STATISTICAL ANALYSIS PLAN

Project Number:	GLPG1205
Study Number:	GLPG1205-CL-220
Study Title:	A Phase II randomized, double-blind, placebo-controlled, 26-week study to evaluate the efficacy, safety and tolerability of GLPG1205 in subjects with idiopathic pulmonary fibrosis

Development Phase:	2		
Status:	Final		
Version:	3.00	Date:	4-Aug-2020
EudraCT:	2017-004302-18	IND:	Not applicable
Sponsor:	Galapagos NV Industriepark Mechelen Noord Generaal De Wittelaan L11 A3 2800 Mechelen, Belgium		
Biostatistician:	[redacted], PhD		
Medical Leader:	[redacted] MD [redacted], MD		

TABLE OF CONTENTS

VERSION HISTORY

SAP Amendment #	Date	Description of changes
SAP Version 1.0	05 Dec 2019	
SAP Version 2.0	05 Dec 2019	Correction of 'draft' status to 'final'. No other changes made.
SAP Version 3.0	4 Aug 2020	Update of the SoA in Section 2.5, addition of an addendum in Section 3.3.3 and addition of a sensitivity analysis in Section 5.2.2.2.4 as a consequence of the implemented USMs due to COVID-19 in CSP Version 3.00 amendment 2. Update of the stratum definition for efficacy analyses in Section 3.4.4. [redacted] Generalization of the planned PK graphs in Section 5.4.

LIST OF ABBREVIATIONS

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}	creatinine clearance
CEAC	clinical endpoint adjudication committee
CI	confidence interval
COVID-19	Corona virus disease 2019
CRF	case report form
CSR	clinical study report
CT	computed tomography
CTCAE	Common Terminology Criteria for Adverse Events
DLCO	diffusing capacity of the lung for carbon monoxide
ECG	electrocardiogram
eCRF	electronic case report form
EOS	end of study
EOT	end of treatment
ETD	early treatment discontinuation
EU	European Union
FAS	full analysis set
FDA	Food and Drug Administration
FEF_{25-75}	forced expiratory flow between 25% and 75% of exhaled volume
FEV_1	forced expiratory volume in 1 second
[redacted]	[redacted]
[redacted]	[redacted]
FVC	forced vital capacity
FSH	follicle-stimulating hormone
Hb	hemoglobin
HR	heart rate
HRCT	high-resolution computed tomography
ICF	informed consent form

ICH	International Council for Harmonization
IDMC	Independent Data Monitoring Committee
IMP	investigational medicinal product
INF	infinity
IPF	idiopathic pulmonary fibrosis
IWRS	interactive web response system
LB	lung biopsy
LLN	lower limit of normal
LS	least squares
MAR	missing-at-random
MNAR	missing not at random
OC	observed cases
[redacted]	[redacted]
PFT	pulmonary function test
PK	pharmacokinetic(s)
PP	per protocol
q.d.	once daily (quaque die)
QTcF	QT interval corrected for the heart rate using Fridericia's formula
SAE	serious adverse event
SAP	statistical analysis plan
SE	standard error
SGRQ	St. George's Respiratory Questionnaire
SOC	standard of care
TEAE	treatment-emergent adverse event
[redacted]	[redacted]
ULN	upper limit of the normal range
US	United States
USM	urgent safety measure

1. INTRODUCTION

This statistical analysis plan (SAP) describes the final and potential interim statistical analyses of study GLPG1205-CL-220 (PINTA). An interim analysis may be performed upon sponsor decision to allow planning of future clinical studies. In addition, an unblinded review of data will be performed on a regular basis by an Independent Data Monitoring Committee (IDMC). The IDMC analyses are detailed in a separate SAP and will be provided to the IDMC by an independent, unblinded statistical support group.

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

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

The statistical analysis will process and present the results following the International Council for Harmonization (ICH) standards, particularly the ICH-E3, ICH-E6, ICH-E9, ICH-R9 (R1) and ICH-E14 guidelines.

2. STUDY DESIGN AND OBJECTIVES

2.1. Study Objectives

2.1.1. Primary Objective

- To evaluate the efficacy of GLPG1205 treatment in subjects with idiopathic pulmonary fibrosis (IPF) on pulmonary function as evaluated by forced vital capacity (FVC) compared to placebo over 26 weeks.

2.1.2. Secondary Objectives

- To evaluate the safety and tolerability of GLPG1205 treatment compared to placebo over 26 weeks.
- To evaluate the impact of GLPG1205 treatment compared to placebo on time to any major events (whichever occurs first) defined as:
  - Mortality (all-cause and respiratory-related)
  - Hospitalization (all-cause and respiratory-related)
- To evaluate the changes from baseline in functional exercise capacity measured by the 6-minute walk test (6MWT), in IPF subjects treated with GLPG1205 compared to placebo at Week 26.
- To evaluate the changes in quality of life measures in IPF subjects treated with GLPG1205 compared to placebo over 26 weeks.
- To evaluate the PK of GLPG1205, nintedanib and pirfenidone in IPF subjects.

2.1.3. Other Objectives

[redacted]

2.2. Study Endpoints

2.2.1. Primary Endpoint

– Change from baseline in FVC (mL) over 26 weeks compared to placebo.

2.2.2. Secondary Endpoints

- Safety and tolerability changes over time (baseline to 26 weeks).
- Time to any of major events (whichever occurs first) defined as:
  - Death (all-cause and respiratory-related)
  - First hospitalization (all-cause and respiratory-related)
- Change from baseline 26 weeks in functional exercise capacity, assessed by the 6MWT at Week 26.
- Change from baseline until 26 weeks in quality of life measures, assessed by the St. George's Respiratory Questionnaire (SGRQ) total score and domains and proportion of SGRQ responders.

– Plasma concentrations of GLPG1205, nintedanib and pirfenidone.

2.2.3. Other Endpoints

[redacted]

2.2.4. Adjudication

The following 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)

Adjudicated endpoints will be used in the time-to-event analyses described in Section 5.2.3.1 and not the corresponding investigator-reported events. Consistency with the investigator-reported information will be investigated.

2.3. Study Design

This study is a randomized, double-blind, parallel group, placebo-controlled, multicenter, Phase 2 study.

Approximately 60 eligible subjects with a centrally confirmed IPF diagnosis are planned to be randomized in a 2:1 ratio to GLPG1205 100 mg q.d. taken as 2 capsules, or matching placebo once daily (q.d.) administered for 26 weeks on top of local standard of care (SOC). Standard of care is defined as receiving nintedanib, pirfenidone, or neither nintedanib nor pirfenidone.

Individual dose reductions of GLPG1205 will be allowed once in case of intolerance or toxicity at the discretion of the investigator. Subjects may have their dose of investigational medicinal product (IMP) reduced from 100 mg (2 capsules) q.d. to 50 mg (1 capsule) q.d. for the remainder of the study. Should the intolerance or toxicity continue, IMP may be discontinued, and the subject may be withdrawn from the study.

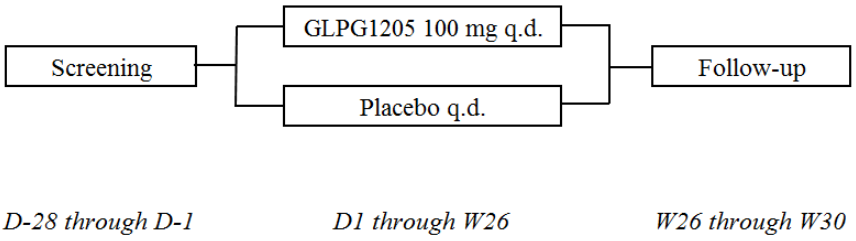
During the course of the study (post randomization) further adaptations of the treatment with nintedanib or pirfenidone due to e.g. intolerance or initiation of either nintedanib or pirfenidone is allowed based on the medical needs of the individual as judged by the investigator.

Enrolled subjects will come to the clinical study center at screening, Day 1 (baseline), Week 2, Week 4, Week 8, Week 12, Week 16, Week 20, Week 26, and, if applicable (i.e. for early IMP discontinuation), the end of treatment visit (EOT). In addition, a follow-up visit will be planned 4 weeks after the last administration of IMP (i.e. Week 30/EoS). The end of the study is reached when the last follow-up visit as planned according the Schedule of Activities (Section 2.5) of the last subject is performed.

Subjects who stop IMP early need to be encouraged to complete the EOT and follow-up visit for safety assessments and pulmonary function, but will not be obliged to do so.

