	36.4
		Not at all	0.0
5	How many times during the past 3 months have you suffered from severe or very unpleasant respiratory attacks?	More than 3 times	86.7
		3 times	73.5
		2 times	60.3
		1 time	44.2
		None of the time	0.0
6	How long did the worst respiratory attack last?	A week or more	89.7
		3 or more days	73.5
		1 or 2 days	58.8
		Less than a day	41.9

7	Over the past 3 months, in a typical week, how many good days (with few respiratory problems) have you had?	No good days	93.3
		1 or 2 good days	76.6
		3 or 4 good days	61.5
		Nearly every day was good	15.4
		Every day was good	0.0
8	If you have a wheeze, is it worse when you get up in the morning?	No	0.0
		Yes	62.0
Part 2 – Section 1			
Q#	**Question**	**Answer**	**Weight**
9	How would you describe your respiratory condition?	The most important problem I have	83.2
		Causes me quite a lot of problems	82.5
		Causes me a few problems	34.6
		Causes me no problems	0.0
10	If you ever held a job:	My respiratory problems made me stop working altogether	88.9
		My respiratory problems interfere with my job or made me change my job	77.6
		My respiratory problems do not affect my job	0.0
Part 2 – Section 2 **Questions about what activities usually make you feel short of breath these days.**			
I#	**Question**	**Answer**	**Weight**
11	Sitting or lying still.	False	0.0
		True	90.6
12	Washing or dressing yourself.	False	0.0
		True	82.8
13	Walking around at home.	False	0.0
		True	80.2
14	Walking outside on level ground.	False	0.0
		True	81.4
15	Walking up a flight of stairs.	False	0.0
		True	76.1
16	Walking up hills.	False	0.0
		True	75.1
17	Playing sports or other physical activities.	False	0.0
		True	72.1

Part 2 – Section 3 Questions about your cough and shortness of breath these days.			
I#	**Question**	**Answer**	**Weight**
18	Coughing hurts.	False	0.0
		True	81.1
19	Coughing makes me tired.	False	0.0
		True	79.1
20	I am short of breath when I talk.	False	0.0
		True	84.5
21	I am short of breath when I bend over.	False	0.0
		True	76.8
22	My coughing or breathing disturbs my sleep.	False	0.0
		True	87.9
23	I get exhausted easily.	False	0.0
		True	84.0
Part 2 – Section 4 **Questions about other effects that your respiratory problems may have on you these days.**			
I#	**Question**	**Answer**	**Weight**
24	My cough or breathing is embarrassing in public.	False	0.0
		True	74.1
25	My respiratory problems are a nuisance to my family, friends or neighbors.	False	0.0
		True	79.1
26	I get afraid or panic when I cannot catch my breath.	False	0.0
		True	87.7
27	I feel that I am not in control of my respiratory problems.	False	0.0
		True	90.1
28	I do not expect my respiratory problems to get any better.	False	0.0
		True	82.3
29	I have become frail or an invalid because of my respiratory problems.	False	0.0
		True	89.9
30	Exercise is not safe for me.	False	0.0
		True	75.7
31	Everything seems too much of an effort.	False	0.0
		True	84.5
Part 2 – Section 5 **Questions about your respiratory treatment.**			
I#	**Question**	**Answer**	**Weight**
32	My treatment does not help me very much.	False	0.0
		True	88.2
33	I get embarrassed using my medication in public.	False	0.0
		True	53.9

34	I have unpleasant side effects from my medication.	False	0.0
		True	81.1
35	My treatment interferes with my life a lot.	False	0.0
		True	70.3
Part 2 – Section 6 **Questions about how your activities might be affected by your respiratory problems.**			
I#	**Question**	**Answer**	**Weight**
36	I take a long time to get washed or dressed.	False	0.0
		True	74.2
37	I cannot take a bath or shower, or I take a long time to do it.	False	0.0
		True	81.0
38	I walk slower than other people my age, or I stop to rest.	False	0.0
		True	71.7
39	Jobs such as household chores take a long time, or I have to stop to rest.	False	0.0
		True	70.6
40	If I walk up one flight of stairs, I have to go slowly or stop.	False	0.0
		True	71.6
41	If I hurry or walk fast, I have to stop or slow down.	False	0.0
		True	72.3
42	My breathing makes it difficult to do things such as walk up hills, carry things up stairs, light gardening such as weeding, dance, bowl or play golf.	False	0.0
		True	74.5
43	My breathing makes it difficult to do things such as carry heavy loads, dig in the garden or shovel snow, jog or walk briskly (5 miles per hour), play tennis or swim.	False	0.0
		True	71.4
44	My breathing makes it difficult to do things such as very heavy manual work, ride a bike, run, swim fast, or play competitive sports	False	0.0
		True	63.5
Part 2 – Section 7 **Questions about how your respiratory problems usually affect your daily life.**			
I#	**Question**	**Answer**	**Weight**
45	I cannot play sports or do other physical activities.	False	0.0
		True	64.8
46	I cannot go out for entertainment or recreation.	False	0.0
		True	79.8
47	I cannot go out of the house to do the shopping.	False	0.0
		True	81.0
48	I cannot do household chores.	False	0.0
		True	79.1
49	I cannot move far from my bed or chair.	False	0.0
		True	94.0

50	Tick the statement which you think best describes how your respiratory problems affects you:	It does not stop me from doing anything I would like to do	0.0
		It stops me from doing one or two things I would like to do	42.0
		It stops me from doing most of the things I would like to do	84.2
		It stops me from doing everything I would like to do	96.7

Domain scores and total score:

- Symptoms score: questions 1-8.
- Activity score: questions 11 to 17 and 36 to 44.
- Impacts score: questions 9-10, 18 to 35 and 45 to 50.

  Total score: all questions.

Each domain score is calculated separately in three steps

- The weights for all items with positive responses are summed.
- The maximum weights for missed items are deducted from the maximum possible weight for each component. The maximum weights for all missed items are deducted from the maximum possible weight for the total score.
- The score is calculated by dividing the summed weights by the adjusted maximum possible weight for that component and expressing the result as a percentage

  $$\text{Score} = 100 \times \frac{\text{summed weights from positive items in that component}}{\text{sum of maximum weights for all non-missing items in that component}}$$

  The total score is calculated in a similar way:

  $$\text{Score} = 100 \times \frac{\text{summed weights from positive items in the questionnaire}}{\text{sum of maximum weights for all non-missing items in the questionnaire}}$$

Sum of maximum possible weights for each score:

- Symptoms: 662.5
- Activity: 1209.1
- Impacts: 2117.8
- Total: 3989.4

Note that the questionnaire requests a single response to questions 1-7, 9-10 and 50. If multiple responses are given to one of these questions then the weights for the positive responses for that question will be averaged.

In case of missing items, the following rules will be applied:

- The symptoms score will only be calculated when no more than 2 items are missing.
- The activity score will only be calculated when no more than 4 items are missing.
- The impacts score will only be calculated when no more than 6 items are missing.
- The total score will only be calculated when the 3 domain scores could be calculated.

Note that no imputation of individual missing items will be done, only the missing domain/total scores will be imputed.

