CONFIDENTIAL

STATISTICAL ANALYSIS PLAN

A Randomised, Double-Blind, Placebo-Controlled Multicentre Clinical Trial of Inhaled Molgramostim

in Autoimmune Pulmonary AlveoLAr Proteinosis Patients

"IMPALA"

Sponsor Study Code: MOL-PAP-002

TFS Project Code: SDX004

Sponsor	Savara ApS.
Product/Compound	Molgramostim nebuliser solution
Phase of the study	II/III
Author	
Company	TFS
Address	Box 165, SE-221-00 Lund, Sweden
Telephone number	+46(0)46 280 18 00
Version	Final 1.0 **Date** 2019-04-25

SIGNATURES

Sponsor Study code: MOL-PAP-002

Author

Name, academic level, title: Jacob Råstam, BSc, Senior Statistician
Company TFS

Signature: DocuSigned by: Jacob Rastam
Signer Name: Jacob Rastam
Signing Reason: I approve this document
Signing Time: 4/29/2019 9:11:27 AM CEST
B4B74E6DAC454D7DA270E23E22BB22CE

Date/time (xxxx-xxx-xx): 29-Apr-2019 | 09:12 CEST
Add time to all dates if on the same day as unblinding.

TFS Reviewing Statistician

Name, academic level, title: Marie Sandgren, BSc, Principle Statistician
Company TFS

Signature: DocuSigned by: Marie Sandgren
Signer Name: Marie Sandgren
Signing Reason: I approve this document
Signing Time: 4/29/2019 9:14:57 AM CEST
B55182FB0E2442A8AD439D7E87A1380D

Date/time (xxxx-xxx-xx): 29-Apr-2019 | 09:15 CEST

Sponsor Reviewing Statistician

Name, academic level, title: Simon Day, PhD, Director
Company Clinical Trials Consulting and Training Ltd

Signature: DocuSigned by: Simon Day
Signer Name: Simon Day
Signing Reason: I approve this document
Signing Time: 29/04/2019 10:06:11 CEST
D3118036641B4CA8AC62AD7AB780CB54

Date/time (xxxx-xxx-xx): 29-Apr-2019 | 01:06 PDT

Sponsor Approval

Name, academic level, title: Cecilia Ganslandt, MD, VP, Head of Clinical Development
Company Savara ApS

Signature: DocuSigned by: Cecilia Ganslandt
Signer Name: Cecilia Ganslandt
Signing Reason: I approve this document
Signing Time: 4/29/2019 10:17:02 AM CEST
13A1B8C665164D6D8853C2941F36A3FF

Date/time (xxxx-xxx-xx): 29-Apr-2019 | 01:17 PDT

TABLE OF CONTENT

ABBREVIATIONS

6MWD	6 Minute Walk Distance
6MWT	6 Minute Walk Test
(A-a)DO_2	Alveolar-arterial Oxygen Difference
ADR	Adverse Drug Reaction
AE	Adverse Event
ALT	Alanine Aminotransferase
ANCOVA	Analysis of Covariance
anti-GM-CSF	Antibodies towards Granulocyte Macrophage Colony Stimulating Factor
aPAP	Autoimmune Pulmonary Alveolar Proteinosis
AST	Aspartate Aminotransferase
ATC	Anatomic Therapeutic Chemical
CEA	Carcinoembryonic Antigen
CI	Confidence Interval
CT	Computer Tomography
CTR	Clinical Trial Report
Cyfra 21-1	Cytokeratin 19 Fragment
DBL	Data Base Lock
DLCO	Diffusion Capacity of the Lung for Carbon Monoxide
DSMB	Data Safety Monitoring Board
DSS	Disease Severity Score
ECG	Electrocardiogram
eCRF	Electronic Case Report Form
EQ-5D-5L	EuroQol-5D 5 level quality of life questionnaire
FAS	Full Analysis Set
FEV_1	Forced Expiratory Volume in one second
FVC	Forced Vital Capacity
IMP	Investigational Medicinal Product
KL-6	Krebs von den Lungen-6
LDH	Lactate Dehydrogenase
LOCF	Last Observation Carried Forward
MA	Marketing Authorisation
MCAR	Missing Completely at Random
MAR	Missing at Random
MNAR	Missing Not at Random

MedDRA	Medical Dictionary for Regulatory Activites
MMRM	Mixed Model for Repeated Measures
PaO_2	Arterial oxygen tension
$PaCo_2$	Arterial partial pressure of carbon dioxide
PAP	Pulmonary Alveolar Proteinosis
PCV/EVF	Packed Cell Volume (PCV) or Erythrocyte Volume Fraction (EVF) also known as haematocrit
PPS	Per Protocol Set
PT	Preferred Term
QoL	Quality of Life
rhGM-CSF	Recombinant Human Granulocyte Macrophage Colony Stimulating Factor
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SGRQ	St Georges Respiratory Questionnaire
SOC	System Organ Class
SOP	Standard Operating Procedure
SP-A, SP-B, SP-C, SP-D	Surfactant Protein A, Surfactant Protein B, Surfactant Protein C, Surfactant Protein D
SpO_2	Oxygen saturation (indirect measurement)
VC	Vital Capacity
WHO	World Health Organisation
WLL	Whole Lung Lavage

1 INTRODUCTION

This Statistical Analysis Plan (SAP) is based on Global Clinical Study Protocol (CSP) Version 5.0, dated 01 December 2017, and Local Versions; 5.3 dated 04 April 2018, only applicable for USA, 5.4 dated 04 April 2018, only applicable for France and version 5.5 dated 08 May 2018, only applicable for Turkey, and in addition Amendment 2.0 dated 14 Dec 2015 and Amendment 3.0 dated 18 Feb 2016 (only applicable for Japan) to Protocol Version: 1.0 dated 15 Oct 2015.

2 STUDY OBJECTIVES

Primary objective:

- To compare efficacy of inhaled molgramostim on the Alveolar-arterial oxygen difference $(A\text{-}a)DO_2$ with placebo after 24-weeks treatment.

Key Secondary objectives:

- To compare efficacy of inhaled molgramostim on tolerance to exercise with placebo after 24-weeks of treatment
- To compare efficacy of inhaled molgramostim on respiratory disease-related quality of life with placebo after 24-weeks of treatment
- To compare efficacy of inhaled molgramostim based on time to Whole Lung Lavage (WLL) with placebo after 24-weeks treatment
- To compare safety of inhaled molgramostim with placebo in terms of reported adverse events (AEs), serious adverse events (SAEs), adverse drug reactions (ADRs), severe AEs and withdrawals due to AEs during 24-weeks treatment

Further Secondary objectives:

- To compare efficacy of inhaled molgramstim on the number of WLL over time with placebo after 24-weeks treatment (Added in SAP).
- To compare efficacy of inhaled molgramostim on Vital Capacity (VC), Diffusion Capacity of the Lung for Carbon Monoxide (DLCO), Forced Expiratory Volume in one (1) second (FEV_1), Forced Vital Capacity (FVC) and Arterial oxygen tension (PaO_2) with placebo after 24-weeks treatment
- To compare efficacy of inhaled molgramostim on the categorical change of $(A\text{-}a)DO_2$, VC, DLCO, FEV_1, FVC, and PaO_2 with placebo after 24-weeks treatment
- To compare efficacy of inhaled molgramostim on categorical change to exercise with placebo after 24-weeks treatment
- To compare efficacy of inhaled molgramostim on dyspnoea, and cough with placebo after 24-weeks treatment
- To compare efficacy of inhaled molgramostim on disease severity by Computer Tomography (CT) scoring with placebo after 24-weeks treatment

Exploratory objectives:

Double-blind treatment period

- To compare efficacy of inhaled molgramostim with placebo on $(A\text{-}a)DO_2$, VC, DLCO, FEV_1, FVC, PaO_2, and on tolerance to exercise after 4-weeks and 12-weeks treatment

- To compare the duration of response of inhaled molgramostim with placebo in (A-a)DO_2 and tolerance to exercise
- To compare efficacy of inhaled molgramostim with placebo on dyspnoea and cough after 4-weeks, and 12-weeks treatment
- To compare efficacy of inhaled molgramostim with placebo on Quality of Life (QoL) after 4 weeks, 12-weeks, and 24-weeks treatment
- To assess pharmacodynamic effects on selected biomarkers in serum after 4-weeks, 12-weeks, and 24-weeks treatment of inhaled molgramostim or placebo
- To assess the effect of molgramostim on Granulocyte Macrophage Colony Stimulating Factor (GM-CSF) concentration in serum after the first dose and after 4-weeks treatment, and to assess levels of antibodies to GM-CSF (anti-GM-CSF) after 4-weeks, 12-weeks and 24-weeks treatment with molgramostim or placebo.
- To assess the requirement for oxygen supplementation therapy during 24-weeks treatment with molgramostim or placebo.
- To assess the change in Disease Severity Score (DSS) from Screening to Week 24.