To enhance the safety of the subjects and integrity of the study data, an IDMC will be implemented and clinical endpoint adjudication will be performed. Details are stipulated in the study protocol and the respective charters.

A schematic diagram of the clinical study design is provided below.



The subjects will be in the study for a duration of approximately 34 weeks: up to 4 weeks of screening, 26 weeks of treatment, and 4 weeks of follow-up.

2.4. Clinical Study Protocol (CSP) and CSP Amendments

This SAP is based on the protocol version 3.0 (amendment 2), dated 30-Apr-2020.

2.5. Schedule of Assessments

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

EVENT	**SCR**	**TREATMENT PERIOD**									**FU**
Study visit	[1]	[2]	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**EOT**	**10/ EOS**
Study week (W) or day (D) ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12 ±4d**	**W16 ±4d**	**W20 ±4d**	**W26[6] ±6d**		**W30[7] ±4d**
Informed consent	✓										
Dispense subject participation card	✓										
Send HRCT (historical or new[2]) for central review	✓										
Send LB (if available) for central review	✓										
Demographics	✓										
Medical history/concurrent illnesses	✓										
Alcohol consumption and smoking habits	✓										
Inclusion/exclusion criteria	✓	✓									
Physical examination	✓[3]	✓	✓	✓		✓		✓	✓[3]	✓[3]	✓[3]
12-Lead ECG	✓	✓	✓			✓		✓	✓	✓	✓
Vital signs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

[1] Screening assessments need to be performed within 28 days before Day 1 and may be split into multiple days or visits.
[2] If there is no historical HRCT available, a new HRCT needs to be planned and performed. Only if for logistic reasons it is not possible to perform a new HRCT within the planned screening period and after approval of the medical monitor it is allowed to extend the screening period with maximum 2 weeks.
[3] At screening include height and weight; at Visit 9/EOT and Visit 10/EOS include weight.

EVENT	SCR	TREATMENT PERIOD									FU
Study visit	[11]	[2]	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**EOT**	**10/ EOS**
Study week (W) or day (D) ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12 ±4d**	**W16 ±4d**	**W20 ±4d**	**W26[6] ±6d**		**W30[7] ±4d**
Spirometry	✓[4]	✓[4]		✓	✓	✓	✓	✓	✓[4]	✓	✓
SGRQ		✓				✓			✓	✓	
Clinical laboratory tests (hematology, coagulation, clinical chemistry, urinalysis)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Serology	✓										
FSH test	✓										
Pregnancy test	✓										
PK blood samples		✓		✓		✓		✓	✓		✓
[redacted]											
[redacted]											
DLCO	✓[redacted]	[redacted]									
6MWT	✓	✓							✓	✓	
[redacted]											
[redacted]											

[4] Spirometry can be repeated once if acceptability and repeatability criteria as specified by ATS/ERS/JRS/ALAT are not met after confirmation by central reader.

[5] [redacted]

[6] In case Visit 9 cannot be performed on site within the visit window due to the COVID-19 pandemic, the study medication intake and the corresponding time window for Visit 9 can be extended up to +56 days or shortened up to -28 days.

[7] In case Visit 10 cannot be performed on site within the visit window due to the COVID-19 pandemic, the corresponding time window for Visit 10 can be extended up to +28 days or shortened up to -7 days.

EVENT	SCR	TREATMENT PERIOD									FU
Study visit	[11]	[2]	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**EOT**	**10/ EOS**
Study week (W) or day (D) ± days (d)	**D-28 to D-1**	**D1**	**W2 ±2d**	**W4 ±2d**	**W8 ±4d**	**W12 ±4d**	**W16 ±4d**	**W20 ±4d**	**W26[6] ±6d**		**W30[7] ±4d**
Randomization		✓									
Dispense subject diary		✓	✓	✓	✓	✓	✓	✓			
Collect subject diary			✓	✓	✓	✓	✓	✓	✓	✓	
Dispense IMP		✓	✓	✓	✓	✓	✓	✓			
Collect returned IMP and review IMP compliance			✓	✓	✓	✓	✓	✓	✓	✓	
Dose IMP		daily throughout the treatment period									
Concomitant medication	throughout the study										
AE assessment	throughout the study										

6MWT=6-Minute Walk Test; AE=adverse event; D=Day; d=days; DLCO=diffusing capacity for the lungs for carbon monoxide; ECG=electrocardiogram; EOS=end of study; EOT=end of treatment visit; [redacted] FSH=follicle stimulating hormone; FU=follow-up; HRCT=high-resolution computed tomography; IMP=investigational medicinal product; [redacted] PK=pharmacokinetic; SAE=serious adverse event; SCR=screening; SGRQ=St George's Respiratory Questionnaire; LB= lung biopsy; W=Week

2.6. Sample Size Justification

See Section 7.1 of the CSP.

2.7. Randomization and Blinding

See Section 4.4 of the CSP.

It is acknowledged that the sponsor may be unblinded by a potential interim analysis. However, all efforts will be made to restrict the access to the study results so that personnel involved with the study conduct remain blinded, at least at an individual subject level.

3. GENERAL METHODOLOGY

3.1. Analysis Sets

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

3.1.1. All Screened Analysis Set

All subjects who signed an informed consent form (ICF) to participate in this study.

3.1.2. All Randomized Analysis Set

All screened subjects who were randomized in this study.

3.1.3. Full Analysis Set

All randomized subjects who have used at least 1 dose of IMP.

3.1.4. Per-Protocol Analysis Set

Subpopulation of the Full Analysis Set (FAS), excluding all subjects/data points with major protocol deviations impacting the efficacy results (determined during the blinded data review meeting). These protocol deviations are determined prior to unblinding and database lock and are entered in the database.

The Per-Protocol analysis is only planned if at least 10% of the subjects are omitted from the Per-Protocol Analysis Set.

If a subject is unblinded (e.g. in case of a serious adverse event [SAE]), all their efficacy data will be disqualified from the Per-Protocol Analysis Set.

3.1.5. Pharmacokinetic Analysis Set

Subset of the FAS, selecting all subjects who have available and evaluable PK plasma concentration data (e.g. excluding all protocol deviations or AEs that may have an impact on the PK analysis).

[redacted]

3.2. Randomized Versus Actual Treatment Group

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

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

Differences between the randomized and actual treatment group will be listed.

3.3. Analysis Periods and Analysis Time Points

3.3.1. Relative Day

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

When the concerned date is before the reference date:

Relative day (days) = concerned date-reference date

When the concerned date is the equal or later than the reference date:

Relative day (days) = concerned date-reference date +1 day

Where:

- The *concerned date* could be the measurement date of the assessment, or the start, or end date of the event.
- The *reference date* default is the date of the first dose of study drug administration, unless specified otherwise.
- *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.

The general terms of this formula also apply when similar relative timings are required in other time units, for example in minutes.

3.3.2. Analysis Periods

All event-type data (e.g. adverse events) and assessments will be allocated to analysis periods according to Table 1.

Table 1 Analysis Periods

Analysis Period	Start Analysis Period	End Analysis Period
Screening	Date of signing the ICF	Date of first study drug administration - 1 day
Treatment	Date of first study drug administration	Date of last study drug administration + 30 days
Follow-up[#]	End date of the preceding analysis period + 1 day	Date of last contact

ICF: informed consent form
[#] If the last study drug administration plus 30 days is after date of last contact, the study will not include a follow-up period. The last analysis period will end on the date of last contact.

If times are collected or other information with respect to timing (e.g. before or IMP intake) this information will be taken into account in addition to the dates. The last analysis period will always end on the date of last contact.

3.3.3. Analysis Windows

All assessments, including data collected on unscheduled visits, will be allocated to analysis windows based on the relative day of the assessment (see Section 3.3.1) and according to the algorithm in Table 2. TLFs will present the analysis windows.

Table 2 Analysis Windows for All Endpoints:

Time point label	Target day	Interval lower bound	Interval upper bound
Baseline	1	-INF	1
Week 2	15	2	22
Week 4	29	23	43
Week 8	57	44	71
Week 12	85	72	99
Week 16	113	100	127
Week 20	141	128	162
Week 26	183	163	197
Week 30	211	198	+INF

INF=infinity

Addendum: alternative approach due to COVID-19 pandemic

The Urgent Safety Measures (USMs dated 02-April-2020) described in CSP Amendment 2 allow to enlarge the visit window at the end of the treatment period (Visit 9, Week 26; from -28 days up to +56 days) including extension of study medication intake, i.e. up to a maximum of 56 days beyond 26 weeks of treatment. Hence, for patients having to extend treatment due to COVID-19, this could potentially result in end-of-treatment assessments being allocated to the FU/W30 analysis window and thus having missing data for the primary analysis time point W26. To assess the impact of this 'missing' data on the efficacy results, sensitivity analyses will be included in which these specific assessments are assigned to W26.