Values and changes from baseline in SGRQ total score, as well as symptom, activity, and impact scores will be summarized per time point and treatment group by descriptive statistics.

A mixed-effects model with treatment, time (as a categorical factor defined by the analysis windows in Section 6.3), treatment*time, and baseline total score as factors in the model will be applied to the SGRQ total score change from baseline. LS means (with 95% CI) will be estimated from the model at each time point for each treatment group as well as the differences from placebo in LS means (and 95% CIs). This model will be used to calculate the p-values for the SGRQ total score (key secondary endpoint) to use in the multiplicity adjustment.

A patient is an SGRQ responder if his absolute change from baseline in SGRQ total score is ≤ -4 points. The number (%) of SGRQ responders at each analysis time point up to the end of the study will be presented and a logistic regression will be used to analyze the data similarly as described above for FVC responders.

In addition, a cumulative incidence table and graph will be presented for absolute change from baseline in total SGRQ score per time point and worst-case up to the end of the study for the categories ≤-4, ≤-3, ≤-2,…, ≤0, …, ≤2, ≤3, ≤4 , by treatment group.

Sensitivity Analyses

A per protocol analysis will be performed for the key secondary analysis of change in total SGRQ score.

If the graphical presentations over time indicate a linear trend a similar slope model as for the similar slope model as for the primary endpoint of rate of decline in FVC may be performed as a supportive analysis for SGRQ.

9.3.6. Other Questionnaires

The following questionnaires will be completed electronically, using a tablet device at the clinical study center.

9.3.6.1. *EQ-5D*

EQ-5D will not be analysed for the abbreviated CSR.

The 3-level version of EuroQOL 5-Dimensions Questionnaire (EQ-5D-3L) consists of two pages:

– *EQ-5D descriptive system:*
The EQ-5D descriptive system is split into five domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels, which results in a 1-digit number that expresses the level selected for that dimension. The digits for the five dimensions can be combined into a 5-digit number that describes the subject's health state (e.g. '11231'). Results will be presented as frequency of subjects having no problem, some

problems, and high problems in each domain. In addition, the 5-digit number will be converted into a single summary index score applying a formula that applies weights to each of the levels in each domain. The formula (value set) is obtained from EuroQol Research Foundation and is derived according to the preferences of the general population of a country / region (EuroQol Research Foundation. EQ-5D-3L User Guide, 2018).

– *EuroQOL visual analogue scale (EQ VAS)*
The EQ VAS records the subject's self-rated health on a vertical VAS and can be used as a quantitative measure of health outcome that reflects the subject's own judgment.

EQ-5D and EQ VAS will be presented using descriptive statistics by treatment group of actual values and changes from baseline at all time points.

9.3.6.2. *Leicester Cough Questionnaire (LCQ)*

LCQ will not be analysed for the abbreviated CSR.

Subjects with IPF commonly present with a (severe) non-productive cough. Cough will be evaluated using the LCQ. The LCQ is a 19-item questionnaire (Appendix 4) split into three domains: physical, psychological, and social. Scores are calculated by domain (range from 1 to 7) by summing the items and dividing by the number of completed items. Then these scores are added to obtain the total score (range from 3 to 21, with higher scores indicating a better health status).

Domains (questions):

– *Physical: 1,2,3,9,10,11,14,15*
– *Psychological: 4,5,6,12,13,16,17*
– *Social: 7,8,18,19*

Domain scores: total score from items in domain/number of items completed in domain (range 1-7).

Total scores: addition of domain scores (range 3–21).

LCQ total score and domain scores will be presented using descriptive statistics by treatment group of actual values and changes from baseline at all time points.

9.3.6.3. *King's Brief ILD Questionnaire (K-BILD)*

K-BILD questionnaire will not be analysed for the abbreviated CSR.

The K-BILD health status questionnaire is a 15-item questionnaire split into three domains: psychological, breathlessness and activities, and chest symptoms. Scores are weighted, using information obtained from the authors (Patel, et al., 2012) *such that every domain score and the total score range from 0 to 100, with higher scores indicating a better health status. If an item is not completed it will not be counted in the numerator.*

The final K-BILD questionnaire consisted of 15 items and three domains:

- *breathlessness and activities (questions 1, 4, 11, 13)*
- *psychological (questions 3, 5, 6, 8, 10, 12, 14)*
- *chest symptoms (questions 2, 7, 9)*

K-BILD total score and domain scores will be presented using descriptive statistics by treatment group of actual values and changes from baseline at all time points.

9.3.6.4. *VAS Cough and Urge to Cough*

VAS Cough and Urge to Cough will not be analysed for the abbreviated CSR.

VAS Cough and Urge to Cough will use a VAS of 100 mm with extremes "no cough" to "worst possible cough", and "no urge" to "highest urge to cough". VAS Cough and Urge to Cough will be presented using descriptive statistics by treatment group of actual values and changes from baseline at all time points.

9.3.7. Functional Exercise Capacity and [redacted]

For the abbreviated CSR only the 6MWT distance and [redacted] will be analysed.

The actual values and changes from baseline in 6MWT total distance walked in 6 minutes, [redacted] will be presented descriptively for each time point by treatment group.

[redacted]

Given [redacted] is measured in a resting state during the scheduled visits and during exercise in the context of the 6-MWT, these will be presented separately. In addition to the descriptive statistics, [redacted] will be presented during the study as well as per visit (see Appendix 8 of the protocol for the definition of the normal range).

9.3.8. DLCO

Change from baseline in DLCO (corrected for Hb, expressed in mmol/min/kPa) and percent predicted from normal (%) will be summarized per time point and treatment group by descriptive statistics.

9.3.9. [redacted]

[redacted]

[redacted]

[redacted]

9.3.10. Subgroup Analyses

Descriptive statistics will be provided within subgroup for the primary endpoint of rate of decline in FVC at Week 52. A similar random coefficient regression model with an additional term for the treatment-by-subgroup interaction will be used to estimate the LS means per treatment group and LS mean differences (with 95% CIs) for each subgroup. Results will be graphically presented for the LS mean difference using a forest plot. The following subgroups will be evaluated:

- *Age (<65 and ≥65 years)*
- *Sex*
- *Race*
- Stratum
- Geographic region (NA, EMEA, Latin-America, Asia-Pacific)

Given the two studies are identically designed, more subgroups may be investigated based on the pooled data to increase the size of certain small subgroups.

Summaries may also be provided for the efficacy and [redacted] *endpoints by* [redacted] *subgroups.*

9.3.11. [redacted]

[redacted]

9.4. Safety Analyses

Safety analyses will be performed using the FAS.

Continuous variables will be summarized using descriptive statistics. Frequencies and percentages will be generated for categorical variables. No formal testing will be performed to compare the treatment groups.

Safety tables will be presented by treatment group, as randomized.

9.4.1. Adverse Events

AEs will be coded according to the most recent version of the Medical Dictionary for Regulatory Activities.

A listing will be provided for all available coding steps between AE verbatim and AE system organ class, mentioning also the subjects who had the AE.

9.4.1.1. Treatment-Emergent Principle.

All adverse events starting on or after the first dose of IP and up to the last dose plus 30 days, are considered treatment-emergent adverse events (TEAE).