Follow-up period

- To compare the requirement for and time to WLL, or other treatment for aPAP after 24-weeks inhaled molgramostim or placebo during a 24-week, or 48-week post-treatment Follow-up period.
- To compare efficacy of 24-weeks inhaled molgramostim with placebo on (A-a)DO_2, VC, DLCO, FEV_1, FVC, PaO_2, and on tolerance to exercise during a 24-week or 48-week post-treatment Follow-up period.
- To compare safety of inhaled molgramostim with placebo during a 24-week or 48-week post-treatment Follow-up period in terms of reported AEs, severe AEs, SAEs, and ADRs.
- To assess levels of anti-GM-CSF at the 12-weeks and 24-weeks post-treatment Follow-up period
- To assess the change in DSS during a 24-week or 48-week post-treatment Follow-up period

2.1 Primary Efficacy Endpoint

Primary endpoint

- Absolute change from baseline of (A-a)DO_2 after 24-weeks treatment

2.2 Secondary Efficacy Endpoints

Key Secondary Endpoints:

- Change from baseline in 6-minute walking distance (6MWD) after 24-weeks treatment
- Change from baseline in SGRQ total score after 24-weeks treatment
- Time to WLL during 24-weeks treatment
- Number of AEs, SAEs, ADRs, severe AEs and AEs leading to treatment discontinuation, including clinically significant changes in laboratory tests and

echocardiographic (ECG) variables, during 24-weeks treatment (hereafter described under Safety Endpoints)

Further Secondary Endpoints:

- Number of WLL over time during 24-weeks treatment (Added in SAP).
- Absolute change from baseline in VC (% predicted), DLCO (% predicted), FEV_1 (% predicted), FVC (% predicted) and relative change from baseline in PaO_2 after 24-weeks treatment
- Number of subjects with >5 mmHg/>0.67 kPa and number of subjects with >10 mmHg/>1.33 kPa improvement in (A-a)DO_2 after 24-weeks treatment
- Number of subjects with >5 percentage points and number of subjects with >10 percentage points improvement in VC (% predicted) after 24-weeks treatment
- Number of subjects with >10 percentage points improvement in DLCO (% predicted) after 24 weeks treatment
- Number of subjects with >10 percentage points improvement in FEV_1(% predicted) after 24-weeks treatment
- Number of subjects with >10 percentage points improvement in FVC (% predicted) after 24-weeks treatment
- Number of subjects with >10% relative improvement in PaO_2 after 24-weeks treatment
- Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m or desaturation <4 percentage points on the 6MWT) after 24-weeks treatment
- Change from baseline in dyspnoea score and cough scores after 24-weeks treatment
- Number of subjects with improved CT score after 24-weeks treatment

Exploratory Endpoints

Double-blind treatment period

- Absolute change from baseline of (A-a)DO_2, VC (% predicted), DLCO (% predicted), FEV_1(% predicted), FVC (% predicted), and relative change from baseline in PaO_2 after 4 and 12-weeks treatment
- Time period during which the (A-a)DO_2 level is maintained below Baseline minus 10mmHg
- Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m or desaturation <4 percentage points on the 6MWT) after 4 and 12-weeks treatment
- Time period during which the improvement in tolerance to exercise is maintained
- Change from baseline in dyspnoea score and cough scores after 4 and 12-weeks treatment
- Number of subjects with improved QoL (reduction of ≥4 units on the SGRQ) after 4, 12, and 24-weeks treatment
- Number of subjects with ‘no problems’ in EQ-5D-5L after 4, 12, and 24-weeks treatment
- Change from baseline in serum concentration of biomarkers: Krebs von den Lungen-6 (KL-6), Carcinoembryonic antigen (CEA), Surfactant Protein A (SP-A), Surfactant

Protein B (SP-B), Surfactant Protein C (SP-C), Surfactant Protein D (SP-D), Cytokeratin 19 Fragment (Cyfra 21-1) and Lactate Dehydrogenase (LDH) after 4, 12, and 24-weeks treatment

- Anti Granulocyte Macrophage Colony Stimulating Factor (GM-CSF) antibody titers after 4, 12 and 24-weeks treatment
- Change in serum concentration of GM-CSF post first dose of trial drug and after 4-weeks of treatment
- Number of subjects in need for oxygen supplement therapy during 24-weeks treatment.
- The distribution of DSS at Screening and at Week 24
- The percentage of subjects with DSS 1 or 2 at Screening and at Week 24.

Follow-up period

- Number of subjects requiring WLL or other treatment for aPAP and number of treatment courses required during 48-weeks follow-up.
- Time to WLL, or other treatment for aPAP during 48-weeks follow-up.
- Absolute change from baseline in (A-a)DO_2 and VC (% predicted), DLCO (% predicted), FEV_1(% predicted) , FVC (% predicted), and and relative change from baseline of PaO_2 after 12, 24, 36 and 48-weeks follow-up
- Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m or desaturation <4 percentage points on the 6MWT) after 12, 24, 36 and 48-weeks follow-up
- Number of AEs, severe AEs, SAEs and ADRs , including clinically significant changes in laboratory tests and ECG variables, during 48-weeks follow- up
- Levels of anti-GM-CSF after 12 and 24 weeks follow-up.
- The distribution of DSS after 24 and 48-weeks follow-up
- The percentage of subjects with DSS 1 or 2 after 24 and 48-weeks follow-up.

2.3 Additional Safety Endpoints

The following safety variables will be assessed:

- Physical examination during 24-weeks of treatment and 48-weeks follow-up
- Electrocardiogram during 24-weeks of treatment and 48-weeks follow-up
- Vital signs during 24-weeks of treatment and 48-weeks follow-up
- Laboratory safety assessments during 24-weeks of treatment and 48-weeks follow-up

3 OVERALL STUDY DESIGN

3.1 Overview of Study Design

This is a parallel, double-blind, multicentre, randomised clinical trial in autoimmune pulmonary alveolar proteinosis subjects, where a planned total of 135 subjects (see section 3.2) are stratified according to whether or not a WLL has been conducted within 2 months prior to the Baseline (according to protocol version 4.0 and later versions), or whether or not

WLL was conducted between the screening visit and the randomization visit (according to protocol versions before version 4.0). Subjects will be randomised 1:1:1 to receive treatment for up to 24 weeks with either;

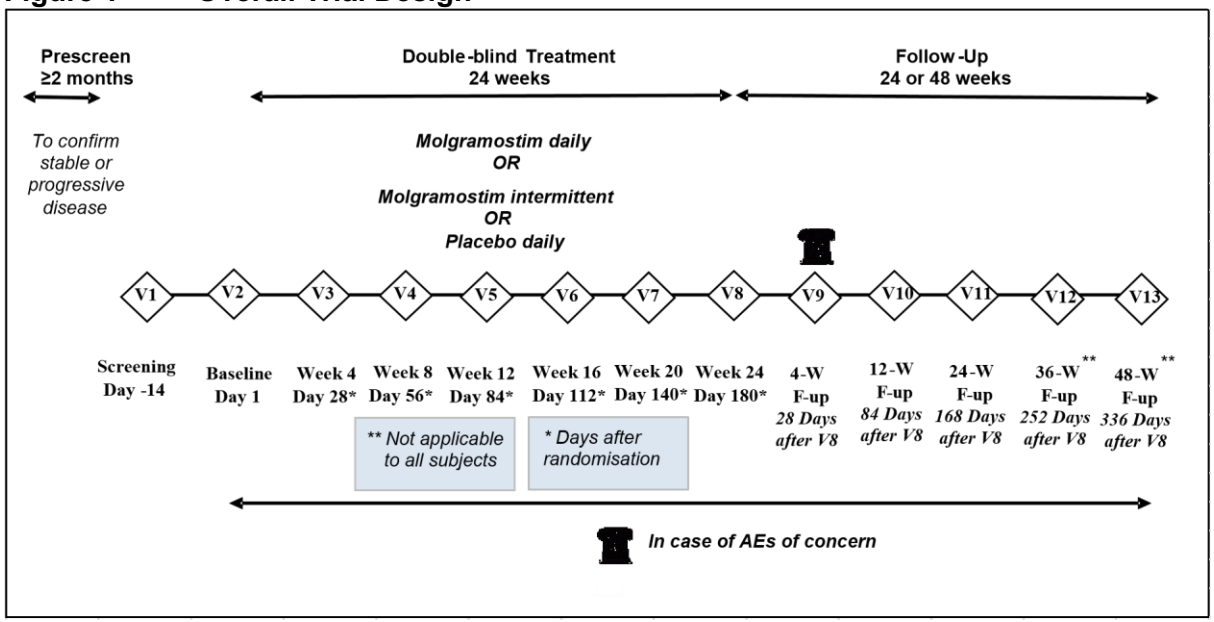
1) Inhaled molgramostim (300 µg) administered once daily,

2) Inhaled molgramostim (300 µg) and matching placebo administered intermittently (12 cycles of seven days molgramostim, seven days placebo; both administered once daily) or

3) Inhaled placebo administered once daily.

The trial includes two phases; a Double-blind treatment period consisting of up to eight trial visits (Screening, Baseline, and at Weeks 4, 8, 12, 16, 20 and 24 after randomisation) and a Follow-up period consisting of up to five trial visits (at Weeks 4, 12, 24, 36 and 48 post-treatment). During the follow-up period all subjects recruited under protocol version 4.0 or later versions will receive treatment with inhaled molgramostim (300 µg) administered once daily in an intermittent regimen for 24 weeks (12 cycles of seven days molgramostim, seven days off). Patients recruited under protocol version 2.0 were only to receive treatment in the follow-up period in case of worsening of aPAP, and patients recruited under protocol version 3.0 were treated (if needed) using the intermittent regimen for 24 weeks and according to investigator's discretion for 24 additional weeks. Likewise, the 36 and 48-weeks post treatment visits are only applicable for subjects included before approval of protocol version 4.0.

Figure 1 Overall Trial Design



3.2 Determination of Sample Size

The trial is a randomised 3-arm trial in which two thirds of the subjects receive active drug and one third receive placebo.

The sample size calculation was based on data in an earlier trial, where the mean (and standard deviation) (A-a)DO_2 were 31.3 (7.4) before treatment and 12.9 (7.6) after treatment

(all measured in mmHg). The minimum clinically relevant difference to detect between active and placebo arms is considered to be 10-12 mmHg. In order to ensure such an effect size can be detected, a delta of 10 mmHg has been used in the sample size calculation which is based on an unpaired *t*-test of mean difference between the two active arms (combined) vs. placebo, using a significance level of 0.01 and a power of 90%.