Definition of Baseline

For all endpoints, except ECG parameters and spirometry, baseline is defined as the last non-missing value before the first study drug administration. If multiple values qualify as last value, the mean of these values will be used in the analysis.

For spirometry endpoints, the mean of the analysis values (detailed in Section 5.2.2 for spirometry) before the first dose of study medication (including screening, baseline and unscheduled values as applicable) will be used to define the baseline value to reduce variability.

For electrocardiogram (ECG), the baseline is defined as the mean of the last recorded triplicate before the first study drug administration. If no triplicate is available before the first dose of study medication intake, the mean of the last duplicate ECG will be used, otherwise, the last ECG value will be considered as baseline.

Selection of Visits

Per parameter and 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, after the previous selection, multiple values still qualify, the mean of these values will be used in the analysis.

3.4. Handling of Data

3.4.1. Handling of Missing Data

3.4.1.1. 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.

3.4.1.2. Handling of Missing Result Data

No imputation will be done of missing result data. That is, an observed cases (OC) analysis will be performed. Details on the sensitivity analyses to investigate the impact of missing data on efficacy parameters are described in Section 6.2.

3.4.1.3. Censoring of Time-to-Event Data

Subjects who do not report the event at the moment of the analysis, are censored at the date of last contact. The date of last contact is defined as the latest date collected in the CRF from the date of death, visit or assessment dates, AE dates, concomitant treatment dates, drug intake dates or lab test dates and trial completion date.

3.4.2. Handling of Values Below or Above a Threshold

Values below (above) the detection limit will be imputed by the value one unit smaller (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".

3.4.3. Handling of Outliers

There will be no outlier detection, all measured values will be included in the analyses.

In the PK part of the analysis, exclusion of abnormal PK concentration results will be formally discussed and determined after data base lock and before running the final statistical analysis in the PK [redacted] Data Review Report.

3.4.4. Stratification Factors

The analyses will be stratified by the type of IPF SOC medication taken at randomization: nintedanib or 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.

The 3 strata could differ in their safety profile, in addition, they might have different effects on the biomarker results (depending on the biomarker) and efficacy results. Therefore, for all analyses, 3 strata will be used (nintedanib, pirfenidone, neither), unless indicated otherwise.

The stratum reported in the case report form (CRF) will be used in the analyses. Any discrepancies in stratification reported in the interactive web response system (IWRS) versus CRF will be listed.

3.5. Presentation of Results

3.5.1. Presentation of Treatment Groups

Results will be presented by treatment group:

- GLPG1205 100 mg
- Placebo

In the section Subject Information, a grand total "All Subjects" will be added to summarize all subjects over all treatment groups in tables and figures.

3.5.2. Calculation of Descriptive Statistics

Descriptive statistics will include:

- the number of non-missing data points (N)
- the arithmetic mean
- the standard error (SE) and standard deviation (SD)
- the median, minimum and maximum
- 95% confidence interval (CI) of the mean (if indicated in the relevant section).

3.5.3. Calculation of Percentages

Frequencies and percentages will be generated for categorical parameters.

For event-type data (e.g. adverse events), the denominator will be all subjects in the analysis set and analysis period. For other data (e.g. worst-case analysis of assessments), the denominator will be all subjects with (post-baseline) data for the parameter, in the analysis set and analysis window/period.

4. INTERIM ANALYSES AND DATA MONITORING COMMITTEE REVIEW

4.1. Independent Data Monitoring Committee

To protect the safety of the subjects in the study and integrity of the study data, an IDMC will be implemented. The IDMC will regularly review the accumulating unblinded safety data and assess the benefit risk of the study by further examining pulmonary function per treatment group as evaluated by spirometry and adjudicated clinical endpoints (if available). At each meeting, the IDMC will provide a recommendation to the sponsor on clinical study continuation, suggestions for clinical study adaptation(s) or early termination. No premature study termination is planned for favorable or unfavorable efficacy treatment effect. Specific details on responsibility, composition, meeting formats, and timing are outlined in an IDMC charter. Analysis for the IDMC are detailed in a separate SAP and will be provided to the IDMC by an independent, unblinded statistical support group.

4.2. Interim Analysis

Given the exploratory nature of the study, an interim analysis may be performed in order to allow sponsor planning of following clinical studies. No premature study termination is planned for favorable treatment effect reason based on this potential interim analysis. It is acknowledged that the sponsor may be unblinded by the potential interim analysis. However, all efforts will be made to restrict the access to the study results so that personnel involved with the study conduct remains blinded.

For the interim analysis, a subset of the final analyses will be provided:

- Subject disposition
- Demographics and baseline disease characteristics
- Listings for prior and concomitant therapies and summary table of change in SOC
- Exposure to study medication and compliance
- Primary and selected secondary efficacy assessments:
    - Change from baseline in FVC (mL)
    - Change from baseline in 6MWT
    - Change from baseline in SGRQ total score and domains, and proportion of SGRQ responders
- Listing of adjudicated events

- Safety:
    - Summary table for treatment-emergent AEs (TEAE) (by stratum)
    - Frequency tables by system organ class and preferred term for
        - TEAEs (by stratum)
        - TEAEs by severity
        - TEAEs of grade 3/4
        - Serious TEAEs
        - TEAEs related to GLPG1205/placebo
        - TEAEs related to SOC (pirfenidone or nintedanib)
    - Listings for
        - All AEs (summary)
        - AEs with severity grade 3 or 4
        - SAEs
        - AEs leading to death
        - AEs leading to respiratory-related hospitalization
        - TEAEs leading to discontinuation of GLPG1205/placebo
        - TEAEs leading to temporary stop or dose reduction of GLPG1205/placebo
        - AEs leading to a permanent stop, temporary stop, or dose reduction of SOC

- Descriptive statistics (and figures) of laboratory data, ECG and vital signs
- Frequency tables and shift tables of treatment-emergent laboratory toxicity grades and abnormalities, ECG abnormalities and vital signs abnormalities
- Abnormality listings for laboratory data, ECG, vital signs and physical examination

Outputs will be presented blinded at the IA, with the exception of demographics, baseline disease characteristics, exposure and efficacy output which will be presented unblinded at the group level. Further details of these analyses are provided in Section 5. PK analyses will be covered in a separate analysis plan.

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

5. STATISTICAL ANALYSES

5.1. Subject Information

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

Subject information will be tabulated with descriptive statistics per treatment group and overall.

5.1.1. Demographic and Baseline Disease Characteristics

Summaries will also be provided by stratum (3 levels; nintedanib, pirfenidone, or neither). The following parameters will be considered:

- Date of ICF signature (listed only)
- Sex
- Age at signing the ICF (years)
- Age, categorized (years):
  - $40 \leq$ age < 50
  - $50 \leq$ age < 60
  - $60 \leq$ age < 70
  - age ≥ 70
- Race and 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 will not be recalculated if already available in the database)
- BMI, categorized (kg/m²):
  - BMI ≤ 18.5

- $18.5 < \text{BMI} \leq 25.0$
- $25.0 < \text{BMI} \leq 30.0$
- > 30.0

– Baseline systolic and diastolic blood pressure and heart rate
– Baseline alcohol status: no, yes ("Does the patient drink alcohol?")
– Baseline smoking status: nonsmoker, current smoker, ex-smoker
– Duration of IPF (years) $= \frac{(\text{date of first intake of study drug}) - (\text{date of initial diagnosis})}{365.25}$. If the date of initial diagnosis is incomplete, then the following rules will be applied: Missing day: use the first of the month. Missing month: use January.
– Duration of IPF, categorized (years):
  - duration < 0.5
  - $0.5 \leq$ duration < 1
  - $1 \leq$ duration < 2
  - $2 \leq$ duration < 3
  - duration ≥ 3
– Baseline background SOC:
  - currently taking pirfenidone
  - currently taking nintedanib
  - never took pirfenidone or nintedanib
  - stopped taking pirfenidone and/or nintedanib
– Baseline spirometry:
  - 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 (assessing the psychosocial impact of the disease).
  - 6MWT distance
  - [redacted]
– DLCO (corrected for Hb)

5.1.2. Baseline Physical Examination

A frequency tabulation per body system of the baseline physical examination results, categorized as normal/abnormal, will be provided by treatment group and overall.