All AEs and changes in attributes (worsening and improvement) of AEs are reported in the database. Only worsenings of a pre-existing adverse event, after first dose of IP up to the last dose plus 30 days are to be considered as a TEAE, improvements after first dose of IP are not. A worsening is defined if at least one of the attributes of the events worsens (severity, seriousness, relationship or action taken).

Selected tables will also be created for all AEs (or worsenings) which started after the first dose of IP (till the end of the study). Adverse events starting before the first dose of IP and improvements will be excluded, but will be in the listings.

9.4.1.2. Allocation of Adverse Events to Analysis Periods

All adverse events records will be placed into analysis periods considering their start date and flag indicating the relative timing versus study (drug) start, aiming to report the incidence of these events only in the analysis period during which they started.

The general rule for allocation of AEs to analysis periods follows:

Analysis period start date ≤ AE start date ≤ analysis period end date

If the start date of an AE is missing or incomplete to a level preventing a clear allocation of the AE to one single analysis period, a worst-case consideration (see below) will be done aiming to allocate the AE record to one single analysis period, if possible. When a worst-case consideration is needed, the end date of the AE record, if and as available, and the flag indicating the relative

timing versus study end should also be considered; if such AEs clearly ends on a given point, this will exclude the possibility to allocate the AE to an analysis period after that point.

- An AE which according to the available information of its start date could belong to the screening as well as to the analysis period with treatment will only be placed in the analysis period with treatment.
- An AE which according to the available information of its start date could belong to an analysis period with treatment as well as to a next analysis period for which no treatment is defined (follow-up) will only be placed in the analysis period with treatment.
- An AE with a missing start date will be allocated to the analysis period with treatment.

9.4.1.3. Treatment Relatedness

Following (ICH-E3), the drug relatedness will be dichotomized as follows:

Drug related: at least possibly drug related as assessed by the investigator, OR with missing drug relatedness (= worst-case)

Not drug related: less than possibly drug related.

In tabulations this dichotomized parameter will be used, but in the listings the original parameter will be presented.

9.4.1.4. Worst-Case Principle

When cross-tabulating AE preferred terms versus an AE attribute (e.g. severity), the worst-case is always applied within each analysis period, i.e. when a subject has multiple times the same AE preferred term starting in the same analysis period, then the subject is reported only once: only with the worst severity. If this happens in two different analysis periods, the AE is reported twice: once in each analysis period.

9.4.1.5. Adverse Event Onset Day and Durations

For each newly reported event, and reported worsening or improvement of an existing event, the start day in the study (the day of the AE start date relative to the date of first IP administration), the start day in the analysis period, and the duration (in days) will be calculated. In addition, the relative days and duration will be derived for the entire event; that is, the full evolution of the event, including the initial reporting and all subsequent worsenings and/or improvements. Relative start- and end days of the entire event will be derived based on the start date of the first event and the end date of the last event respectively; and the duration of the entire event will be calculated by subtracting the start date of the first event of the end date of the last event + 1 day.

See section 6.1 for the calculation of relative days.

9.4.1.6. Analyses

9.4.1.6.1. Treatment-emergent AEs during the treatment period

A summary table of the number of subjects with at least one TEAE, at least one IP/SOC related TEAE, one serious TEAE, a TEAE leading to death, TEAEs by worst severity, and TEAE leading to IP/ SOC discontinuation, down-titration and interruption per treatment group will be provided. The table will also include the difference in percentage of subjects with an TEAE from placebo for each of the two dose groups, together with the 95% CI using the method of Miettinen and Nurminen (Miettinen & Nurminen, 1985).

The number (percent) of subjects with a TEAE by system organ class and preferred term will be presented. Similar tables will be provided by worst severity, by IP/SOC relationship and, by worst severity and relationship. Furthermore serious TEAEs, deaths, TEAEs leading to IP/SOC down-titration, TEAEs leading to IP/SOC interruption and TEAEs leading to IP/SOC discontinuation will be tabulated by system organ class and preferred term.

A table will also be presented, with the difference from placebo and 95% CI based on the method of Miettinen and Nurminen for the number (percent) of subjects with a TEAE by system organ class and preferred term, for events with an incidence of at least 5% in at least one of the treatment groups (600, 200 mg or placebo).

Kaplan-Meier estimates will be presented graphically and in tabular format for the following time to first event endpoints, by treatment, as well as by treatment and stratum:

- first TEAE
- first serious TEAE
- first TEAE leading to early IP discontinuation

Treatment-emergent COVID-19 infections and suspected infections will be tabulated and listed.

9.4.1.6.2. AEs from First Dose of IP up to the End of the Study

A summary table of the number of subjects with at least one AE, at least one IP/SOC related AE, one serious AE, an AE leading to death, AEs by worst severity, and AE leading to IP/SOC discontinuation, down-titration and interruption per treatment group will be provided for all AEs starting after the first dose of IP (up to the end of the study). The table will also include the difference in percentage of subjects with an AE from placebo for each of the 2 dose groups, together with the 95% CI using the method of Miettinen and Nurminen.

The number (percent) of subjects with an AE by system organ class and preferred term will also be presented, for AEs, deaths, SAEs, AEs leading to IP/SOC discontinuation down-titration and interruption.

A table will also be presented, with the difference from placebo and 95% CI based on the method of Miettinen and Nurminen for the number (percent) of subjects with an AE by system organ

class and preferred term, for events with an incidence of at least 5% in at least one of the treatment groups (600, 200 mg or placebo).

COVID-19 infections and suspected infections will also be tabulated and listed.

Listings will be provided for the serious adverse experiences (including AEs leading to hospitalization), deaths, AEs leading to death and AEs leading to down-titration, interruption and early treatment discontinuation of IP, Pirfenidine and Nintedanib.

9.4.1.6.3. EudraCT Adverse Events Reporting

For the purpose of EudraCT reporting, the following frequency tabulations will be added:

Frequency tabulations, by system organ class and preferred term, of the number (percent) of subjects with serious TEAE and the number of serious TEAEs will be presented. A similar table will be provided for the all non-serious TEAEs and for non-serious TEAEs reported in at least 5% of the subjects in any treatment group.

9.4.1.6.4. Adverse Events Related to Nintedanib or Pirfenidone

To investigate the events expected to be associated with pirfenidone and nintedanib, Kaplan-Meier curve and table, as well as a counts table by system organ class and preferred term, by treatment, as well as by stratum and treatment will be provided for the preferred terms of events of specific interest to nintedanib (e.g. diarrhea, nausea, vomiting, dehydration, weight decrease and decreased appetite adverse events) and pirfenidone (e.g. photosensitivity reactions, nausea, diarrhea, dyspepsia, vomiting, gastro-esophageal reflux disease, and abdominal pain). The full list of preferred terms to be investigated will be detailed before unblinding for the CSR.

9.4.1.7. Subgroup Analyses

For each subgroup, a summary table of the number of subjects with at least one TEAE, at least one related TEAE, a serious TEAE, leading to death, TEAEs by worst severity, and TEAE leading to down-titration, interruption and discontinuation of IP/SOC per treatment group will be provided. The number (percent) of subjects with a TEAE by system organ class and preferred term will also be presented per subgroup.