The trial will test the null hypothesis

H_0: $\mu_{diff} = 0$

Versus the alternative hypothesis

H_1: $\mu_{diff} \neq 0$.

The calculation was based on an unpaired *t*-test using nQuery Advisor® 7.0. In versions of the protocol including Version 4 and earlier, it was intended that an analysis of the two active dose groups combined, versus placebo, would be carried out, hence the sample size calculation was made based on a 2:1 allocation.

Table 1 Two Group T-Test of Equal Means (Unequal n's)

Test significance level, α	0.010
1 or 2-sided test?	2
Difference in means, $\mu_1 - \mu_2$	10.000
Common standard deviation, σ	7.500
Effect size, $\delta = \lvert\mu_1 - \mu_2\rvert / \sigma$	1.333
Power (%)	90
n_1	14
n_2	28
Ratio: n_2 / n_1	2.000
$N = n_1 + n_2$	42

Based on the assumption listed in Table 1, above, 42 subjects would initially be required to be randomised.

Based on a subsequent evaluation of variability and plausible effect sizes for the key secondary endpoints, the sample size was increased to 90 subjects (30 subjects in each treatment arm) in Substantial Protocol Amendment 11 (06 June 2017), to increase the power to identify statistically significant treatment effects also on one or more of the key secondary endpoints.

Very limited data on these secondary endpoints are available but with a sample size of 30 subjects per group, it would be possible to show a treatment effect of 50 m on the 6MWD (assuming a SD of 50) or of 10 points on the SGRQ (assuming a SD of 10) with approximately 90% power.

Because of the limited data available, a fully blinded sample size re-estimation procedure was carried out to assess the standard deviations of the 6MWD and the SGRQ, and the overall event rate of WLL. This was planned to be done in January 2018 when approximately 50 patients were expected to have reached 24 weeks of treatment.

The change from baseline to week 24 in 6MWD [and SGRQ] was analysed in an ANCOVA model including WLL (stratification) and baseline 6MWD [SGRQ] in order to estimate the SD. For the purposes of the sample size re-estimation process, only data on patients who had baseline and week 24 data was used, and no imputations were made for missing values. Additionally, the event rate for WLL was determined including any patients who met the criteria for WLL any time prior to week 24, but excluding any patients who withdrew prior to week 24 where WLL had not been required.

Based on a literature review of different pulmonary indications but with main focus on Idiopathic Pulmonary Fibrosis carried out in November 2017, the minimal clinically important effect sizes for 6MWD and SGRQ in aPAP were considered to be in the order of 40-50 m and 7-10 points, respectively. From each ANCOVA model, the respective residual variance for 6MWD [and SGRQ] were used to calculate sample size based on the effect sizes of 50 m and 10 points, respectively. Similarly, the sample size for comparing WLL rates of 5% (active) versus 20% (placebo) at 24 weeks was evaluated, based on the observed (pooled) WLL rate. All calculations were made based on a 2-sided 5% significance levels, and 80% and 90% power.

If the maximum of these sample size calculations were to be no more than 33 patients per group, then no change to the target sample size would be made. Otherwise, the sample size would be increased to attain 90% power for at least 2 of the 3 key secondary endpoints, subject to the total study size not exceeding 150 patients.

From these sample size calculations conducted in January 2018 it was clear that the two key secondary endpoints with highest power to detect a clinically relevant difference between treatment groups based on the least number of subjects were SGRQ Total Score and 6MWD. Based on the pre-defined criterion that the sample size would be increased to attain 90% power for at least 2 of the 3 key secondary endpoints, the results indicate that the critical endpoint is the 6MWD, which would require that the sample size should be increased to at least 43x3=129 randomised subjects. After reviewing the results from the sample-size recalculation it was decided by Savara ApS to increase the sample size to 45x3=135 randomised subjects. Any subjects already in screening at the time of reaching 135 randomised subjects would also be randomised, if otherwise eligible for the trial.

The detailed results of the sample size re-estimation were documented in a separate report and are available in Appendix 2 to the SAP.

4 DATA SETS TO BE ANALYSED

The following analysis sets will be used for the statistical analysis and presentation of data during the DB (Double Blind) period:

Safety set (DB-SAF): All randomised subjects who received at least one dose of the trial drug, will be analysed according to actual treatment.

Full analysis set (DB-FAS): All randomised subjects, will be analysed according to randomised treatment.

Per-protocol set (DB-PPS): All randomised subjects who have completed Visit 8 and are deemed to have no important protocol deviations that could interfere with the objectives of this trial, will be analysed according to actual treatment.

Important deviations of eligibility criteria and other deviations from the protocol will be assessed by TFS in cooperation with the sponsor. Important deviations from the protocol may lead to exclusion of a subject from the DB-PPS. All such decisions will be made and documented before the trial database is unblinded.

If not otherwise stated, all efficacy endpoints will be analysed based on DB-FAS. Although analyses of primary and key efficacy endpoints will be performed on both the DB-FAS and the DB-PPS, the DB-FAS will be considered the primary analysis population.

Safety summaries will be performed on the safety set (DB-SAF).

Analysis sets to be applied for the follow-up period will correspond to the analysis sets for the double-blind period, although the follow-up period analysis sets will include only subjects not withdrawing from the trial before the follow-up period begins and who has at least one follow-up period data-point. The follow-up period FAS and SAF will be denoted FU-FAS and FU-SAF, respectively. In addition, the subgroup of subjects receiving the intermittent dosing regimen during the follow-up period (i.e. the vast majority of follow-up period subjects) will be evaluated alone (FU-MOL-INT).

5 STATISTICAL AND ANALYTICAL PLANS

The planned tables, figures and listings are presented in Appendix 1.

5.1 Changes in the Planned Analyses

This SAP is based on protocol versions and amendments described in section 1. Compared to these, the following changes have been implemented in the SAP:

- All model-based analyses will include adjustment for a dicotomeous categorization of country as Japan vs other countries.
- A number of additional subgroup analyses were added, as described in section 5.8.
- An additional sensitivity analysis for the primary endpoint was added for handling of subjects who had a WLL (section 5.3.6)
- Number of WLLs over time, was included as a further secondary endpoint in the double-blind period, as described in section 5.3.7.

5.2 Blind Review

The trial will be unblinded and the primary analysis will be conducted after all subjects have completed the 24-week Double-blind period. Data will be reviewed on an ongoing basis before the database is locked and unblinded. Clarifications of specific data decisions, having an impact on the planned analyses, which are identified in connection to the blind review, after finalization of the SAP, will be included in a separate “pre-analysis review” document, which will be signed before unblinding. Final classification of subjects according to study populations will be also documented and included in the pre-analysis-review document.

In this trial, analyses related to the double-blind (DB) treatment period will be reported after last subject has completed the treatment period, thus a partial DB data base lock will be held and all data up to that point will be cleaned and locked before unblinding. Further data review and cleaning, related to the open label follow-up period, will be done and any new decisions will be documented in a separate pre-analysis review document before a final complete FU data base lock. Locked data for the double-blind period cannot be changed after the partial data base lock, except for any ongoing safety evaluations that have not yet been resolved at the time point of the partial DB data base lock.

5.3 Hypotheses and Statistical Methods

5.3.1 General

For all efficacy analyses, primary and secondary, where statistical tests are conducted (see separate sections below) the null hypothesis is that there is no treatment difference, which

will be tested against the alternative hypothesis, that there is a treatment difference. All tests will be two-sided.

In general, all summary statistics will be presented for all treatment groups.

In this trial, for subjects withdrawing from treatment early, attempts will be made to continue follow-up and collect the planned data, as described in protocol. If a subject is withdrawn from the trial early, the investigator should attempt to complete all required trial assessments (such as those at Week 24 if withdrawn during the Double-blind treatment period). Assessments for subjects withdrawing from trial early, and where any measurements related to endpoints to be statistically analysed are collected on the Week 24 Visit in the eCRF, (i.e. if intermittent visits are missing), then these assessments will be reallocated to the visits closest in time to the time point of withdrawal. This is done in order to avoid an unintended partial LOCF approach to analysis of data.

For statistical analyses of the primary and key secondary efficacy endpoints, the type-1 error rate will be controlled, as described in section 5.4. In addition a number of sensitivity analyses will be conducted, as described in the following sections. An overview of the planned statistical analyses and how inflation of alpha, due to multiple testing, will be prevented can be seen in figure 2.

Results applicable to the follow-up period will be summarised both according to randomised treatment arms under the double-blind treatment period and in total.

5.3.2 Definitions

Baseline	Baseline measurements refers to the last non-missing assessment made before randomisation. If a measurement is intended to be collected at the baseline visit, but is missing, then a corresponding measurement taken at the screening visit may be used to replace the missing assessment at the baseline visit.
Relative day	The relative day of an event is derived as follows: Relative day = (Start date)-(Date of first administration of IMP)+1. For days before the start date, calculate as Relative day = (Start date)-(Date of first administration of IMP). In this way, there will be no day=0. So, Day 1 is the same day as the day of first administration of IMP, and Day -1 is the day before.
(A-a)DO_2	Alveolar-arterial oxygen difference is a measure of the difference between the alveolar concentration (A) of oxygen and the arterial (a) concentration of oxygen
DLCO	Uptake of carbon monoxide from a single inspiration in standard time.
FEV_1	The volume of air that can forcibly be blown out in one second, after full inspiration.
FVC	The maximum amount of air a person can expel from the lungs after a maximal inhalation and during a forceful expiration.
SpO_2	Indirect measurement of oxygen saturation using a finger probe, ear sensor or similar device.
6MWT	Test used to measure the distance that a subject can walk quickly on a flat hard surface in a period of 6 minutes (at an individually predetermined oxygen supplementation).
VC	The maximum amount of air a person can slowly expel from the lungs after a maximum slow inhalation.
% predicted	Calculation of % predicted of spirometry parameters collected during the double-blind period will be made centrally for statistical analysis using the same equation for all sites.
Region	For summaries and figures split by region ‘Europe’ will include: Denmark, France, Germany, Greece, Italy, Netherlands, Portugal, Russia, Slovakia, Spain, Switzerland and United Kingdom. ‘Other’ will include: Australia, Israel, South Korea and Turkey while Japan and the United States will be defined as their own regions.