5.1.3. Disposition Information

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

- The number of subjects screened, randomized (treated and not treated), not-randomized and the number (percent) of subjects in each analysis set as defined in Section 3.1.
- The number (percent) of subjects who completed/discontinued the study/treatment and the reasons for discontinuation. This table will also be provided by stratum.
- The number (percent) of subjects per analysis window as defined in Section 3.3.3.
- Number (percent) of subjects randomized per country and investigator.

In addition, the following listings will also be provided:

- Subject identification and randomization (country, site number, investigator, subject number, randomization number, planned and actual randomization group and stratum, with flags for discrepancies).
- Subjects excluded from the FAS and the reason for exclusion.
- Date of the earliest ICF signed for this study and date of last scheduled visit performed in this study.

5.1.4. Protocol Deviations and Eligibility

Major protocol deviations are determined and recorded while the study is ongoing, and the list is finalized prior to database lock (and unblinding). For more details, please refer to the Protocol Deviations Plan.

The number (percent) of subjects with major protocol deviations will be tabulated, overall and per class of deviation, by treatment group and overall.

All available information concerning major protocol deviations, violations on eligibility criteria, and subjects not treated will be listed. The major protocol deviations leading to exclusion from the Per-Protocol Analysis Set (see Section 3.1.4) will be flagged.

5.1.5. Medical History and Concomitant Diseases

Frequency tabulations per system organ class and preferred term 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 concomitant diseases (i.e. conditions present at the start of the study).

5.1.6. Prior and Concomitant Therapies

5.1.6.1. Coding of Reported Terms

All prior and concomitant therapy terms will be coded in the database using the World Health Organization drug coding dictionary.

5.1.6.2. 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 (e.g. before, after, ongoing:

- Prior only: when the record ended before the first dose of IMP.
- Concomitant only: when the record started on or after the first dose of IMP.
- Prior and concomitant: when the record started before the first dose of IMP, and ended on or after this point, or continued.

Records without a start date are assumed to have started before the date of first study drug 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 (and no flags indicating relative timing are available), the date of first IMP administration will be considered to the same level of information provided by these incomplete dates to categorize the timing of these records. This means that a record only having month and year will be categorized comparing only to the month and the year of the date of first IMP administration. If the start date is unknown but the medication is flagged to have started prior to the first dose of IMP, this information will be taken into account.

5.1.6.3. Calculation of Relative Days

For both the start and the end dates of the concomitant therapy records, their day relative to the day of first study drug administration will be calculated as described in Section 4.3.1.

5.1.6.4. Presentation of Results

A frequency tabulation per treatment group and overall of the Anatomical Therapeutic Classification (ATC) classes level 4 by therapeutic subgroup (ATC level 2) and generic term of the prior medications (defined as ‘prior only’ and ‘prior and concomitant’) will be provided as well as of the concomitant medications (defined as ‘concomitant only’ and ‘prior and concomitant’).

Additional tables will be provided 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.

A frequency table per treatment group, summarizing the IPF 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).

In addition, a listing will be provided of the subjects who had any change in IPF 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.

5.1.7. Exposure to Study drug and Compliance

5.1.7.1. Derivation Rules

Derived Parameters: Extent of Exposure to Study Drug

- *Total treatment duration* (days) = last study drug administration date – first study drug administration date + 1 day.
- *Total treatment duration, excluding days off study drug*: Number of days with any study drug administration.

- *Total treatment duration, fully compliant* (days): Number of days with study drug administration of exactly 2 capsules.

Derived Parameters: Compliance

Study drug compliance is defined by the planned treatment dosage (100 mg q.d.; 2 tablets of 50 mg per day) without taking into account potential down-titrations (as allowed by the protocol).

- *Overall compliance* (%) $= 100 \times \frac{number\ of\ capsules\ actually\ used}{number\ of\ capsules\ that\ should\ have\ been\ used}$ (according to randomization)

- *Percent days with any intake* (%) $= 100 \times \frac{total\ treatment\ duration, excluding\ days\ off\ drug}{total\ treatment\ duration}$
- *Percent days fully compliance* (%) $= 100 \times \frac{total\ treatment\ duration, fully\ compliant}{total\ treatment\ duration}$

5.1.7.2. Presentation of Results

Summary statistics per treatment group and overall will be provided for each compliance and extent of exposure parameter. Frequency tables will be provided for the compliance parameters, using the following categories: <80%; 80% ≤ x < 100%; 100%; 100% < x ≥ 120%; > 120%.

A listing will be provided of the subjects who took incorrect study medication (GLPG1205 instead of placebo, or vice versa) for at least part of the treatment period, all GLPG1205/placebo records are to be included to provide an overview of the amount of cross-treatment relative to the total treatment period.

In addition, a listing of the subjects who had their study medication down-titrated or interrupted for at least part of the treatment period will be provided. All GLPG1205/placebo records are to be included to provide an overview of the amount of down-titration/interruption relative to the total treatment period.

Finally, a listing of the 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) will also be provided.

5.2. Efficacy Analyses

Efficacy analyses will be performed on the FAS, excluding all data collected after lung transplant. To investigate the impact of non-compliance and protocol violators a Per-Protocol analysis will also be performed for the primary endpoint, if at least 10% of the subjects are omitted from the Per-Protocol Analysis Set.

Tabulations will be shown per treatment group.

5.2.1. Level of Significance

Statistical tests will be done at a 2-sided significance level of 5%.

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

No multiplicity adjustment will be applied for the secondary endpoints, since this is a non-confirmatory proof-of-concept study.

5.2.2. Pulmonary Function by Spirometry

Pulmonary function will be assessed through spirometry at the study center. All spirometry evaluations should be performed pre-bronchodilator*. Spirometry data taken post-bronchodilator will be excluded from the analysis. Values will be used as provided by the central reading vendor.

* pre-bronchodilator is defined as:

- short-acting more than (>) 6 hours prior to the spirometry assessment
- long-acting bronchodilator at least (≥) 12 hours prior to the spirometry assessment

Note that a spirometry assessment with no use of bronchodilator is equivalent to a pre-bronchodilator assessment.

The data will be coded as "acceptable", "borderline acceptable" and "unacceptable". "Unacceptable" values will be excluded from analyses. The following quality grades as provided in the database will be used.

Table 3 Spirometry Quality Grades Based on Acceptability and Repeatability criteria

Quality	**Criteria**
Acceptable	≥ 3 efforts that meet the American Thoracic Society / European Respiratory Society (ATS/ERS) Guidelines (Raghu, Collard, Egan, & al, 2011; Raghu, Remy-Jardin, Myers, Richeldi, & Ryerson, 2018) and the criteria for Acceptability and Repeatability
Borderline acceptable	≥ 2 efforts that meet the ATS/ERS Guidelines and the criteria for Acceptability and Repeatability, or ≥ 3 efforts that meet the ATS/ERS Guidelines and the criteria for Acceptability but not Repeatability
Unacceptable	≤ 1 effort that meets the ATS/ERS Guidelines and the criteria for Acceptability and Repeatability
Acceptability:	

- Time to Peak Expiratory Flow < 250 ms
- Absence of artifact or cough
- Exhalation ≥ 6 seconds however demonstrated a 1 second plateau
- Lack of technical problems during effort

Repeatability:

- Will be based on the computer selected Best effort (highest) compared to next Best effort (second highest) based upon FVC values to be within 150 ml or within 100 ml when FVC ≤ 1L

Source: Pulmonary Function Test (PFT) OverRead Guidelines for GLPG1205-CL-220

The following parameters will be presented as part of the spirometry assessment:

- FVC (mL)
- Percent predicted FVC (%)
- FEV_1 (mL)
- Percent predicted FEV_1 (%)
- FEV_1/FVC ratio
- FEF_{25-75}

Derived values:

- Change from baseline:
  Change at time point t = (time point t value) – (baseline value)
- Percent change from baseline:
  Percent change at time point t = 100 x (change at time point t) / (baseline value)

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

5.2.2.1. Primary Endpoint: Change from Baseline in FVC over 26 Weeks

The change from baseline in FVC over 26 weeks will be compared between treatment groups by means of an analysis of covariance (ANCOVA) model at each planned time point (as specified in Section 2.5 and 3.3.3) including treatment, sex, and stratum as categorical covariates and age, height, and baseline FVC value as continuous covariates. LS means and 95% CIs will be presented per time point and treatment group as well as the LS mean differences between treatment groups with 95% CI.