The following subgroups will be investigated:

- Age (<65 and ≥65 years),
- Race,
- Sex,
- Stratum,
- Geographic region (NA, EMEA, Latin-America, Asia-Pacific)

Since the two studies are identically designed, more subgroups may be investigated in the pooled safety analysis in the ISS/SCS to increase the size of small subgroups.

9.4.2. Laboratory Safety

Since all scheduled clinical laboratory evaluations will be performed by the central laboratory, only data from the central laboratory will be included, information from local labs will not be used. Lab tests that are not part of the planned test panels according to the protocol will only be listed.

9.4.2.1. Laboratory Units

The statistical analysis will only present results in Conventional Units.

9.4.2.2. Derived Lab Tests

Estimated creatinine clearance, calculated according to Cockcroft-Gault calculation (C_{Cr}):

C_{Cr} (in mL/min) = [(140-age) x weight]/[72 x S_{Cr}] (x 0.85 for women), with S_{Cr} = serum creatinine in mg/dL, age in years and weight in kg

9.4.2.3. Handling Non-fasted Results

For laboratory tests like glucose, triglycerides and total serum bile acid that are sensitive to fasting, descriptive statistics and tabulations of Common Terminology Criteria for Adverse Events (CTCAE) toxicities (fasting only) will be shown by fasting status. Tabulations of scorings according to normal ranges will be created for both fasting and non-fasting results pooled. Laboratory results for which the fasting status is missing, will be considered as taken fasted

9.4.2.4. Scoring According to Normal Ranges

All values will be compared to their matching normal ranges. The normal ranges provided by the central laboratory will be used and are available in the database. Values will be scored as abnormally low (L), normal (N) or abnormally high (H): the variable LBNRIND in SDTM.LB dataset will be used in the analysis.

For glucose, triglycerides and total serum bile acid in non-fasting state the following limits are to be used:

Tests	Thresholds
Glucose (non-fasting)	>= 200 mg/dL
Triglycerides (non-fasting)	>= 200 mg/dL
Total serum bile acid (non-fasting)	>x 3 ULN[1]

Only for tests for which no toxicity gradings are specified in Section 9.4.2.5, analysis of the classification according to normal ranges will be done. Except for fasting sensitive tests like glucose and triglycerides. These will be summarized with scorings based on CTCAE/other

[1] ULN: upper limit of normal range, LLN: lower limit of normal range

gradings for the fasted samples only, but also with scorings according to normal ranges for fasting and non-fasting samples pooled.

Any clinical significance flags related to these values will only be listed.

9.4.2.5. Scoring Based on CTCAE/Other Gradings

CTCAE version 5.0 classification is only for the tests listed below.

Tests	Thresholds (Conventional Units)		Thresholds (SI Units)		CTCAE v5.0
Hematology					
Hemoglobin (anemia) decreased	<LLN-10	g/dL	<LLN-100	g/L	G1
	<10 - 8	g/dL	<100 -80	g/L	G2
	<8	g/dL	<80	g/L	G3
Hemoglobin increased (increase above ULN)	Increase >0-2	g/dL	Increase >0-20	g/L	G1
	Increase >2-4	g/dL	Increase >20-40	g/L	G2
	Increase>4	g/dL	Increase>40	g/L	G3
White blood cell (WBC) count decreased	<LLN-3000	/mm3	<LLN-3.0	* 10^9/L	G1
	<3000 – 2000	/mm3	<3.0–2.0	* 10^9/L	G2
	< 2000-1000	/mm3	< 2.0-1.0	* 10^9/L	G3
	<1000	/mm3	<1.0	* 10^9/L	G4
White blood cell (WBC) count increased	>100,000	/mm3	>100.0	* 10^9/L	G3
Neutrophils decreased	< LLN-1500	/mm3	< LLN-1.5	* 10^9/L	G1
	< 1500-1000	/mm3	< 1.5-1.0	* 10^9/L	G2
	< 1000-500	/mm3	< 1.0-0.5	* 10^9/L	G3
	< 500	/mm3	< 0.5	* 10^9/L	G4
Lymphocytes decreased	< LLN-800	/mm3	< LLN-0.800	* 10^9/L	G1
	< 800-500	/mm3	< 0.8-0.5	* 10^9/L	G2
	< 500-200	/mm3	< 0.5-0.2	* 10^9/L	G3
	< 200	/mm3	< 0.2	* 10^9/L	G4
Lymphocytes increased	> 4000	/mm3	> 4.0	* 10^9/L	G2
	> 20,000	/mm3	> 20.0	* 10^9/L	G3
Eosinophils absolute count	>ULN		>ULN		G1
Eosinophils relative count	> 7%		> 7%		G1
Platelets decreased	< LLN-75000	/mm3	< LLN-75.0	* 10^9/L	G1

	< 75000 – 50000	/mm3	< 75.0 – 50.0	* 10^9/L	G2
	< 50000-25000	/mm3	< 50.0-25.0	* 10^9/L	G3
	< 25000	/mm3	< 25.0	* 10^9/L	G4
Coagulation					
Activated partial thromboplastin time (aPTT)	> ULN - 1.5 x ULN		> ULN - 1.5 x ULN		G1
	> 1.5 - 2.5 x ULN		> 1.5 - 2.5 x ULN		G2
	> 2.5 x ULN		> 2.5 x ULN		G3
Prothrombin time (PT)	> ULN - 1.5 x ULN		> ULN - 1.5 x ULN		
	> 1.5 - 2.5 x ULN		> 1.5 - 2.5 x ULN		/
	> 2.5 x ULN		> 2.5 x ULN		
INR increased	1.2 – 1.5 xULN		> 1.2 – 1.5 xULN		G1
	> 1.5 - 2.5 xULN		> 1.5 - 2.5 xULN		G2
	> 2.5 xULN		> 2.5 xULN		G3
Chemistry					
Glucose decreased	< LLN – 55	mg/dL	< LLN - 3.0	mmol/L	G1
	< 55 – 40	mg/dL	< 3.0-2.2	mmol/L	G2
	< 40 – 30	mg/dL	< 2.2 -1.7	mmol/L	G3
	< 30	mg/dL	< 1.7	mmol/L	G4
Glucose increased (fasting)	> 130	mg/dL	> 7.2	mmol/L	/
Creatinine increased	> ULN - 1.5 x ULN		> ULN - 1.5 x ULN		G1
	>1.5-3x ULN		>1.5-3x ULN		G2
	> 3 – 6x ULN		> 3 – 6x ULN		G3
	> 6x ULN		> 6x ULN		G4
eGFR / Creatinine Clearance ml/min/1.73 m2	< LLN-60	ml/min	< LLN-60	ml/min	G1
	<60-30	ml/min	<60-30	ml/min	G2
	< 30 -15	ml/min	< 30 -15	ml/min	G3
	< 15	ml/min	< 15	ml/min	G4
Sodium decreased	<LLN - 130	mmol/L	<LLN - 130	mmol/L	G1
	< 130 – 125	mmol/L	< 130 – 125	mmol/L	G2
	< 125 – 120	mmol/L	< 125 – 120	mmol/L	G3
	< 120	mmol/L	< 120	mmol/L	G4
Sodium increased	> ULN - 150	mmol/L	> ULN - 150	mmol/L	G1
	> 150 - 155	mmol/L	> 150 - 155	mmol/L	G2
	> 155 - 160	mmol/L	> 155 - 160	mmol/L	G3