5.3.3 Summary Statistics

Data will be summarised by means of summary statistics. For continuous data the following summary statistics will be presented: number of observations, mean value, standard deviation, minimum, first quartile, median, third quartile and maximum value. Categorical data will be presented as counts and percentages. The data will be presented by visit. Baseline, demographics, primary endpoint, and key secondary endpoints will also be summarised by region.

Summary statistics will be presented by treatment group and assessment time and/or visit, as applicable.

5.3.4 Subject Data Listings

Data collected in the eCRF, as well as any derived assessments used for statistical analyses, will generally be listed in Appendix 16.2 (see section 7.2). eCRF check questions [e.g. lab samples taken (Yes/No)] and reminders may not be listed. Listings will primarily be sorted by centre, subject-number and visit and/or any other natural ordering related to the applicable assessment.

In CRF modules where a date is recorded, the date and the relative day (in relation to date of randomisation) will be printed in the corresponding listing. In modules where both a start date and stop date is recorded, also a duration will be included in the listings.

5.3.5 Demographic and other Baseline Characteristics

A subject disposition will be produced, including number of subjects screened, randomised, withdrawn from trial (with reason for withdrawal), number of subjects withdrawing from treatment but remaining in trial, number of subjects completing double-blind treatment period and follow-up, number of subjects included in the safety analysis, full analysis and per protocol sets. Demographics, standard baseline characteristics and aPAP history data will be presented using summary statistics. Data that is measured both before baseline and at other visits following randomisation will in general be presented in other sections in by visit displays.

5.3.6 Primary Efficacy analysis

The primary efficacy variable, (A-a)DO_2, will be collected at Screening, Baseline (randomisation), Week 4, 12 and 24 during treatment period, and at Week 12, 24, 36 and 48 during the follow-up period. All (A-a)DO_2 values reported in the CRF will be recalculated for statistical analyses using the equation:

$$Aa\ Gradient = \left(F_iO_2(P_{atm} - P_{H_2O}) - \frac{P_aCO_2}{0.8} \right) - P_aO_2$$

where F_iO_2 is fraction of inspired oxygen

P_{atm} is ambient atmospheric pressure

P_{H_2O} is saturated vapour pressure of water at body temperature (set to 47 mmHg / 6.266 kPa)

P_aCO_2 is arterial partial pressure of carbon dioxide

P_aO_2 is arterial partial pressure of oxygen

Any of these assessments reported in kPa will be converted to a result in mmHg before the calculations, this will be done using a factor k=7.5.

The primary efficacy endpoint, absolute change from baseline of (A-a)DO_2 after 24-weeks treatment, will be analysed based on an ANCOVA model including treatment group, WLL (stratification), Japan vs other countries, and baseline (A-a)DO_2. The data from the three treatment groups will be analysed all together (all treatment groups included in one model). Contrasts will be set up to estimate treatment differences in change from baseline with 95% confidence intervals (CI) and the corresponding p-value for Molgramostim OD vs. Placebo and for Molgramostim intermittently vs. Placebo.

For each analysis two sided tests will be performed where the null hypothesis is that there is no difference between treatment groups, which will be tested against the alternative that

there is a difference. Control for type-1 error inflation due to multiple testing will be handled as described in section 5.4.

A de-facto estimand approach will be applied, where for the primary analysis, all observed data will be included, irrespective of subject adherence to the randomised treatment. Subjects who withdraw from their randomised treatment before week 24 will be followed up through to week 24, and their week 24 data will be considered as primary, irrespective of treatment discontinuation, treatment interruption(s), use of rescue medication (WLL or other), etc.

All assessed (A-a)DO_2 measurements, and the corresponding changes from baseline will be summarised by visit and treatment group. In addition, similar summaries of (A-a)DO_2 will also be produced excluding data collected during treatment discontinuation, treatment interruption(s) and use of rescue medication (WLL or other). The decision of what data will be excluded and the reason will be identified during the blind review of data and documented in the “Pre-analysis-review”-document before data base lock (DBL).

Mean profiles for (A-a)DO_2 observed values and change from baseline over time will be provided.

Sensitivity Analyses

The following sensitivity analyses for the primary endpoint will be done in order to evaluate the robustness of the results in relation to influence of any missing data. The sensitivity analyses should be considered supportive analyses and will thus not be part of the procedures applied to control type-1 error rate for testing hypotheses for primary and key secondary efficacy endopints as described in section 5.4.

- A mixed model for repeated measures (MMRM) will be applied including fixed, categorical effects for treatment, WLL (stratification), visit, treatment-by-visit interaction, Japan vs other countries, and fixed covariates for baseline (A-a)DO_2. An unstructured covariance matrix will be applied to model the within-subject errors, if possible. If this analysis fails to converge, compound symmetry will be tested. Kenward-Roger approximation will be used to estimate denominator degrees of freedom. This methodology does not require imputation of missing data. Instead missing data is modelled based on the subjects’ available data and on other subjects’ data over time (this approach is valid under assumption that data is missing according to ‘Missing at Random’ (MAR)) [2].
- Subjects who withdraw from the study early and who do not have week 24 data available will have their (A-a)DO_2 measurement imputed using multiple imputation (MI). This will be done by first ensuring a monotone missing data pattern by imputation of intermediate missing data (baseline, Week 4 and Week 12) by a Markov chain Monte Carlo (MCMC) approach, where imputations are done by treatment group Japan vs other countries, and WLL (stratification) if possible (strata may be dropped depending on the number of observations per combination). The imputation will be done generating 10 data sets. When monotone missingness has been accomplished, then the remaining missing data, up to and including Week 24, will be imputed using a monotone regression approach using treatment and WLL (stratification) as class variables. Each replication of the data, including the imputed missing values, will then be analysed using the ANCOVA used in the primary analyisis, and estimates will be pooled using proc MIANALYZE SAS®. (this approach is valid under assumption that data is missing according to MAR).

- If superiority has been determined for any of the comparisons planned for the primary endpoint, a tipping point analysis (TP) will be applied where subjects with a missing endpoint (Week 24) in the group for which superiority has been determined will be assigned progressively worse (A-a)DO_2. This will be done by generating imputed data according to the previous step and using the MNAR option in proc MI SAS® to adjust imputed Week 24 data with an appropriate sized shift paramter in the superior arm in an iterative manner until the results shift from significant (P<0.05) to non-significant (P>0.05). From this analysis, results will be presented including overall estimates at the tipping point, as well as descriptive statistics for imputed data vs non-imputed data.
- A “reasonably worst case” scenario. For the primary analysis, patients in either of the active arms will have the value of the 10th percentile from their dose arm imputed (i.e. approximately the 3rd worst case) whilst patients in the placebo arm will have the median for the placebo arm imputed. This analysis will be based on the same ANCOVA model as applied for the primary analysis.

The above sensitivity analyses can be considered as de-facto approaches. The following sensitivity analyses will also be conducted as de-jure estimation approaches to evaluating data:

- A sensitivity analysis will be done in order to evaluate the robustness of the results due to subjects having WLL, open-label molgramostim or rituximab (rescue treatment). For these subjects, all (A-a)DO_2 measurements following the rescue treatment will be excluded and then re-imputed using multiple imputation (in the same way as described above for imputing data for subjects withdrawn from the study early, and who do not have week 24 data available). These subjects will contribute with their data available prior to the rescue treatment and imputed data following the rescue treatment
- A completer analysis will be done, including only data for subjects who completed the 24 weeks and who did not discontinue treatment. This analysis will be based on the same ANCOVA model as applied for the primary analysis.
- The primary analysis will be repeated for PPS.

5.3.7 Secondary Efficacy Analyses

Key Secondary Efficacy Analyses

The following key secondary efficacy endpoints are considered confirmatory endpoints and will therefore be part of the testing procedure in order to control for multiplicity as described in section 5.4.

Absolute change from baseline in 6 minute walking distance and in SGRQ total score after 24-weeks treatment

6 minute walking distance (6MWD)

The 6 minute walking distance (6MWD) is captured as part of the 6 minute walk test (6MWT) (See general description of 6MWT under Further Secondary Efficacy Analyses). Absolute change from baseline in 6 minute walking distance, will be summarized by treatment group and visit. The key secondary endpoint, absolute change from baseline after 24 weeks of treatment, will be analysed as described below.

St Georges Respiratory Questionnaire (SGRQ)

The St Georges Respiratory Questionnaire (SGRQ) is a respiratory specific questionnaire. The symptoms component with 3-month symptoms recall is used. The three domain scores, Symptoms, Activity, and Impacts and the Total score will be calculated. Scores are expressed as a percentage of overall impairment where 100 represents worst possible health status and 0 indicates best possible health status. The standard scoring algorithms will be applied, as described in the St George's Respiratory Questionnaire Manual Version 2.3, June 2009.

The three derived domain scores and total score, and the corresponding changes from baseline will be summarised by treatment group and visit. Individual item responses will only be listed. The key secondary efficacy endpoint, Change from baseline in SGRQ total score after 24-weeks of treatment will be analysed as described below.