Missing data will not be imputed and thus this analysis is conducted on OC only.

Percent predicted FVC will be analyzed using a similar ANCOVA model.

5.2.2.2. Sensitivity Analyses to the Primary Analysis

The following sensitivity analyses are planned to investigate the impact of missing data and protocol violators. 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.

5.2.2.2.1. Linear Mixed Model

A random coefficient regression model including sex, and stratum as categorical fixed effects and age, and height as continuous fixed effects and a random intercept and slope will be used to analyze the yearly rate of decline in FVC. The treatment effect will be determined by using estimated slopes for each study group on the basis of the time-by-treatment interaction term from the mixed model. The model will be fit using the SAS Proc Mixed procedure, details are in Appendix II. The assumptions of the model will be investigated graphically.

This approach provides another estimate for the primary endpoint. In addition, this model implicitly takes into account missing data under the assumption of missing-at-random (MAR).

5.2.2.2.2. Repeated Measures Analysis

The change from baseline in FVC over 26 weeks will also be compared between treatment groups by means of a mixed effect repeated measures model (MMRM) including sex, stratum and a treatment-by-time point interaction (planned time points as specified in Section 2.5 and 3.3.3) as categorical fixed effects, age, height, and baseline FVC as continuous fixed effects and correlated within-subject residuals. LS means and 95% CIs will be presented per time point and treatment group as well as the LS mean difference with 95% CI. The model will be fit using the SAS Proc Mixed procedure, details are in Appendix II.

This approach provides another estimate for the primary endpoint. In addition, this model implicitly takes into account MAR.

5.2.2.2.3. Per-Protocol Analysis

To investigate the impact of non-compliance and protocol violators a Per-Protocol analysis may also be performed for the primary endpoint using the ANCOVA model described in Section 5.2.2.1. Protocol violations potentially impacting the results will be excluded and identified before database lock and unblinding for the final analysis.

This Per-Protocol analysis is only planned if at least 10% of the subjects are omitted from the Per-Protocol Analysis Set.

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

5.2.2.2.4. Impact COVID-19

To investigate the impact on the primary endpoint of V9 assessments being assigned to the W30 analysis window due to COVID-19, a sensitivity analysis in which these assessments are assigned to the W26 analysis window will be performed for the descriptive statistics, the summary plots over time and the ANCOVA model described in Section 5.2.2.1.

5.2.3. Secondary Endpoints

5.2.3.1. Time to Major Events

A time to (first) event analysis will be conducted for major events, defined as

- Death (all-cause and respiratory-related)
- Hospitalization (all-cause and respiratory-related)

Deaths and 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 and proportion of subjects with an event during the study will be summarized per treatment group. Treatment effect on time to first event will be assessed using the log-rank test. In addition, time to first event will be graphically displayed and tabulated (at each planned time point as described in Section 3.3.3, Week 2, Week 4, etc, and last available time point) per treatment group using Kaplan-Meier estimates.

In addition, for all-cause hospitalization and all-cause mortality a frequency table will be provided for the event classification categories as determined by the CEAC. These categories are described in the CEAC charter.

Censoring

Time-to-event endpoints are defined from the date of first dose of study medication.

If the subject did not have an event, the time to event will be censored at the last contact date, as defined in Section 3.4.1.3.

5.2.3.2. Change from Baseline in 6MWT over 26 Weeks

The actual values, changes from baseline, and percent changes from baseline in 6MWT total distance walked in 6 minutes [redacted] will be presented descriptively for each time point by treatment group. Results will also be graphically displayed by means of subject profile plots, boxplots over time by treatment group for the actual values, and line plots of the mean (+/- SE) over time by treatment group for (percent) changes from baseline.

The change from baseline in 6MWT over 26 weeks will be compared between treatment groups by means of an ANCOVA model at Week 26 (as defined in Section 3.3.3) including treatment and stratum as categorical covariates and baseline 6MWT distance as continuous covariates. LS

means and 95% CIs will be presented per treatment group as well as the LS mean difference between treatment groups with 95% CI.

5.2.3.3. Change from Baseline in SGRQ over 26 Weeks

The St. George's Respiratory Questionnaire (SGRQ) is a 50-item paper questionnaire split into 3 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 4 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 2 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 3 steps

- The weights for all items 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.

  Sum of maximum possible weights for each score:

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

- 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

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

  The total score is calculated in a similar way:

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

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 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.

Source: St. George's Respiratory Questionnaire Manual (2009).

For actual values and changes from baseline in SGRQ total score, as well as symptom, activity, and impact scores, descriptive statistics will be provided per time point and treatment group. Results will also be graphically displayed by means of subject profile plots, boxplots over time

by treatment group for the actual values, and line plots of the mean (+/- SE) over time by treatment group for changes from baseline.

The change from baseline in SGRQ score (total, symptom, activity and impact) over 26 weeks will be compared between treatment groups by means of an ANCOVA model at each planned time point (as specified in Section 2.5 and 3.3.3) including treatment and stratum as categorical covariates and baseline SGRQ score as continuous covariates. LS means and 95% CIs will be presented per time point and treatment group as well as the LS mean differences between treatment groups with 95% CI.

The proportion of SGRQ responders at Week 26 (defined as absolute change from baseline in SGRQ total score ≤-4 %, at least once) will also be tabulated per treatment group. Treatment effect on SGRQ responders will be assessed using the Fisher's exact test.

5.2.4. Other Endpoints

[redacted]

[redacted]

5.2.5. Subgroup Analyses

Descriptive statistics of the spirometry parameters will be provided by stratum (3 strata; pirfenidone, nintedanib, neither). In addition, for the primary endpoint a similar ANCOVA model with an additional term for treatment-by-stratum interaction will be used to estimate the LS means per treatment group and LS mean differences (with 95% CIs) for each stratum.

[redacted]

Summaries may also be provided for the efficacy [redacted] endpoints [redacted] data will be available in time.

5.3. Safety Analyses

Safety analyses will be performed on the FAS.

Safety tables will be presented by treatment group.

Safety parameters will be analyzed descriptively (see Section 3.5.2 and Section 3.5.3). No formal testing will be performed to compare the treatment groups.

5.3.1. Adverse Events

All AEs and changes in attributes (worsening and improvement) of AEs are reported in the database. An identification number serves to link the records considered by the investigator as describing the evolution of one and the same event.

5.3.1.1. Definition of Treatment-Emergent Adverse Events

The analysis of AEs will be based on treatment-emergent events (TEAE).
TEAEs are defined as

AEs (or worsening[6] of an existing event) having a start date equal or after the date of the first administration of study drug and before the last dosing date plus 30 days. Improvements are not considered treatment-emergent.

5.3.1.2. Coding of Reported Terms

All AE terms will be coded in the database using the Medical Dictionary for Regulatory Activities (MedDRA) coding dictionary.

All tables will show the AE preferred terms grouped into system organ class. Subject listings will also show the reported terms. Any other coding levels will only be shown in a listing summarizing coding unless explicitly mentioned otherwise.

5.3.1.3. Allocation of Adverse Events to Analysis Periods

All AEs will be placed into analysis periods considering their start date, 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 and no flag indicating timing relative to study medication is available, a worst-case consideration (see below) will be done aiming to allocate the AE to one single analysis period, if possible. When a worst-case consideration is needed, the end date of the AE, if and as available, should also be considered; if such AEs clearly end 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 treatment period will only be placed in the treatment period.

[6] 'Worsening' is defined as worsening in at least one of the following attributes: seriousness, severity, relationship and/or action taken.

- An AE which according to the available information of its start date could belong to the treatment period as well as to the follow-up period will only be placed in the treatment period.
- An AE with a missing start date will be allocated to the treatment period.

5.3.1.4. Treatment Relatedness

Following the guideline ICH-E3 Structure and Content of Clinical Study Reports (Step 4 Version), the originally reported relatedness to study drug of an AE will be dichotomized as follows:

- *Not study drug related*: all non-missing weaker levels of relatedness than 'possibly drug related'.
- *Study drug related*: 'possibly drug related' and all stronger levels of relatedness (this class also includes any missing drug relatedness, as a worst-case consideration).

Only this dichotomized relatedness will be used in tables and can apply to different study drugs when relatedness has been collected separately per study drug; relatedness as originally reported will only be listed.