	> 160	mmol/L	> 160	mmol/L	G4
Potassium decreased	< LLN - 3 < 3 - 2.5 < 2.5	mmol/L mmol/L mmol/L	< LLN - 3 < 3 - 2.5 < 2.5	mmol/L mmol/L mmol/L	G1 G3 G4
Potassium increased	> ULN – 5.5 > 5.5 – 6 > 6 – 7 > 7	mmol/L mmol/L mmol/L mmol/L	> ULN – 5.5 > 5.5 – 6 > 6 – 7 > 7	mmol/L mmol/L mmol/L mmol/L	G1 G2 G3 G4
Calcium decreased	< LLN - 8 < 8-7 < 7-6 < 6	mg/dl mg/dl mg/dl mg/dl	< LLN - 2.0 < 2-1.75 < 1.75-1.5 < 1.5	mmol/L mmol/L mmol/L mmol/L	G1 G2 G3 G4
Calcium increased	> ULN - 11.5 > 11.5 - 12.5 > 12.5 - 13.5 > 13.5	mg/dl mg/dl mg/dl mg/dl	> ULN - 2.9 > 2.9 - 3.1 > 3.1 - 3.4 > 3.4	mmol/L mmol/L mmol/L mmol/L	G1 G2 G3 G4
AST	>ULN-3 x ULN > 3 -5 xULN > 5 -20 xULN > 20 xULN		>ULN-3 x ULN > 3 -5 xULN > 5 -20 xULN > 20 xULN		G1 G2 G3 G4
ALT	>ULN-3 x ULN > 3 -5 xULN > 5 -20 xULN > 20 xULN		>ULN-3 x ULN > 3 -5 xULN; > 5 -20 xULN > 20 xULN		G1 G2 G3 G4
GGT	> ULN – 2.5 x ULN >2.5 -5 x ULN > 5 -20 *ULN > 20 *ULN		> ULN – 2.5 x ULN >2.5 -5 x ULN > 5 -20 *ULN > 20 *ULN		G1 G2 G3 G4
Total bilirubin	> ULN – 1.5 x ULN > 1.5 -3 x ULN > 3-10 x ULN > 10 x ULN		> ULN – 1.5 x ULN > 1.5 -3 x ULN > 3-10 x ULN > 10 x ULN		G1 G2 G3 G4
Alkaline phosphatase (aP)	> ULN – 2.5 x ULN >2.5 - 5 x ULN >5 -20 x ULN > 20 x ULN		> ULN – 2.5 x ULN >2.5 - 5 x ULN >5 -20 x ULN > 20 x ULN		G1 G2 G3 G4
Albumin	< LLN - 3 < 3 - 2 < 2	g/dL g/dL g/dL	< LLN - 30 < 30 - 20 < 20	g/L g/L g/L	G1 G2 G3

Triglycerides (fasting)	> 150 - 300 > 300 - 500 > 500 – 1000 > 1000	mg/dL mg/dL mg/dL mg/dL	> 1.71 - 3.42 > 3.42 - 5.7 > 5.7 – 11.4 > 11.4	mmol/L mmol/L mmol/L mmol/L	G1 G2 G3 G4
Cholesterol	> ULN - 300 > 300 - 400 > 400 - 500 > 500	mg/dL mg/dL mg/dL mg/dL	> ULN - 7.75 > 7.75 - 10.34 > 10.34 - 12.92 > 12.92	mmol/L mmol/L mmol/L mmol/L	G1 G2 G3 G4
High density lipoprotein (HDL)	< 60	mg/dL	< 1.554	mmol/L	
Low density lipoprotein (LDL)	> 160	mg/dL	> 4.144	mmol/L	
Creatine kinase (CK)	> ULN - 2.5 x ULN >2.5 -5 x ULN > 5 – 10 x ULN > 10 x ULN		> ULN - 2.5 x ULN >2.5 -5 x ULN > 5 – 10 x ULN > 10 x ULN		G1 G2 G3 G4

LLN=lower limit of normal range, ULN=upper limit of normal range

The following criteria will be used for C_{cr}.

Tests	Thresholds	Grade	Terms
C_{Cr} (in mL/min)	≥90	G1	Normal or high
	60 to <90	G2	Mildly decreased
	45 to <60	G3a	Mildly to moderately decreased
	30 to <45	G3b	Moderately to severely decreased
	15 to <30	G4	Severely decreased
	<15	G5	Kidney failure

9.4.2.6. Handling of Urinalysis Results

Results of urinalysis with qualitative results will be tabulated by time point. No toxicity grading or non-graded abnormalities will be derived.

9.4.2.7. Worst-Case Abnormality

9.4.2.7.1. According to Normal Range

The worst-case postbaseline will be determined per subject, per laboratory test (and sense, if below and above) and for each analysis period. All non-missing postbaseline values (including unscheduled and follow-up visits, but excluding local laboratory results) will be used to derive the following worst-case:

- L = low: at least one post-baseline result is classified as L.
- N = normal: all post-baseline results are classified as N.

– H = high: at least one post-baseline result is classified as H.

If, for a subject, both L and H are reported, the subject will be counted twice in the table: once with a worst-case L and once with a worst-case H.

Tables will be created for treatment-emergent abnormalities, i.e. postbaseline abnormalities which differ from (and are not better than) the baseline result. If the baseline result is missing, a postbaseline abnormality L or H will always be considered as treatment-emergent.

9.4.2.7.2. According to CTCAE Toxicity Grades

The worst-case postbaseline toxicity grade 0, 1, 2, 3 or 4 will be determined per subject, per laboratory test (and sense, if below and above) and for each analysis period, using all non-missing postbaseline records (including unscheduled and follow-up visits, but excluding local laboratory results).

The worst-case toxicity grade is the highest toxicity grade scored for the laboratory test (in each sense, if below and above).

Tables will be created for treatment-emergent toxicity grades. A postbaseline toxicity grade 1, 2, 3 or 4 is defined as treatment-emergent when higher than the toxicity grade of the baseline result. If the baseline result is missing, a postbaseline toxicity grade 1, 2, 3 or 4 will always be considered as treatment-emergent.

9.4.2.8. Analyses

No formal inferential statistics (p-values) will be derived.

For each laboratory test, descriptive statistics and graphical presentation (line plot of the mean and SE) will be provided for the actual laboratory values and change from baseline over time-by-treatment group.

A summary table per laboratory test category (hematology, biochemistry, urinalysis and coagulation), laboratory test and unit, treatment group and time point, showing only the mean value of actual values and changes from baseline will also be provided.

All laboratory abnormalities after the first dose of IP up to 30 days after the last dose will be included, consistently with AE reporting.