Analysis of change from baseline after 24 weeks of treatment in 6 minute walking distance (6MWD) and Total SGRQ

Absolute change from baseline in 6MWD and in SGRQ will be analyzed using the same ANCOVA model as for (A-a)DO_2, including all data irrespective of adherence to treatment regimens as the main approach, and the results from these analyses will be part of the procedures to control for multiple testing (see section 5.4).

Sensitivity analyses for absolute change from baseline in 6MWD and SGRQ will be done using similar approaches as those used for the sensitivity analyses for (A-a)DO_2, but will be limited to:

- Mixed model for repeated measurements (MMRM),
- ANCOVA after multiple imputation,
- ANCOVA completer analysis
- ANCOVA based on the per protocol set (PPS).

Mean profiles for 6MWD and SGRQ over time will be provided.

<u>Time to WLL during 24-weeks treatment</u>

Kaplan-Meier plots will be used to assess time to WLL (time from randomisation to documentation of fulfilment of criterion for requiring WLL) and treatment comparisons corresponding to those made for primary endpoint, will be analysed by a logrank test adjusting for the randomisation strata (WLL). Subjects who fulfil the criteria for rescue treatment (i.e., worsening of aPAP based on symptoms, reduced exercise capacity and/or findings of hypoxemia or desaturation according to the Investigator's judgement), will be counted as if WLL was conducted at the date of fulfilment of rescue treatment criteria. In this analysis, subjects withdrawing from the trial will be handled as censored at time of discontinuation, and subjects who reach week 24 and have not needed WLL will be censored at time point for Visit at week 24. Survival estimates for time to WLL will be tabulated for subjects in each group who require WLL, by relevant time points, along with 95% Cis.

Sensitivity analyses for time to WLL will be carried out as follows:

- Any subject withdrawing from trial before week 24 and before fulfilling criteria for WLL will be handled as having WLL the day after discontinuation date. This is done to

evaluate the impact of missing data and censoring due withdrawals. The same analyses as planned above will be conducted in order to evaluate the robustness of results with regards to the handling of censoring.

- A Cox regression will be applied where the hazard ratio and its 95% confidence interval will be estimated, while including adjustment for treatment, Japan vs other countries and WLL (stratification). The analysis will be done using the same approach for handling censoring as done in the main analysis.

Further Secondary Efficacy Analyses

The further secondary efficacy endpoints will be statistically analysed, although the analyses will be considered supportive rather than confirmatory, and will therefore not be part of the testing procedures implemented for primary and key secondary efficacy endpoints to control for multiplicity.

All data will be tabulated using summary statistics including levels at each visit and change from baseline for continuous data, and counts and shifts including percentage of subjects within categories for frequency data. Tables will be based on all observed data applicable to the FAS population.

Number of WLLs over time

A Generalized Linear Model, based on the negative binomial distribution, is planned for analysis using the number of WLLs as the dependent variable and will include treatment group and the stratification factor of whether or not a WLL was conducted within 2 months prior to the Baseline (according to protocol version 4.0 and later versions), or whether or not WLL was conducted between the screening visit and the randomization visit (according to protocol versions before version 4.0), and country (Japan vs. all others) as fixed effects. The analysis will be adjusted for each subject's length of follow-up in the trial. The rate ratio (each active dose group: placebo) will be derived from the model along with the 95% 2-sided Confidence Intervals. Model adjusted mean rates for each treatment group will be derived. In this analysis each lung lavaged contributes as an event in the count per subject, regardless of whether both lungs are lavaged on different days or on the same day. Events with the date of criteria met in the double-blind period will be counted as being conducted in the double-blind period.

6 minute walking test (6MWT)

Each item related to the six minutes walk test (6MWT) will be summarized by visit. Summaries will include distance and duration of walk during the test, the pre and post walk Borg CR10, the derived change in dyspnoea score (post walk Borg CR10 score – pre walk Borg CR10 score), the SpO_2 at start and the worst SpO_2 during the walk, blood oxygen desaturation derived as (SpO_2 at start – Worst SpO_2 during the walk). Change from baseline after 24 weeks of treatment in; blood oxygen desaturation, difference in pre- to post-test dyspnoea score and distance and duration of walk will in addition be summarized separately including categorical variables corresponding to improved tolerance to exercise (derived as increase in distance walked ≥50 m (compared to baseline) or desaturation <4 percentage points on the 6MWT, respectively), see below.

CT scan

CT scans will be centrally evaluated by two independent and blinded assessors at the 24 weeks treatment visit where change in relation to the baseline assessment will be graded as: *Improved*, *Worsened*, *No change*, *Data missing* or *Impossible to evaluate*. This evaluation, as well as the Ground-Glass Opacification (GGO) score for each anatomical area (upper, middle and lower), (graded as: *no GGO*, *less than 5%*, *5-24%*, *25-49%*, *50-74%*, *75% or more GGO*) will be summarized in frequency tables by treatment group including proportion

of subjects reaching each grade. The number of subjects with, and without, improvements will be analyzed (see, below).

Continuous endpoints

The following continuous endpoints will be analysed using the same main model (ANCOVA) as applied for primary endpoint, including all collected data irrespective of adherence to treatment regimens.

- Absolute change in VC (% predicted), haemoglobin corrected DLCO (% predicted), FEV_1 (% predicted), and FVC (% predicted), and relative change in PaO_2 from baseline after 24 weeks of treatment.
- Change from baseline in pre and post-exercise dyspnoea scores and cough scores (severity and frequency VAS scores) after 24-weeks treatment.

Categorical Endpoints

Number of subjects classifed as having improvements in parameters, as defined below, will be summarised by visit in freqency tables, including cumulative counts. Logistic regression models will be applied to statistically analyse if proportion of subjects having improvements differ between treatment groups according to the same groups as applied for analyses of primary and key secondary endpoints. The logistic regressions will include the same covariates as for main analysis of primary endpoint, although with the baseline covariates corresponding to each respective endpoint. The complement in these analyses, subjects not reaching improvements according to the categorisations, will be based on the total number of subjects applicable to the FAS population.

- Number of subjects with >5 mmHg/>0.67 kPa and number of subjects with >10 mmHg/>1.33 kPa improvement in (A-a)DO_2 after 24-weeks treatment
- Number of subjects with >5 percentage points and number of subjects with >10 percentage points improvement in VC (% predicted) after after 24-weeks treatment
- Number of subjects with >10 percentage points improvement in DLCO (% predicted) after 24-weeks treatment
- Number of subjects with >10 percentage points improvement in FEV_1 (% predicted) after 24-weeks treatment
- Number of subjects with >10 percentage points improvement in FVC (% predicted) after 24-weeks treatment
- Number of subjects with >10% relative improvement in PaO_2 after 24-weeks treatment
- Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m or desaturation <4 percentage points on the 6MWT) compared to baseline after 24-weeks treatment
- Number of subjects with improved CT score after 24-weeks treatment

5.3.8 Exploratory Efficacy Analyses

Exploratory secondary efficacy endpoints will only be summarised descriptively by visit and treatment group unless otherwise specified below.

Double-blind treatment period

Absolute change from baseline in Blood Gas and Lung Function Tests

Absolute change from baseline of (A-a)DO_2, VC (% predicted), DLCO (% predicted), FEV_1 (% predicted), FVC (% predicted), and relative change from baseline of PaO_2 after 4 and 12-weeks treatment will be summarised by visit according to description in previous sections.

Time period during which the (A-a)DO_2 level is maintained below Baseline minus 10mmHg

This time period will be defined as the difference between the date of the visit where the (A-a)DO_2 level first reaches below the baseline level minus 10 mmHg and the date of the first visit where the where the (A-a)DO_2 level first reaches above the baseline level minus 10mmHg. For subjects never reaching the first threshold before ending the double-blind treatment period, the time period will be set to zero, and for subjects never reaching the second treshold, the time period will be set to the duration between the visit where the first treshold it reached and the last visiti of the double-blind treatment period. This time period, in days, will be summarised descriptively.

Number of subjects with improved tolerance to exercise

Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m (compared to baseline) or desaturation <4 percentage points on the 6MWT, respectively) after 4 and 12-weeks treatment will be summarised by visit according to description in previous sections.

Time period during which the improvement in tolerance to exercise is maintained

The time period during which the improvement in tolerance to exercise is maintained will be defined as the difference between the date of the visit where subjects are for the first time classified as having improvement in tolerence to exercise, defined as increase in distance walked ≥50 m (compared to baseline), and the date of the first visit where the increase in distance walked is above ≥50 m compared to baseline on the 6MWT. This time period, in days, will be summarised descriptively.

Dyspnoea score and cough scores

Change from baseline in dyspnoea score and cough scores (severity and frequency) after 4 and 12-weeks treatment wil be summarised as described in previous sections for 24 weeks of treatment.

Number of subjects with improved QoL (change of ≥4 units on the SGRQ The St Georges Respiratory Questionnaire (SGRQ) was described under Key secondary efficacy endpoints. The three derived domain scores and total score, and the corresponding changes from baseline will be summarised by visit. Individual item responses will be listed.

Subjects who have an improvement of more than 4 units from baseline on the SGRQ total score will be counted by visit and treatment group. The number of subjects with improvement of more than 4 units from baseline on the SGRQ total score will be summarised in the way that other categorical data are summarised.

Number of subjects with ‘no problems’ in EQ-5D after 4, 12, and 24-weeks treatment

The EuroQol-5D (EQ-5D) will be used to measure health outcome in terms of mobility,self care, usual activities, pain/discomfort, anxiety/depression. The 5 level instrument (EQ-5D-5L) which uses a 5-point scale (no problems, slight problems, moderate problems, severe problems, and extreme problems) will be used in the local language. A VAS-scale is also included in the EQ-5D questionnaire. Each dimension/item score and the number of subjects in each state will be summarised by visit including the corresponding change from baseline.