5.3.1.5. Worst-Case Selections

When cross-tabulating AE preferred terms versus an AE attribute (e.g. severity), only the worst-case within each analysis period will be considered, i.e. when a subject has more than once the same AE preferred term reported in the same analysis period, the subject will be counted only once and will be shown under the worst outcome (e.g. the worst severity for that AE in the concerned analysis period). If this happens in 2 different analysis periods, the AE is reported twice: once in each analysis period.

5.3.1.6. Calculation of Relative Days and Duration

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 study drug administration), the start day in the analysis period, and the duration (in days) will be calculated. In addition, the relative day and duration will be derived for the entire event; that is, the full evolution of the event, including the initial reporting and all subsequent worsening and/or improvement.

Relative days and durations will only be listed.

See Section 3.3.1 for the calculation of relative days.

5.3.1.7. Presentation of Results

The analysis will focus on AEs reported during the treatment period. AEs reported during the screening or follow-up period will only be listed.

All adverse events tables will show the number of subjects with TEAEs.

A summary table will be provided, showing the number (percent) of subjects with at least one:

- TEAE
- serious TEAE
- TEAE leading to death
- TEAEs by worst severity
- TEAE that was considered related to GLPG1205/placebo
- TEAE that was considered related to pirfenidone
- TEAE that was considered related to nintedanib
- TEAE for which GLPG1205/placebo was reduced
- TEAE for which GLPG1205/placebo was temporarily stopped
- TEAE for which GLPG1205/placebo was permanently stopped
- TEAE for which pirfenidone was reduced
- TEAE for which pirfenidone was temporarily stopped
- TEAE for which pirfenidone was permanently stopped
- TEAE for which nintedanib was reduced
- TEAE for which nintedanib was temporarily stopped
- TEAE for which nintedanib was permanently stopped

For each line, the difference in percentage of subject with an AE between placebo and GLPG1205 will be included, together with the 95% CI of the difference using the method of Miettinen and Nurminen.

Frequency tabulations, by system organ class and preferred term, of the number (percent) of subjects with a TEAE will be presented. Similar tables will be provided (by worst severity) for TEAEs of grade 3/4, TEAEs related to GLPG1205/placebo and for TEAEs related to SOC.

Serious adverse events (including AEs leading to hospitalization), AEs leading to death, leading to respiratory-related hospitalization (as reported by the investigator), of grade 3/4, leading to discontinuation of GLPG1205/placebo, leading to a temporary stop or dose reduction of GLPG1205/placebo, leading to a permanent stop, temporary stop, or dose reduction of SOC, pre- and post-treatment adverse events will be listed.

5.3.1.8. Adverse Events Related to Nintedanib or Pirfenidone

To investigate the events expected to be associated with pirfenidone and/or nintedanib a frequency table by system organ class and preferred term, by treatment group, as well as by stratum and treatment group, 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 final CSR.

5.3.1.9. EudraCT Adverse Events Reporting

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

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 all non-serious TEAEs.

5.3.1.10. Subgroup Analyses

The AE summary table, the frequency table by system organ class and preferred term, and the frequency table for events of specific interest will be provided for each stratum (3 strata; nintedanib, pirfenidone, neither).

5.3.2. Laboratory Safety

5.3.2.1. Available Data

Laboratory tests scheduled are described in the protocol. In addition, the derived laboratory tests described in Section 5.3.2.2 will be added.

The statistical analyses will only present results in Standard International (SI) units. Other units will not be presented.

Only data provided by the central laboratory will be used.

5.3.2.2. Derivation Rules

Derived Laboratory Tests

Creatinine clearance (C_{Cr}), estimated according to Cockcroft-Gault calculation:

> 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

For the calculation of estimated creatinine clearance, the age at screening and the weight, closest to the assessment for which creatinine clearance is calculated, will be used.

Fasted and Non-Fasted Results

Laboratory tests that are sensitive to the fasting status: glucose, triglycerides.

For these laboratory tests, 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.

5.3.2.3. Definition of Toxicity Grades

Toxicity grades will only be derived for laboratory tests specified in this section. For other tests, only the classification according to normal ranges will be used (Section 5.3.2.4).

Toxicity grades will be determined as implemented in the attached table (appendix) based on CTCAE, v5.0: November 27, 2017.

Note: when no explicit grade (e.g. "G1") is included, the categories will be labeled as indicated by the thresholds.

In addition, the following criteria will be used:

Table 5 Additional Toxicity Grading

Tests	Thresholds (Conventional Units)		Thresholds (SI Units)	
Hematology				
Eosinophiles relative count	>5%		>5%	
Coagulation				
Prothrombin time (PT)	>ULN - 1.5 x ULN >1.5 - 2.5 x ULN >2.5 x ULN		>ULN - 1.5 x ULN >1.5 - 2.5 x ULN >2.5 x ULN	
Chemistry				
Random glucose increased (fasting)	>130	mg/dl	> 7.2	mmol/L
High density lipoprotein (HDL)	<60	mg/dL	<1.554	mmol/L
Low density lipoprotein (LDL)	>160	mg/dL	>4.144	mmol/L

To investigate elevated liver function, the following additional grading (based on the exclusion and discontinuation criteria in the CSP) is defined.

Table 6 Additional Toxicity Grading for Liver Function

Tests	Thresholds
AST	AST >1.5 - <=3 x ULN AST >3 - <=5 x ULN AST >5 - <=8 x ULN AST >8 x ULN
ALT	ALT >1.5- <=3 x ULN ALT >3 - <=5 x ULN ALT >5 - <=8 x ULN ALT >8 x ULN

Tests	Thresholds
AST / ALT combination	ALT and/or AST >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
AST / ALT / total bilirubin combination	Bilirubin >2 x ULN AND AST or ALT >3 x ULN
AST / ALT / INR combination	INR >1.5 AND AST or ALT >3 x ULN
AST / ALT / eosinophiles relative count combination	Eosinophils >5% AND AST or ALT >3 x ULN

5.3.2.4. Definition of Non-Graded Abnormalities

For laboratory tests provided by the laboratory, the position of the actual analysis values versus their normal ranges will be determined directly by using the position indicator provided in the database as reported by the central laboratory, expressing the classes for these analysis values as low (L), normal (N) or high (H). L, N, and H are further referred to as non-graded abnormalities.

Only for tests for which no toxicity gradings are specified in Section 5.3.2.3, analysis of the classification according to normal ranges will be done.

5.3.2.5. Urinalysis Tests with Categorical Results

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

5.3.2.6. Treatment-Emergent Principle

Toxicity Grades

A post-baseline 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 post-baseline toxicity grade 1, 2, 3 or 4 will be considered as treatment-emergent.

Non-graded Abnormalities

A post-baseline non-graded abnormality class L or H is defined as treatment-emergent when it differs from the abnormality class of the baseline result. If the baseline result is missing, a post-baseline abnormality L or H will be considered as treatment-emergent.

5.3.2.7. Worst-Case Principle

Toxicity Grading

The worst-case post-baseline 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 post-baseline records (including unscheduled and follow-up visits, but excluding local lab results).

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

Non-graded Abnormalities

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

- 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.

5.3.2.8. Presentation of Results

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

Continuous laboratory tests will be summarized by means of descriptive statistics of the actual values and changes from baseline (including 95% CI of the mean change) by laboratory test, treatment group, and analysis window.

Boxplots of the actual values over time and of the mean (+- SE) change from baseline will be prepared for all laboratory tests.

The analysis of abnormalities will focus on assessments reported during the treatment period. Results reported during the screening or follow-up period will only be listed.

Abnormalities and toxicities grades of the actual values will be presented as shift tables of the worst-case abnormality/toxicity grade versus the baseline abnormality/toxicity grade. The table will be created per laboratory test and treatment group. The results of non-graded abnormalities and toxicities grades will be shown separately.

A frequency table of the number (percent) of subjects with treatment-emergent worst-case abnormalities/toxicity grade per laboratory test and treatment group will be presented. The results of non-graded abnormalities and toxicities grades will be shown separately.

A listing per treatment group, per subject, and per time point of all post-baseline time points scored as treatment-emergent and out-of-normal-range, plus also the baseline reference time point will be provided. In addition, laboratory parameters with insufficient data to be presented in the tables will be listed.

5.3.3. Electrocardiogram

5.3.3.1. Available Data

The following electrocardiogram (ECG) parameters will be analyzed: PR interval, heart rate (HR), RR interval, QRS duration, QT interval, and QTcF interval.