A shift table per laboratory test category (hematology, biochemistry, urinalysis and coagulation), laboratory test and treatment group will be provided for laboratory tests with normal ranges, for which CTCAE grades are not available. The table will present the worst shift in abnormality (L/N/…) during the treatment period versus the baseline abnormality (L/N/H).

Frequency table will be provided of the worst treatment-emergent laboratory abnormalities per laboratory test category (hematology, biochemistry, urinalysis and coagulation), laboratory test, and treatment group. This table will be created for those laboratory tests for which CTCAE grades are not available.

A shift table per laboratory test category (hematology, biochemistry and coagulation), laboratory test and treatment group for CTCAE grades will also be provided for selected tests only. The table will present the worst shift in CTCAE toxicity grade during the treatment period versus the baseline value.

A frequency table of the worst treatment-emergent laboratory CTCAE toxicity grades per laboratory test category (hematology, biochemistry and coagulation), laboratory test, and treatment group will be provided for selected tests only.

Shift table per laboratory test category (hematology, biochemistry, and urinalysis), laboratory test, and treatment group for laboratory tests with categorical data. The table will present the worst shift in value during the treatment period versus the baseline value.

9.4.2.9. Elevated Liver Function Test

Time to first liver enzyme elevation in conjunction with bilirubin elevation (alanine aminotransferase (ALT) and/or aspartate aminotransferase (AST) ≥3x upper limit of normal (ULN) and bilirubin ≥ 1.5xULN) will be presented by a graph and table.

A graphical and tabular presentation (Kaplan-Meier) of the time to first liver enzyme elevations (per level) as defined in the protocol, ALT and/or AST

- ≥ 1.5xULN
- ≥ 3xULN
- ≥ 5xULN
- ≥ 8xULN

In addition a similar shift table and frequency table as for the CTCAE grades defined in Section 9.4.2.8 will be provided using the grades shown below:

Tests	Thresholds
AST/ALT combination	AST 1.5 - <3 x ULN AST >= 3 - <5 x ULN AST >= 5 - <8 x ULN AST >= 8 x ULN ALT 1.5 - <3x ULN ALT >= 3 - <5 x ULN ALT >= 5 - <8 x ULN ALT >= 8 x ULN AST and/or ALT 1.5 - <3 x ULN AST and/or ALT >= 3 - <5 x ULN AST and/or ALT >= 5 - <8 x ULN AST and/or ALT >= 8 x ULN

Tests	Thresholds
AST/ ALT / bilirubin combination	Bilirubin >= 1.5 x ULN and AST or ALT >= 3 x ULN Bilirubin >= 2 x ULN and AST or ALT >= 3 x ULN

9.4.3. ECG

9.4.3.1. Calculated Parameters

Fridericia's cube-root corrected QT (Fridericia, 1920) (QTcF) will be provided by the central vendor and will be used as available in the database and will not be re-derived.

9.4.3.2. Handling Triplets

If ECG is collected in triplicates, the following approach will be taken.

First, any derivation of ECG parameters will always be done before handling ECG triplicates. Next, the mean of the triplicate ECG values will be calculated for each individual ECG parameter, without rounding the result. These calculated means will constitute the analysis values; any derivation (e.g. change from baseline, assignment of abnormalities) and statistic will be based on the mean value of the triplets.

The values of the original members of a triplicate will be listed only.

9.4.3.3. Normal Ranges

For the QT and QTcF parameters, the following categorizations will be used:

- of the actual values:
  - <= 450 ms,
  - >450 - ≤480 ms,
  - >480- ≤500 ms,
  - >500 ms;
- of the changes from baseline:
  - <= 30 ms (including all decreases in QT),
  - >30- ≤60 ms,
  - >60 ms.

For PR interval, the following categorizations will be done:

- Low: < 120 ms,
- Normal: ≥120- ≤ 220 ms,
- High: >220 ms;

For heart rate, the following categorizations will be done:

- Low: < 50 bpm,

- Normal: ≥50- ≤ 100 bpm,
- High: >100 bpm;

9.4.3.4. Worst-case Abnormality of QT and QTcF

The worst-case post-baseline categorized actual analysis value and the worst-case categorized change from baseline for QT and QTcF will be determined per subject, per parameter, and for each analysis period, using all non-missing post-baseline records (including unscheduled and follow-up visits).

The worst-case categorized actual analysis value is the category corresponding to the highest post-baseline actual value.

The worst-case change from baseline is the category corresponding to the largest increase (positive change) from baseline.

Tables will be created for treatment-emergent abnormalities. For actual values, an abnormal post-baseline abnormality is defined as treatment-emergent when the abnormality is worse compared to the abnormality at baseline. When the baseline value is missing, post-baseline abnormalities are considered as treatment-emergent.

An abnormal category for change from baseline is always treatment-emergent.

9.4.3.5. Worst-case Abnormality of PR and HR

The following worst-case post-baseline abnormalities L, N or H will be determined per subject, per parameter and for each analysis period, using all non-missing post-baseline records (including unscheduled and follow-up visits):

– L = low: at least one post-baseline result is classified as L.
– N = normal: all post-baseline results are classified as N.
– H = high: at least one post-baseline result is classified as H.

If, for a subject, both L and H are reported, the subject will be counted twice in the table: once with a worst-case L and once with a worst-case H.

Tables will be created for treatment-emergent abnormalities, i.e. postbaseline abnormalities which differ from (and are not better than) the baseline result. If the baseline result is missing, a postbaseline abnormality L or H will always be considered as treatment-emergent.

9.4.3.6. Analyses

No formal inferential statistics (p-values) will be derived.

Descriptive statistics will be summarized of the actual values and change from baseline (including 95% CI of the mean change) per parameter and unit, treatment group and time point. The actual value at baseline will also be presented in this table. Only pre-dose assessments will be used for this table. A similar summary table as for labs will also be provided.

In addition to the predose measurements, ECGs are also taken 2-3 hours postdose on selected visits (Day 1, Week 2, 4, 12 and 34 for ECG). These postdose measurements will not be taken into account in the windowing for the above tables, but an additional table will be created summarizing the pre- and postdose values at these selected visits.

All ECG abnormalities after the first dose of IP up to 30 days after the last dose will be included for consistency with AE reporting.

Abnormalities of the actual values will be presented as shift tables of the worst-case abnormality versus the baseline abnormality. The table will be created per parameter (QT, QTcF, PR and heart rate), treatment group and analysis period.

A frequency table of the number (percent) of subjects with treatment-emergent worst-case abnormalities and the worst change per parameter (QT, QTcF, PR and heart rate), treatment group and analysis period will be presented.

Frequency table per treatment group and time point of the ECG interpretation scores as recorded in the eCRF. Note that more than one of these items can be ticked.

No figures are planned for ECGs.

Listing will be provided per treatment group, per subject and per time point of all postbaseline data scored as out-of-normal-range or postbaseline clinically significant, plus also the baseline reference time point.

9.4.4. Vital Signs

Summaries will be provided for the vital signs heart rate, diastolic and systolic blood pressure, temperature, respiratory rate and oxygen saturation.

9.4.4.1. Normal Ranges

Normal ranges are defined in Appendix 8 of the Clinical Study Protocols.