Number of subjects where each item on the EQ-5D are scored with the highest outcome, indicating no problems, will likewise be counted by visit and treatment group. The number of subjects with each item, scored as “no problems” on EQ-5D will be summarised in the way that other categorical data are summarised.

Change in serum concentration of possible biomarkers

The possible biomarkers KL-6, CEA, SP-A, SP-B , SP-C, SP-D, Cyfra 21-1 and LDH will be summarised by visit after 4, 12, and 24-weeks treatment including both observed values and change from baseline.

Levels of anti-drug antibodies (anti-GM-CSF, anti-Polyethylene glycol (anti-PEG) and anti-recombinant Human Albumin (anti-rHA)

Number of subjects with and without measured anti-drug antibodies (positive/negative), number of subjects with neutralising antibodies where tested, actual anti-drug antibody titers, and any titer increase or decrease >2 or >4 fold over baseline titer, will be summarised by visit using standard descriptive summary statistics.

Sampling for anti-drug antibodies in France is more frequent than for the other countries, thus additional summaries only including data collected in France will be prepared.

Similar summaries will be done for double-blind treatment period and the Follow-up period.

Change in serum concentration of GM-CSF post first dose of trial drug and after 4-weeks treatment

Pre and 2h post dose serum concentration of GM-CSF, and the corresponding changes (2h post dose measurement – pre dose measurement) will be summarised for first dose of trial drug and after 4 weeks of treatment.

In US sites, additional sampling is done 12 and 24 weeks after randomization, and at week 12 during follow-up, including pre and 2h post dose. These US sampling data will be summarised separately, in the same way as described above.

For sites in Japan additional sampling is done, including pre-dose and samples at 1, 2, 4 and 8 hours post dose. These sampling profiles for Japan will be summarised separately. If applicable (sufficient number of samples have GM-CSF concentrations above LLOQ) standard PK parameters will be derived and evaluated, and the results will be provided in a separate PK-report.

Number of subjects in need for oxygen supplement therapy

Number of subjects in need for oxygen supplement therapy, will be summarised by visit during 24-weeks treatment. The number of subjects in need for oxygen supplement therapy will be defined based on actual use of oxygen supplementation as recorded in concomitant therapy. Oxygen therapy will be summarised separately according to start and stop date of therapy, in the same way that other concomitant medication is presented. Proportion of subjects with supplemental oxygen therapy will be compared between treatment groups using a logistic regression model including treatment, baseline use of supplemental oxygen therapy, strata, and Japan vs Other countries.

The distribution of DSS at Screening and at Week 24

Shift tables will be produced for observed DSS scores at Screening/Baseline vs Week 24, displaying both the visit distributions and the shifts between visits. Similar shift tables will also be produced displaying number of subjects with a DSS score of 1 or 2, vs number of subjects not reaching a DSS score of 1 or 2.

Follow-up period

The following endpoints related to visits in the follow-up period will be summarised, both by randomised treatment during the double-blind treatment period and in total, in the same way that the corresponding endpoints during the double-blind treatment period is summarised.

- Number of subjects requiring WLL, or other treatment for aPAP and number of treatment courses required during 24 and 48-weeks follow-up
- Time to WLL, or other treatment for aPAP during 24 and 48 weeks follow-up
- Absolute change from baseline in (A-a)DO_2 and VC (% predicted), DLCO (% predicted), FEV_1 (% predicted), FVC (% predicted), and relative change from baseline of PaO_2 after 12, 24, 36 and 48-weeks follow-up
- Number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m (compared to baseline) or desaturation <4 percentage points on the 6MWT, respectively) after 12, 24, 36 and 48-weeks follow-up
- Levels of anti-drug antibodies (anti-GM-CSF, anti-Polyethylene glycol (anti-PEG) and anti-recombinant Human Albumin (anti-rHA)) after 12 and 24 weeks follow-up
- The distribution of DSS after 24 and 48-weeks follow-up
- The percentage of subjects with DSS 1 or 2 after 24 and 48-weeks follow-up

In addition to the above endpoints described in the protocol, (A-a)DO_2 values collected during the follow-up period will be summarised only including values assessed based on arterial blood gas samples, excluding capillary samples.

5.3.9 Exposure to Treatment

Exposure to treatment

DB period: exposure will be calculated as the number of days from the date of the first dose (as recorded in Trial Drug part of CRF) to the date of Visit 8 or until date of discontinuation of double-blind Treatment, whichever occurs first.

FU period: exposure for patients enrolled under protocol versions 1 through 3 will be calculated from start and stop dates in the Concomitant Medication part of the CRF. Exposure for patients enrolled under protocol versions 4 through 5 will be calculated from the date of Visit 8 to the date of Visit 11 or until discontinnuation of treatment.

Duration of exposure will be summarized/tabulated by treatment group using descriptive statistics, including the number and proportion of subjects exposed in defined intervals (at least one day to 4 weeks; more than 4 to 8 weeks; more than 8 to 12 weeks; more than 12 to 16 weeks; more than 16 to 20 weeks, and more than 20 to 24 weeks in the DB period, and similarly but in 4-weekly intervals in the FU period).

Exposure as defined above will also be listed by site and subject (including information on premature withdrawal from trial/discontinuation of treatment where relevant).

Compliance

Compliance will be calculated as the total number of empty vials returned divided by the total number of days in the trial.

Compliance will be summarized/tabulated by treatment group using descriptive statistics, including the number and proportion of subjects in defined compliance intervals (90% or more; 80% or more; 70% or more; 50 to 69%; 49% or less). Compliance will also be listed by site and subject (including information on premature withdrawal from trial/discontinuation of treatment where relevant).

Additional data that will only be listed include: answer to CRF question 'Medication taken according to instruction' and nebulizer tracking data (not part of the CRF and only available in a subgroup of subjects).

Rescue medication

Administration of any rescue medication, the type of, reason for, and the corresponding start time and time period of rescue medication will be listed separately.

5.3.10 Concomitant Medication

Concomitant medication and concomitant therapy will be summarised as number of subjects being treated with each type of medication/therapy classified according to Anatomical Therapeutic Chemical (ATC) level 3 and WHO Drug Dictionary preferred term. The safety set will be used for these presentations.

The medications will be divided between those ongoing before baseline (start and end date before baseline) or as ongoing after baseline (starting before or after baseline, and stop date, or still ongoing, after baseline). Listing of concomitant medication will include all medications including an indicator of above defined time intervals.

If the start date of a concomitant medication is missing, it will be assumed that the medication started before treatment. If the date is partially missing, the year and month, if available, will be used to decide whether the medication started before baseline or after baseline. In summary tables, each subject is only counted once for each drug.

The safety set will be used for these presentations.

5.3.11 Adverse Events

AEs will be coded according to the Medical Dictionary for Regulatory Activities (MedDRA) system and tabulated by System Organ Class (SOC) and by Preferred Term (PT).

A summary table will be presented with No. of subjects with AEs, No. of subjects with serious AEs, No. of subjects with severe AEs, No. of subjects with ADRs (probable or possible), Number of subjects with at least one AE leading to discontinuation, and also the number of AEs, severe AEs, serious AEs, ADRs (where a certain AE is only counted once within each subject on a preferred term level).

The total number of subjects with at least one AE and the number of AEs will be derived and summarised by SOC and PT.

Similar tables will be created for serious adverse events.

AEs will also be tabulated by severity (mild, moderate and severe) and relationship to treatment (probable, possible, unlikely or not applicable). Adverse events occurring or worsening after first treatment will be included in summary tables. Any AEs reported before first treatment will only be included in listings.

In listings of AEs the relative day counted from first administration of IMP (Day 1) will be presented together with the relative day from the latest administration of IMP (the actual dates will also be included).

5.3.12 Other Safety Assessments

Any safety assessments described in this section will be summarised in a similar manner for the double-blind and follow-up period, if applicable.

Vital Signs

For vital signs parameters (systolic and diastolic blood pressure, heart rate, respiratory rate and oral body temperature) summary statistics will be produced for observed values by visit.

In addition, the difference from baseline to each visit at which they were assessed will be derived and presented.

Shift tables that show the number of subjects that changed from normal, abnormal (not clinically significant) or abnormal (clinically significant) at baseline to the state at each post-baseline time of assessment will be presented.

Clinical Laboratory Measurements

Clinical laboratory measurements will be assessed at screening, baseline and each visit during the double-blind treatment and follow-up period.

Haematology includes: Haemoglobin, Red Blood Cell Count (RBC), Red Blood Cell Distribution Width (RDW), Haematocrit (PCV/EVF), Mean cell volume (MCV), Mean cell haemoglobin (MCH), Mean cell haemoglobin concentration (MCHC), Platelet count, White cell count, and white cell differential absolute count and percentage: Neutrophils, Lymphocytes, Monocytes, Eosinophils, Basophils, and Prothrombin Time International Normalised Ratio (PT-INR).

Clinical Chemistry includes: Aspartate Aminotransferase (AST), Alanine Aminotransferase (ALT), Gamma Glutamyl Transpeptidase (GGT), Alkaline Phosphatase, Bilirubin, Urea, Creatinine, Potassium, Sodium, Calcium, Chloride, Phosphate, Total protein, Albumin, LDH, C-Reactive Protein (CRP), and Glucose (non-fasting).

Specifically for France, Urinalysis includes: Dipstick measuring for proteinuria, haematuria, glucosuria and nitrite; and microalbuminuria test.

For all laboratory parameters, summary statistics will be produced for observed values by visit. In addition, the difference from baseline to each visit at which they were assessed will be derived and presented. The number of normal, abnormal and clinically significant observations will be tabulated for each treatment group by visit, and in corresponding shift tables from baseline to each following visit. Abnormal values will be flagged in listings.