5.3.3.2. Derivation Rules

Derived Parameters

The QTcF values will be used as available in the database.

Handling of ECGs Measured in Triplicate

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

The mean of the triplicate/duplicate 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 triplicates/duplicates.

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

5.3.3.3. Abnormalities

The actual analysis values and changes from baseline of the QT and QTcF parameters will be categorized into the abnormality classes as defined in ICH-E14 The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs (Step 4 Version). For PR and HR the categorization as specified in the protocol is used.

Table 7 Abnormalities on ECG Parameters

Parameter	Abnormality	Limits
Abnormalities on actual values		
QT and QTcF (ms)	QT≤ 450 450 < QT ≤ 480 480 < QT ≤ 500 QT > 500	≤ 450 450 < value ≤ 480 480 < value ≤ 500 > 500
PR (ms)	L N H	< 120 120 ≤ value ≤ 220 > 220
HR (bpm)	L N H	< 50 50 ≤ value ≤ 100 > 100
Abnormalities on change from baseline		

QT and QTcF (ms)	QT change ≤ 30	≤ 30
	30 < QT change ≤ 60	30 < change ≤ 60
	QT change > 60	> 60

Worst-Case Abnormality

The worst-case post-baseline categorized actual analysis value and the worst-case categorized change from baseline 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).

For QT and QTcF, 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.

For PR and HR, the worst-case post-baseline abnormality L, N or H is determined as follows:

- 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 a table: once with a worst-case L and once with a worst-case H.

Treatment-Emergent Abnormalities

For QT and QTcF, an abnormal post-baseline category for actual values 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.

For PR and HR, post-baseline abnormalities L or H are defined as treatment-emergent when they differ from the abnormality class of the baseline result. When the baseline result is missing, post-baseline abnormalities L or H are considered as treatment-emergent.

5.3.3.4. Presentation of Results

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

Continuous parameters will be summarized by means of descriptive statistics (including 95% CI of the mean change) by parameter, treatment group, and analysis window. Actual values and changes from baseline will be tabulated separately.

Boxplots of the actual values over time and of the mean (+- SE) change from baseline will be prepared for all parameters.

The analysis of abnormalities will focus on assessments reported during the treatment period. Results reported during the screening or follow-up period will only be listed.

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 HR) and treatment group.

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

A frequency table per treatment group and time point of the ECG interpretation findings as transferred from the central reader may be provided.

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

5.3.4. Vital Signs

5.3.4.1. Available Data

The following vital signs parameters will be analyzed: heart rate, respiratory rate, diastolic and systolic blood pressure, oral or tympanic body temperature. If blood pressure is collected standing and supine, then results are presented by assessment position.

5.3.4.2. Abnormalities

Vital signs data will be evaluated for abnormalities based on the normal ranges as defined in Appendix 4 of the CSP.

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.

5.3.4.3. Treatment-Emergent Principle

A post-baseline abnormality class L or H is defined as treatment-emergent when it differs from the abnormality class at baseline. If the baseline result is missing, a post-baseline abnormality L or H will be considered as treatment-emergent.

5.3.4.4. 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.

5.3.4.5. Presentation of Results

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

Continuous parameters will be summarized by means of descriptive statistics (including 95% CI of the mean change) by parameter (and position if applicable), treatment group, and analysis window. Actual values and changes from baseline will be tabulated separately.

Boxplots of the actual values over time and of the mean (+- SE) change from baseline will be prepared for all parameters.

The analysis of abnormalities will focus on assessments reported during the treatment period. Results reported during the screening or follow-up period will only be listed.

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) and treatment group.

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

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

The vital signs measurements collected at the start and finish of the 6MWT will be listed separately.

5.3.5. Physical Examinations

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

5.4. Pharmacokinetic Assessments

Blood samples for the PK assessment of GLPG1205, pirfenidone and nintedanib should be collected as specified in the Schedule of Activities in section 2.5.

PK concentrations of GLPG1205, pirfenidone, and nintedanib will be described graphically. In addition, listings of all individual plasma concentration data will be created. [redacted]

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

[redacted]

Data after lung transplant will be excluded from all efficacy analyses.

[redacted]

In addition to descriptive statistics, ANCOVA models were included for the analysis of 6MWT distance, SGRQ score[redacted]

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

6. REFERENCES

Cockcroft DW, Gault MH. (1976). Prediction of creatinine clearance from serum creatinine. Nephron 16(1):31-41. PMID 1244564.

Fridericia, L. (1920). Die Systolendauer im Elektrokardiogramm bei normalen Menschen und bei Herzkranken. Acta Medica Scandinavia 53:469-486.

ICH E3 Structure and Content of Clinical Study Reports (Step 4 Version).

ICH E9 Statistical Principles for Clinical Trials (Step 4 Version).

ICH E14 The Clinical Evaluation of QT/QTc Interval Prolongation and Proarrhythmic Potential for Non-Antiarrhythmic Drugs (Step 4 Version).

ICH E6 Intergrated Addendum to ICH E6 (R1): Guideline for good clinical practice (Step 4 Version).

ICH E9(R1) Estimands and Sensitivity Analysis in Clinical Trials (Step 2 Version).

Jones, P.W. and Forde Y. (2009) St. George's Respiratory Questionnaire Manual. Version 2.3.

APPENDIX

APPENDIX I: LABORATORY: TOXICITY GRADING

Implementation of Common Terminology Criteria for Adverse Events (CTCAE), v5.0: November 27, 2017.

Notes:

- Laboratory results for which the fasting status is missing will be considered as taken non-fasted.
- Analysis results scoring below the lowest grade limits (< grade 1) are defined as grade 0.

Table 8 Laboratory Toxicity Grades based on CTCAE v5.0

Tests	Thresholds (Conventional Units)		Thresholds (SI Units)		Grade
Hematology					
Hemoglobin 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	>ULN - ULN + 2	g/dL	>ULN - ULN + 20	g/L	G1
	>ULN + 2 - ULN + 4	g/dL	>ULN + 20 - ULN + 40	g/L	G2
	>ULN + 4	g/dL	>ULN + 40	g/L	G3
White blood cell (WBC) count decreased	<LLN - 3000	/mm^3	<LLN - 3.0	* 10^9/L	G1
	<3000 - 2000	/mm^3	<3.0 - 2.0	* 10^9/L	G2
	<2000 - 1000	/mm^3	< 2.0 - 1.0	* 10^9/L	G3
	<1000	/mm^3	<1.0	* 10^9/L	G4
WBC count increased	>100000	/mm^3	>100.0	* 10^9/L	G3
Neutrophils decreased	<LLN - 1500	/mm^3	<LLN - 1.5	* 10^9/L	G1
	<1500 - 1000	/mm^3	<1.5 - 1.0	* 10^9/L	G2
	<1000 - 500	/mm^3	<1.0 - 0.5	* 10^9/L	G3
	<500	/mm^3	<0.5	* 10^9/L	G4
Lymphocytes decreased	<LLN - 800	/mm^3	<LLN - 0.800	* 10^9/L	G1
	<800 - 500	/mm^3	<0.8 - 0.5	* 10^9/L	G2
	<500 - 200	/mm^3	<0.5 - 0.2	* 10^9/L	G3
	<200	/mm^3	<0.2	* 10^9/L	G4
Lymphocytes increased	>4000	/mm^3	>4.0	* 10^9/L	G2
	>20000	/mm^3	>20.0	* 10^9/L	G3
Eosinophiles absolute count	>ULN		>ULN		G1
Platelets decreased	<LLN - 75000	/mm^3	<LLN - 75.0	* 10^9/L	G1
	<75000 - 50000	/mm^3	<75.0 – 50.0	* 10^9/L	G2
	<50000 - 25000	/mm^3	<50.0 - 25.0	* 10^9/L	G3
	<25000	/mm^3	<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
INR (international normalized ratio) increased	>1.2 - 1.5		>1.2 - 1.5		G1
	>1.5 - 2.5		>1.5 - 2.5		G2