Values equal to the boundaries are still considered normal (N). A value is classified as abnormally low (L) when the value < lower limit of the normal range. A value is classified as abnormally high (H) when the value > upper limit of the normal range.

9.4.4.2. Worst-Case Abnormality

The following worst-case post-baseline abnormalities L, N or H will be determined per subject, per parameter and for each analysis period, using all non-missing post-baseline records (including unscheduled and follow-up visits):

- L = low: at least one post-baseline result is classified as L.
- N = normal: all post-baseline results are classified as N.
- H = high: at least one post-baseline result is classified as H.

If, for a subject, both L and H are reported, the subject will be counted twice in the table: once with a worst-case L and once with a worst-case H.

Tables will be created for treatment-emergent abnormalities, i.e. postbaseline abnormality which differ from (and are not better than) the baseline result. If the baseline result is missing, a postbaseline abnormality L or H will always be considered as treatment-emergent.

9.4.4.3. Analyses

No formal inferential statistics (p-values) will be derived.

Descriptive statistics will be summarized of the actual values and change from baseline (including 95% CI of the mean change) per parameter and unit, treatment group and time point. The actual value at baseline will also be presented in this table. A similar summary table as for labs will also be provided.

All vital signs abnormalities after the first dose of IP up to 30 days after the last dose will be included.

Abnormalities of the actual values will be presented as shift tables of the worst-case abnormality versus the baseline abnormality. The table will be created per parameter (and position if applicable), treatment group and analysis period.

A frequency table of the number (percent) of subjects with treatment-emergent worst-case abnormalities per parameter, treatment group and analysis period will be presented.

No figures are planned for vital signs.

A listing will be created per treatment group, per subject and per time point of all postbaseline time points scored as treatment-emergent out-of-normal-range, plus also the baseline reference time point.

9.4.5. *Physical Examinations*

A counts table will be created for the number (%) of subjects with at least one treatment-emergent abnormality by system organ class.

A listing will be created per treatment group, per subject and per time point of the selection of all abnormal findings.

9.5. Pharmacokinetic Assessments

Blood samples for the PK assessment of GLPG1690, pirfenidone and nintedanib should be collected as specified in the Schedule of Activities in section 2.5.

Subject profile plots over time will be provided for the PK concentrations of GLPG1690, pirfenidone and nintedanib. In addition, listings of all individual plasma concentration data will be created. More detailed population PK[redacted] analyses will be described in a separate SAP.

10. REFERENCES

EuroQol Research Foundation. EQ-5D-3L User Guide. (2018). Retrieved from https://euroqol.org/publications/user-guides

Fridericia, L. (1920). Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. *J Int med, 53(1)*, 469-486.

Halperin, M., Lan, K., Wright, E., & Foulkes, M. (1987). Stochastic curtailing for comparison of slopes in longitudinal studies. *Controlled Clinical Trials, 8(4)*, 315-326.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika, 75*, 800-2.

Jennison, C., & Turnbull, B. (2000). *Group Sequential Methods with Applications to Clinical Trials.* Boca Raton: Chapman & Hall/CRC.

Keene, O., Jones, M., Lane, P., & Anderson, J. (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat, 6(2)*, 89-97.

Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine, 4*, 232-226.

Miller, M., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R. C., Casaburi, R., & Coates, A. (2005). Standardisation of spirometry. *Eur Respiratory J, 26(2)*, 319-338.

Patel, A., Siegert, R., Brignall, K., Gordon, P., Steer, S., & Desai, S. (2012). The development and validation of the King's Brief Interstitial Lung Disease (K-BILD) health status questionnaire. *Thorax, 67(9)*, 804-810.

Ratitch B, O. M. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. . *Pharm Stat , 40(6)*, 12 (6):337-47.

Rubin, D. (1987). *Multiple Imputation for nonresponse in surveys.* New York: John Wiley and Sons.

11. APPENDICES

1. ADDITIONAL DETAILS ON PRIMARY ANALYSIS

For both parameters (FVC [in mL] or percent predicted FVC, separately), the annual rate of decline will be analyzed using a random coefficient regression model (linear slope model) including sex, age, height and stratification factor (stratum: pirfenidone, nintedanib and neither) as covariates and a random intercept and slope. The treatment effect is determined by using estimated slopes for each treatment group on the basis of the time-by-treatment interaction term from the mixed model. All available FVC values, as defined by the analysis windows will be used, including FVC measurements at the follow-up visit for subjects who discontinue the IP prematurely and do not complete the study visits. The model will be fit using the SAS Proc Mixed procedure, details are below. The assumptions of the model will be investigated.

In the primary analysis only data up to the Week 52 will be used, before a lung transplant. In a secondary analysis all data will be used up to the end of the study, before a lung transplant.

The statistical model can be written as follows:

$Y_{itk} = (\alpha + a_i) + (\gamma + \beta_{T1}T_{k1} + \beta_{T2}T_{k2} + g_i)t + \beta_g$ Gender $+ \beta_a$ age $+ \beta_h$ Height $+ \beta_{sP}$ Stratum (P)$+ \beta_{sN}$ Stratum (N) $+ \varepsilon_{it}$

Where Y_{itk} is the FVC value for the i^{th} subject at time t in treatment group k, T_{k1}=0 for placebo and T_{k1}=1 for GLPG1690 200 mg, T_{k2}=0 for placebo and T_{k2}=1 for GLPG1690 600 mg, β_{T1}, β_{T2} are the effects of GLPG1690 on the slope.

α and γ are the fixed effects for the intercept and slope, respectively

a_i and g_i are random effects for the intercept and slope for subject i, these are assumed to be normally distributed with mean 0 and unstructured variance-covariance matrix.

β_g, β_a, β_{sP}, β_{sN} and β_h are subject specific demographic coefficients for gender, age, stratification factor (standard of care) and height at baseline.

ε_{it} is the random error for subject i at time t, assumed to be normally distributed with mean 0 and variance σ_ε^2.

An unstructured variance-covariance matrix will be used to model the variance-covariances for the random slope, random intercept. The variance-covariance matrix for the error terms is assumed to have a variance-component structure.

If this model fails to converge, other structures will be tested for the variance-covariance matrix for the error terms (e.g. heterogeneous Toeplitz). Depending on the structure used, the random slope and random intercept may need to be dropped from the model. The structure with the best fit, based on Akaike's information criterion, will then be used.

The linearity assumption of the model will be explored graphically be the lineplots over time of the observed FVC values (mL) as well as change from baseline and LS means at each time point.

The following SAS code is to be used:

```
proc SORT data = FVC ;
  By subject visit ;
run ;
Proc MIXED data = FVC order = internal;
   CLASS subject treatment visit gender stratum;
   MODEL FVC = treatment * time gender age height stratum / solution CL ddfm=KR ;
   RANDOM intercept time / type=UN subject = subject ;
   REPEATED visit / type= Simple subject = subject ;
   ESTIMATE "GLPG 1690 600 mg qd - Placebo" treatment*time -1 0 1 / e CL ;
   ESTIMATE "GLPG 1690 200 mg qd - Placebo" treatment*time -1 1 0 / e CL ;
run ;
```

2. IMPLEMENTATION OF MULTIPLE IMPUTATIONS FOR PATTERN MIXTURE APPROACH

The following steps provide a guidance on how missing FVC data at Week 52 will be imputed in the pattern mixture approach for subjects of pattern 3, 4 and 5 using multiple imputation.