For laboratory values which are below the limit of quantification, the value corresponding to the limit of quantification will be used when summarising data (e.g. if the result is <0.5 then the value 0.5 will be used in the statistical analysis).

ECG

Electrocardiogram parameters (heart rate, QRS, PR and QT intervals) will be summarised, together with changes from Baseline. Corrected QT will be calculated using Fridericia's formula (QTcF) and summarised, together with changes from Baseline.

Overall interpretation of ECG, normal, abnormal not clinically significant or clinically significant abnormalities will be tabulated by visit and data will be presented in shift tables, where data before baseline is cross-tabulated versus data at later assessments.

Physical Examination

Any abnormal findings in physical examination are reported as medical history or adverse events. Overall interpretation of whether any abnormalities where observed (Yes/No) will be presented in shift tables where data before baseline is cross-tabulated versus data at later assessments.

CT Scan

Occurrence of adverse findings at week 24 compared to Baseline (Yes/No) will be summarised. Description of any adverse findings will be listed.

5.4 Level of Significance, Multiple Comparisons and Multiplicity

For all significance tests, exact p-values (to four decimal places) will be calculated and quoted.

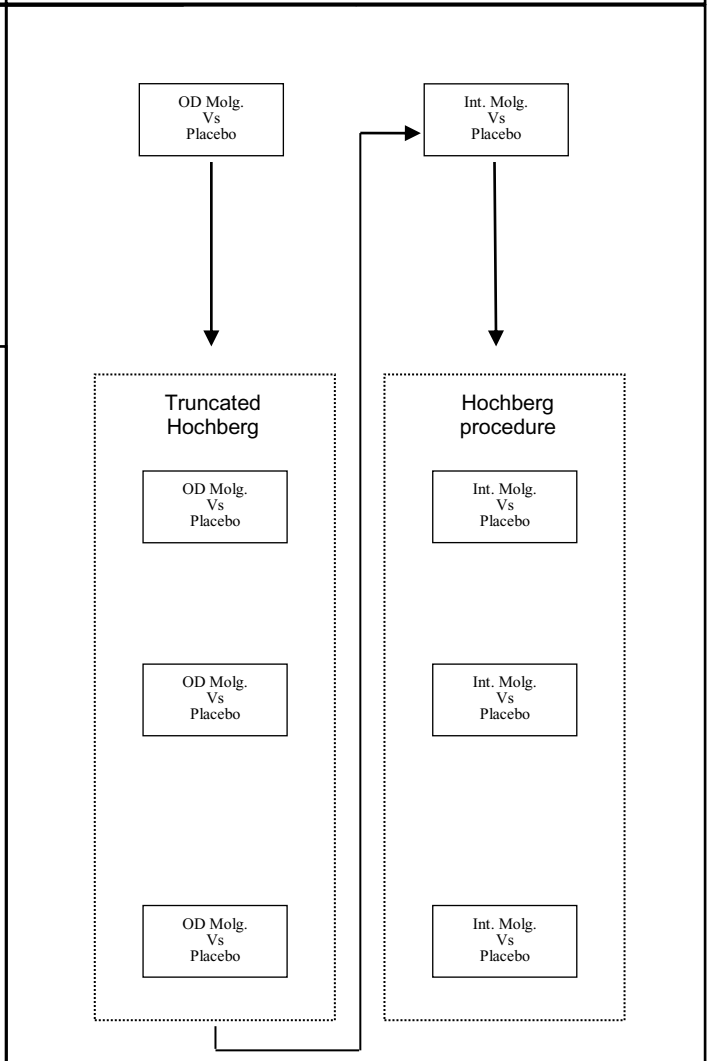
There are two dose regimens and one primary efficacy parameter, and three key secondary efficacy parameters, which are intended to support conclusions based on the primary parameter. For these analyses type-1 error rate will be controlled as described below. In addition, sensitivity analyses and several other endpoints will be considered as supportive and descriptive, and no adjustments for multiplicity will be made for these analyses.

Two sets of comparisons (for the primary and key secondary endpoints) will be carried out. The first will be of once daily dosing vs. placebo, and the second will be of intermittent dosing vs. placebo. In order to account for multiplicity, an overall hierarchical testing procedure will be applied in the steps between primary and secondary endpoints, while a Hochberg procedure will be applied among the secondary endpoints. An overall significance level of 0.05 will be applied.

The first step will be the test of once daily dosing vs. placebo for the primary endpoint. If the null hypothesis is rejected, then once daily dosing vs. placebo will be tested for the three key secondary endpoints using a Truncated Hochberg procedure (applying a truncation factor of 0.75). If at least one null hypothesis is rejected among the key secondary endpoints, then the hierarchical testing will proceed with comparing intermittent dosing vs. placebo for the primary endpoint based on a significance level depending on the number of successes in previous step, as applicable to the truncation factor used. Last, if the null hypothesis is rejected for the comparison of primary endpoint in intermittent dosing vs placebo, then the key secondary endpoints will be compared for intermittent dosing vs placebo using the regular Hochberg and based on an overall significance level corresponding to the previous step. The truncated Hochberg procedure using a truncation factor of 0.75 among the tests in the family of secondary tests between Molgramostim OD and Placebo, implies ordered alpha levels of 0.0417, 0.0229, 0.0167, while preserving alpha 0.05 for the first test between Molgramostim Int. and Placebo if all null hypotheses are rejected, and 0.0083 and 0.0042 if only two or one of the ordered hypotheses are rejected, respectively. The steps of the testing procedure applied to account for multiple testing among the primary and key secondary endpoints are illustrated in figure 2.

All of the analyses will be tabulated separately (irrespective of statistical significance), although an overview of the results of the steps in the testing procedure will be also tabulated.

Figure 2 Overview of Analyses for Primary and Key secondary endpoints

Endpoint / Analysis	Hierarchical testing procedure
Primary Endpoint Change from baseline of (A-a)DO2 after 24-weeks **Main analysis** • ANCOVA (FAS) **Sensitivity analyses** • MMRM (FAS) • ANCOVA MI (FAS) • ANCOVA Tipping Point (FAS) • ANCOVA Worst case (FAS) • ANCOVA (Completer) • ANCOVA (PPS)	
Key Secondary Endpoints **EP 1)** Change from baseline in 6MWD after 24-weeks **Main Analysis** • ANCOVA (FAS) **Sensitivity analyses** • MMRM (FAS) • ANCOVA MI (FAS) • ANCOVA (Completer) • ANCOVA (PPS) **EP 2)** Change from baseline in SGRQ total score after 24-weeks **Main Analysis** • ANCOVA (FAS) **Sensitivity analyses** • MMRM (FAS) • ANCOVA MI (FAS) • ANCOVA (Completer) • ANCOVA (PPS) **EP 3)** Time to WLL during 24-weeks **Main Analysis** • Logrank test (FAS) **Sensitivity analyses** • Cox regression (FAS) • Logrank test (PPS) • Logrank test change in censoring (FAS)	

Results from analyses denoted “Main analysis” will be part of testing procedures where adjustments to control for multiple testing will be implemented.

MI – Multiple imputation,

5.5 Adjustment for Covariates

In all analyses of ‘change from baseline’, the corresponding baseline value of the parameter being analysed will be included in an analysis of covariance (ANCOVA) model. The WLL (stratification) will be included in all analysis models. For endpoints analysed by a mixed model for repeated measurements (MMRM), baseline will be included as a covariate in the model. Geographical region (as defined for subgroups), unique country or centre will not be adjusted for in the statistical analyses, instead adjustment will be made for Japan vs all other countries in all parametric analyses.

5.6 Handling of Dropouts and Missing Data

Different approaches will be applied for evaluation of impact of missing data on the results for primary and key secondary endpoints, each is described in previous sections.

5.7 Multicentre Studies

Summary statistics will be presented for the whole population. For baseline demographics and disease characteristics, primary endpoint, and key secondary endpoints, summary statistics will also be presented by geographical region. The results for the primary and key secondary endpoints will be presented graphically as well as in tables. Centre will not be included as a stratification factor in either the randomisation procedure or data analysis. All subject listings will be sorted by centre.

5.8 Examination of Subgroups

For the primary endpoint, key secondary efficacy endpoints, and AEs summary statistics will be presented by subgroups. In addition to subgroups by region, as stated in protocol, summary statistics of observed data, during the double-blind treatment period, will be prepared for subgroups defined by age, sex, aPAP severity, aPAP duration since diagnosis and WLL conducted previously, for the assessments as described in Table 2.

Table 2: Subgroup Analyses

	Age Groups	**Sex**	**aPAP severity**	**aPAP diagnosis**	**Previous WLL**
	</≥ 65	M/F	DSS 2/3/4/5	</≥ 6m	Y/N
Subject Disposition	+	+	+	+	+
Exposure	+	+	+	+	+
Efficacy					
(A-a)DO2	+	+	+	+	+
6MWD	+	+	+	+	+
SGRQ total score	+	+	+	+	+
Time to WLL	+	+	+	+	+
FVC			+		
DLCO			+		
KL-6			+		
Safety					
AEs (SOC/PT)	+	+			

Furthermore, estimated differences with corresponding 95% Cis will be presented for the efficacy endpoints split by the subgroups in forest plots. The estimated differences and the 95% Cis will be calculated by applying the general ANCOVA.

5.9 Interim Analysis

Analyses will be conducted when all subjects have completed the double-blind treatment period, before all subjects have completed the open-label follow-up period. This does not constitute a statistical interim analysis.