	>2.5		>2.5		G3
Chemistry					
Random glucose decreased (fasting)	<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
Creatinine increased	>ULN - 1.5 x ULN		>ULN - 1.5 x ULN		G1
	>1.5 - 3x ULN		>1.5 - 3 x ULN		G2
	> 3 - 6 x ULN		> 3 - 6 x ULN		G3
	>6 x ULN		>6 x ULN		G4
eGFR (estimated glomerular filtration rate)/creatinine clearance ml/min/1.73 m^2	<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	mmol/L	<LLN - 3	mmol/L	G1
	<3 - 2.5	mmol/L	<3 - 2.5	mmol/L	G3
	<2.5	mmol/L	<2.5	mmol/L	G4
Potassium increased	>ULN - 5.5	mmol/L	>ULN - 5.5	mmol/L	G1
	>5.5 - 6	mmol/L	>5.5 - 6	mmol/L	G2
	>6 - 7	mmol/L	>6 - 7	mmol/L	G3
	>7	mmol/L	>7	mmol/L	G4
Calcium decreased	<LLN - 8	mg/dl	<LLN - 2.0	mmol/L	G1
	<8 - 7	mg/dl	<2 - 1.75	mmol/L	G2
	<7 - 6	mg/dl	<1.75 - 1.5	mmol/L	G3
	<6	mg/dl	<1.5	mmol/L	G4

Calcium increased	>ULN - 11.5	mg/dl	>ULN - 2.9	mmol/L	G1
	>11.5 - 12.5	mg/dl	>2.9 - 3.1	mmol/L	G2
	>12.5 - 13.5	mg/dl	>3.1 - 3.4	mmol/L	G3
	>13.5	mg/dl	>3.4	mmol/L	G4
AST (asparatate aminotransferase)	>ULN - 3 x ULN		>ULN - 3 x ULN		G1
	>3 - 5 x ULN		>3 - 5 x ULN		G2
	>5 - 20 x ULN		>5 - 20 x ULN		G3
	>20 x ULN		>20 x ULN		G4
ALT (alanine aminotransferase)	>ULN - 3 x ULN		>ULN - 3 x ULN		G1
	>3 - 5 x ULN		>3 - 5 x ULN		G2
	>5 - 20 x ULN		>5 - 20 x ULN		G3
	>20 x ULN		>20 x ULN		G4
GGT (gama-glutamyl transferase)	>ULN - 2.5 x ULN		>ULN - 2.5 x ULN		G1
	>2.5 - 5 x ULN		>2.5 - 5 x ULN		G2
	>5 - 20 x ULN		>5 - 20 x ULN		G3
	>20 x ULN		>20 x ULN		G4
Total bilirubin	>ULN - 1.5 x ULN		>ULN - 1.5 x ULN		G1
	>1.5 - 3 x ULN		>1.5 - 3 x ULN		G2
	>3 - 10 x ULN		>3 - 10 x ULN		G3
	>10 x ULN		>10 x ULN		G4
Alkaline phosphatase (aP)	>ULN - 2.5 x ULN		>ULN - 2.5 x ULN		G1
	>2.5 - 5 x ULN		>2.5 - 5 x ULN		G2
	>5 - 20 x ULN		>5 - 20 x ULN		G3
	>20 x ULN		>20 x ULN		G4
Albumin	<LLN - 3	g/dL	<LLN - 30	g/L	G1
	<3 - 2	g/dL	<30 - 20	g/L	G2
	<2	g/dL	<20	g/L	G3
Triglycerides (fasting)	>ULN - 500	mg/dL	>ULN - 5.7	mmol/L	G1
	>500 - 1000	mg/dL	>5.7 - 11.4	mmol/L	G3
	>1000	mg/dL	>11.4	mmol/L	G4
Cholesterol	>ULN - 400	mg/dL	>ULN - 10.34	mmol/L	G1
	>400 - 500	mg/dL	>10.34 - 12.92	mmol/L	G3
	>500	mg/dL	>12.92	mmol/L	G4
Creatine kinase (CK)	>ULN - 2.5 x ULN		>ULN - 2.5 x ULN		G1

	>2.5 - 5 x ULN	>2.5 - 5 x ULN	G2
	>5 - 10 x ULN	>5 - 10 x ULN	G3
	>10 x ULN	>10 x ULN	G4

G=grade; LLN=lower limit of normal; ULN=upper limit of normal; NA=not applicable

APPENDIX II: ADDITIONAL DETAILS ON MIXED MODEL PRIMARY ANALYSIS

Linear Mixed Model

As a sensitivity analysis for the primary analysis, the rate of decline in FVC over 6 months will be analyzed using a random coefficient regression model including sex, and stratum as categorical fixed effects and age, and height as continuous fixed effects and a random intercept and slope. The treatment effect will be determined by using estimated slopes for each study group on the basis of the time-by-treatment interaction term from the mixed model. Only post-baseline FVC values at planned time points, as specified in Section 2.5 and 3.3.3, will be used. The model will be fit using the SAS Proc Mixed procedure, details are below. The assumptions of the model will be investigated graphically.

The statistical model can be written as follows:

$$Y_{ij} = (\alpha + a_i) + (\theta + (\beta_T\theta) * Trt_i + g_i) * t_{ij} + \beta_S * Sex_i + \beta_A * Age_i + \beta_H * Height_i + \beta_{SOC} * (Standard\ of\ care)_i + \varepsilon_{ij}$$

Where

- Y_{ij} is the FVC value for subject i at time t_{ij},
- α and θ 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, with $\boldsymbol{b_i} = (a_i, g_i)$ assumed to be independent and identically normally distributed with mean 0 and unstructured variance-covariance matrix for all i,
- $(\beta_T\theta)$ is the effect of GLPG1205 on the slope,
- β_S, β_A, β_{SOC} and β_H are subject specific demographic coefficients for sex, age, stratification factor (SOC; 2 levels) and height at baseline,
- $Trt_i = 0$ for placebo and $Trt_i = 1$ for GLPG1205, $(Standard\ of\ care)_i = 0$ for neither and $(Standard\ of\ care)_i = 1$ for pirfenidone or nintedanib,
- ε_{ij} is the random error for subject i at time t_{ij}, assumed to be independent and identically normally distributed with mean 0 and variance σ_ϵ^2 for all i,j.

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 by the line plots 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 sex stratum;
  MODEL FVC = treatment * time sex age height stratum / solution CL ddfm=KR;
  RANDOM intercept time / type=UN subject = subject;
  REPEATED visit / type= Simple subject = subject;
  ESTIMATE 'GLPG1205 – Placebo' treatment*time -1 1 / e CL;
run;
```

Repeated Measures Analysis

As a sensitivity analysis for the primary analysis, the change from baseline in FVC over 26 weeks will also be compared between treatment groups by MMRM including sex, stratum, and a treatment-by-time point interaction as categorical fixed effects, age, height, and baseline FVC as continuous fixed effects and correlated within-subject residuals. Only post-baseline FVC values at planned time points, as specified in Section 2.5 and 3.3.3, will be used. The model will be fit using the SAS Proc Mixed procedure, details are below.

The statistical model can be written as follows:

$$Y_{ij} = \alpha + (\theta_j + (\beta_T\theta)_j * Trt_i) + \beta_S * Sex_i + \beta_A * Age_i + \beta_H * Height_i + \beta_{SOC} * (Standard\ of\ care)_i + \beta_B * Baseline_i + \varepsilon_{ij}$$

Where

- Y_{ij} is the FVC change from baseline for subject i at time point j (excluding baseline),
- α is the intercept and θ_j is the placebo time effect at time point j,
- $(\beta_T\theta)_j$ is the effect of GLPG1205 at time point j,
- β_S, β_A, β_{SOC} and β_H are subject specific demographic coefficients for sex, age, stratification factor (SOC; 2 levels) and height at baseline,

- β_B is the coefficient for baseline FVC value,
- $Trt_i = 0$ for placebo and $Trt_i = 1$ for GLPG1205, $(Standard\ of\ care)_i = 0$ for neither and $(Standard\ of\ care)_i = 1$ for pirfenidone or nintedanib,
- ε_{ij} is the random error for subject i at time point j, with $\boldsymbol{\varepsilon_i}$assumed to be independent and identically normally distributed with mean **0** and unstructured variance-covariance matrix, for all i.

If this model fails to converge, other structures will be tested for the variance-covariance matrix for the error terms (e.g. heterogenous Toeplitz, AR(1),…). The structure with the best fit, based on Akaike's information criterion, will then be used.

The following SAS code is to be used:

```
proc SORT data = FVC;
  BY subject avisit;
run;
proc MIXED data = FVC order = internal;
  CLASS subject treatment avisit sex stratum;
  MODEL FVC_change = treatment * avisit sex age height stratum FVC_baseline / solution CL
                                ddfm=KR;
  REPEATED avisit / type= UN subject = subject;
  LSMEANS treatment*avisit /cl diff;
run ;
```

[redacted]