1. Run the random coefficient regression model for the primary analysis to get the slope estimates by treatment group and stratum.
2. Assume that the true slopes follow a normal distribution with mean and standard error obtained in step 1.
3. Impute missing values multiple times (1000 imputations per subject). To do so, draw random slopes from the distributions in step 2. Considering that the withdrawal of a subject leading to missing data can occur at any time during the study, the time point 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 + $\widehat{\beta_{ij}}$ (time between last available FVC and Day 365 in days)

   where i denotes the indicator of the i^{th} subject, j the indicator of the j^{th} imputation and β_{ij} denotes a random slope sampled from the distribution mentioned in step 2.
   Depending on the sensitivity analysis, use either random slopes drawn from the distribution for the slope obtained for subjects randomized to active treatment or placebo.
4. Run the same random coefficient regression model for each of the 1000 imputed datasets.
5. Combine the estimates of the 1000 analyses in step 4 using SAS PROC MIANAYZE

3. ADDITIONAL DETAILS ON MIXED-EFFECT MODEL FOR REPEATED MEASURES ANALYSIS

A mixed-effects model for repeated measures (MMRM) with treatment*time (categorical) , stratum, gender, age, height and baseline value as factors in the model will be applied to the change from baseline in FVC value in mL. Least squares (LS) means (with 95% CI) will be estimated from the model at each time point for each treatment group as well as the differences from placebo in LS means (and 95% CIs).

The SAS code for the MMRM model for FVC change from baseline is given below. The same model will be used for SGRQ.

```
proc SORT data= FVC;
        BY patient visit;
run;
proc MIXED data=FVC order=internal covtest ;
        CLASS patient treatment visit;
        MODEL change_from_baseline= treatment*visit baseline  gender age height stratum
/solution ddfm=KR;
        REPEATED visit/type=UN subject =patient;
        LSMEANS treatment*visit / cl diff;
run;
```

If the model does not converge with an unstructured variance-covariance model, other structures will be tried (e.g. heterogenous Toepliz, AR(1),…). The structure with the best fit, based on Akaike's information criterion, will then be used.

4. LEICESTER COUGH QUESTIONNAIRE

This questionnaire is designed to assess the impact of your cough on various aspects of your life. Read each question carefully and answer by CIRCLING the response that best applies to you. Please answer ALL questions, as honestly as you can.

1. In the last 2 weeks, have you had chest or stomach pains as a result of your cough?

1	2	3	4	5	6	7
All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time

2. In the last 2 weeks, have you been bothered by phlegm production when you cough?

1	2	3	4	5	6	7
Every time	Most times	Several times	Sometimes	Occasionally	Rarely	Never

3. In the last 2 weeks, have you been tired because of your cough?

1	2	3	4	5	6	7
All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time

4. In the last 2 weeks, have you felt in control of your cough?

1	2	3	4	5	6	7
None of the time	Hardly any of the time	A little of the time	Some of the time	A lot of the time	Most of the time	All of the time

5. How often during the last 2 weeks have you felt embarrassed by your coughing?

1	2	3	4	5	6	7
All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time

6. In the last 2 weeks, my cough has made me feel anxious

1	2	3	4	5	6	7
All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time

7. In the last 2 weeks, my cough has interfered with my job, or other daily tasks

	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
8.	In the last 2 weeks, I felt that my cough interfered with the overall enjoyment of my life						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
9.	In the last 2 weeks, exposure to paints or fumes has made me cough						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
10.	In the last 2 weeks, has your cough disturbed your sleep?						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
11.	In the last 2 weeks, how many times a day have you had coughing fits?						
	1	2	3	4	5	6	7
	All of the time (continuously)	Most times during the day	Several times during the day	Sometimes during the day	Occasionally through the day	Rarely	None
12.	In the last 2 weeks, my cough has made me feel frustrated						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
13.	In the last 2 weeks, my cough has made me feel fed up						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
14.	In the last 2 weeks, have you suffered from a hoarse voice as a result of your cough?						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
15.	In the last 2 weeks, have you had a lot of energy?						
	1	2	3	4	5	6	7
	None of the time	Hardly any of the time	A little of the time	Some of the time	A lot of the time	Most of the time	All of the time
16.	In the last 2 weeks, have you worried that your cough may indicate a serious illness?						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
17.	In the last 2 weeks, have you been concerned that other people think something is wrong with you because of your cough?						
	1	2	3	4	5	6	7
	All of the time	Most of the time	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
18.	In the last 2 weeks, my cough has interrupted conversations or telephone calls						
	1	2	3	4	5	6	7
	Every time	Most times	A lot of the time	Some of the time	A little of the time	Hardly any of the time	None of the time
19.	In the last 2 weeks, I feel that my cough has annoyed my partner, family or friends						
	1	2	3	4	5	6	7
	Every time I cough	Most times when I cough	Several times when I cough	Sometimes when I cough	Occasionally when I cough	Rarely	Never

Thank you for completing this questionnaire.

5. SUMMARY OF THE FUTILITY APPROACH

In view of the termination of the trial, the below section is not applicable.

As mentioned in Section 8, an interim analysis will be performed to assess futility based on the effect of treatment with GLPG1690 on the annual rate of decline in FVC using the pooled interim data from studies GLPG1690-CL-303 and GLPG1690-CL-304. The IDMC will review the results of this analysis and make a recommendation to the sponsor on the continuation of both studies. To allow collection of as much safety information as possible, a study will only be considered futile when both doses show futility. In addition, a study will only be stopped, when both studies meet the condition for futility.

Futility of a dose in a study will be assessed by a conditional power approach (Jennison & Turnbull, 2000). Conditional power is defined as the probability to conclude a significant effect for the final analysis, given the interim data and an assumption for the effect in the unobserved data after the interim. In the primary analysis (see Section 9.3.2.1) all data up to and including the

52-week time window will be included, therefore conditional power in the context of stochastic curtailing for longitudinal data will be implemented (Halperin, Lan, Wright, & Foulkes, 1987). A dose will be considered futile if the conditional power is lower or equal to a certain stopping boundary, assuming the effect size in the unobserved data is equal to the point estimate from the interim analysis.

To determine the best choice for the timing of the interim analysis and the stopping boundary, simulations were run under different scenarios of the true treatment effect for a range of timings of the interim and stopping boundaries. Based on the results of these simulations, values were selected such that the probability to stop in the interim is high if there is no effect and low for effects around 80 mL and higher. This led to the selection of a stopping boundary of 40% .i.e. a dose will be considered futile if the conditional power is lower than 40%, the interim analysis will be planned when approximately 70% of total information is collected or when at least 25% of subjects from the two studies combined have completed 52 weeks of treatment (whichever comes latest). Note that the term ‘information’ in this context is defined as the information level associated with the primary test statistic (i.e. the inverse of the variance of the treatment estimate at interim) (Jennison & Turnbull, 2000).