5.10 Data Monitoring

An independent DSMB comprised of pulmonologists from lung centres who are experts in interstitial lung disease, and in conducting clinical studies were appointed and assigned for safety surveillance of the trial. The DSMB members were not investigators in the study. All members of the DSMB were unblinded to the treatment received by the subject and the subject identification.

All study team members remained blinded. Unblinded safety and tolerability data listings for the DSMB review were provided by a statistician from TFS who is independent of the study.

All details of the procedures related to independent unblinding, delivery and review of data are described in the DSMB Charter, V4.0, dated 18 Sep 2018.

6 REFERENCES

[1] Jones P, Spencer S, Adie S. The St George's Respiratory Questionnaire Manual; Version 2.3; June 2009; St George's Hospital Medical School; London, UK

[2] Panel on Handling Missing Data in Clinical Trials; National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials; 2010; The National Academies Press, Washington DC

APPENDIX 1

Data from subjects screened (SCR) but not randomised in the study will only be presented in listings of inclusion/exclusion evaluations and the number of subjects will be included in the subject disposition.

Tables, Figures and Listings to be Produced for the Clinical Study Report (Section 14 and 16 according to ICH E3)

In general, data will be summarised in the same way for the DB and FU period, although will depend on the availability of data in each period. Summaries of efficacy data during the FU period will include data collected during the DB period, while relevant safety output reported at FU may contain data collected during both periods and/or and or data only collected during the FU period. Columns below, SCR (Screened), DB-FAS, DB-PPS, DB-SAF, FU-FAS, FU-MOL INT and FU-SAF indicate the combination of period and population that the corresponding output will be based on. Treatment groups will be abbrevaiated as 'MOL OD' (Molgramostim Once Daily), 'MOL INT' (Molgramostim Intermittently) and 'PBO' (Placebo) for output based on the DB period. For output based on FU period, the abbreaviations 'MOL OD/MOL FU', 'MOL INT/MOL FU' and 'PBO/MOL FU' will be used for the corresponding randomised treatment groups followed by molgramostim in FU period. The final numbering of output may change in the CSR.

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
		14.1 DEMOGRAPHIC DATA							
Table	14.1.1	Subject Disposition Analysis Sets and Reason for Exclusions	X						X
Table	14.1.2	Subject Disposition Analysis Sets and Reason for Exclusions by Age Group	X						
Table	14.1.3	Subject Disposition Analysis Sets and Reason for Exclusions by Sex	X						
Table	14.1.4	Subject Disposition Analysis Sets and Reason for Exclusions by aPAP severity	X						
Table	14.1.5	Subject Disposition Analysis Sets and Reason for Exclusions by aPAP diagnosis	X						
Table	14.1.6	Subject Disposition Analysis Sets and Reason for Exclusions by previous WLL	X						
Table	14.1.7	Subject Discontinuation		X	X	X	X	X	X
Table	14.1.8	Number of Subjects by Visit		X	X	X	X	X	X
Table	14.1.9	Demographics: Age Sex and Race		X	X	X	X	X	X
Table	14.1.10	Demographics: Age Sex and Race by Region		X	X	X			
Table	14.1.11	Demographics: Weight Height and BMI		X	X	X	X	X	X
Table	14.1.12	Demographics: Weight Height and BMI by Region		X	X	X			
Table	14.1.13	aPAP history		X	X	X	X	X	X
Table	14.1.14	aPAP history by Region		X	X	X			
Table	14.1.15	Medical history		X	X	X	X	X	X

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
		14.2 EFFICACY DATA							
		Primary Endpoint							
Table	14.2.1.1	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment		X	X				
Figure	14.2.1.2	Mean profiles of (A-a)DO2 over time		X					
Figure	14.2.1.3	Mean change profiles of (A-a)DO2 over time		X					
Table	14.2.1.4	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 while on treatment by visit during 24 weeks treatment		X					
Table	14.2.1.5	Analysis of primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA)		X					
Table	14.2.1.6	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (MMRM)		X					
Table	14.2.1.7	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA) multiple imputation analysis		X					
Table	14.2.1.8	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA) multiple imputation of measurements following WLL		X					
Table	14.2.1.9	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA) tipping point analysis		X					
Table	14.2.1.10	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA) worst case analysis		X					
Table	14.2.1.11	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA) completer analysis		X					
Table	14.2.1.12	Sensitivity analysis for primary endpoint change from baseline in (A-a)DO2 after 24 weeks (ANCOVA)			X				
Table	14.2.1.13	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment by Age Group		X					
Table	14.2.1.14	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment by Sex		X					
Table	14.2.1.15	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment by aPAP severity		X					
Table	14.2.1.16	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment by aPAP diagnosis		X					
Table	14.2.1.17	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment – by strata		X					
Figure	14.2.1.18	Mean profiles of (A-a)DO2 over time by region		X					
Table	14.2.1.19	Summary of (A-a)DO2 and change from baseline in (A-a)DO2 by visit during 24 weeks treatment – by region		X					
Figure	14.2.1.20	Mean profiles of (A-a)DO2 over time by strata		X					
Figure	14.2.1.21	Forest Plot for analyses of change from baseline in (A-a)DO2, Molgramostim OD vs Placebo by subgroups		X					
Figure	14.2.1.22	Forest Plot for analyses of change from baseline in (A-a)DO2, Molgramostim Int. vs Placebo by subgroups		X					
		Key Secondary Efficacy Endpoints							
Table	14.2.2.1	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit during 24 weeks treatment		X	X				
Figure	14.2.2.2	Mean profiles of six minutes walking distance (6MWD) over time		X					
Figure	14.2.2.3	Mean change profiles of six minutes walking distance (6MWD) over time		X					
Table	14.2.2.4	Analysis of change from baseline in six minutes walking distance (6MWD) after 24 weeks (ANCOVA)		X					
Table	14.2.2.5	Sensitivity analysis of change from baseline in six minutes walking distance (6MWD) after 24 weeks (MMRM)		X					
Table	14.2.2.6	Sensitivity analysis of change from baseline in six minutes walking distance (6MWD) after 24 weeks (ANCOVA) – multiple imputation		X					

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Table	14.2.2.7	Sensitivity analysis of change from baseline in six minutes walking distance (6MWD) after 24 weeks (ANCOVA) – completer analysis		X					
Table	14.2.2.8	Sensitivity analysis of change from baseline in six minutes walking distance (6MWD) after 24 weeks (ANCOVA)			X				
Table	14.2.2.9	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit during 24 weeks treatment by Age Group		X					
Table	14.2.2.10	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit during 24 weeks treatment by Sex		X					
Table	14.2.2.11	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit during 24 weeks treatment by aPAP severity		X					
Table	14.2.2.12	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit during 24 weeks treatment by aPAP diagnosis		X					
Table	14.2.2.13	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit and strata during 24 weeks treatment		X	X				
Figure	14.2.2.14	Mean profiles of six minutes walking distance (6MWD) over time by strata		X					
Table	14.2.2.15	Summary of six minutes walking distance (6MWD) and change from baseline in 6MWD by visit and region during 24 weeks treatment		X					
Figure	14.2.2.16	Mean profiles of six minutes walking distance (6MWD) over time by region		X					
Figure	14.2.2.17	Forest Plot for analyses of change from baseline in minutes walking distance (6MWD), Molgramostim OD vs Placebo by subgroups		X					
Figure	14.2.2.18	Forest Plot for analyses of change from baseline in minutes walking distance (6MWD), Molgramostim Int. vs Placebo by subgroups		X					
Table	14.2.2.19	Summary of SGRQ total score and change from baseline in SGRQ total score by visit during 24 weeks treatment		X	X				
Figure	14.2.2.20	Mean profiles of SGRQ total score over time		X					
Figure	14.2.2.21	Mean change profiles of SGRQ total score over time		X					
Table	14.2.2.22	Analysis of change from baseline in SGRQ total score after 24 weeks (ANCOVA)		X					
Table	14.2.2.23	Sensitivity analysis of change from baseline SGRQ total score after 24 weeks (MMRM)		X					
Table	14.2.2.24	Sensitivity analysis of change from baseline in SGRQ total score after 24 weeks (ANCOVA) – multiple imputation		X					
Table	14.2.2.25	Sensitivity analysis of change from baseline in SGRQ total score after 24 weeks (ANCOVA) – completer analysis		X					
Table	14.2.2.26	Sensitivity analysis of change from baseline in SGRQ total score after 24 weeks (ANCOVA)			X				
Table	14.2.2.27	Summary of SGRQ total score and change from baseline in SGRQ total score by visit during 24 weeks treatment by Age Group		X					
Table	14.2.2.28	Summary of SGRQ total score and change from baseline in SGRQ total score by visit during 24 weeks treatment by Sex		X					
Table	14.2.2.29	Summary of SGRQ total score and change from baseline in SGRQ total score by visit during 24 weeks treatment by aPAP severity		X					
Table	14.2.2.30	Summary of SGRQ total score and change from baseline in SGRQ total score by visit during 24 weeks treatment by aPAP diagnosis		X					
Table	14.2.2.31	Summary of SGRQ total score and change from baseline in SGRQ total score by visit and strata during 24 weeks treatment		X					
Figure	14.2.2.32	Mean profiles of SGRQ total score over time by strata		X					
Table	14.2.2.33	Summary of SGRQ total score and change from baseline in SGRQ total score by visit and region during 24 weeks treatment		X					
Figure	14.2.2.34	Mean profiles of SGRQ total score over time by region		X					
Figure	14.2.2.35	Forest Plot for analyses of change from baseline in SGRQ total score, Molgramostim OD vs Placebo by subgroups		X					

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Figure	14.2.2.36	Forest Plot for analyses of change from baseline in SGRQ total score, Molgramostim Int. vs Placebo by subgroups		X					
Table	14.2.2.37	Summary of number of subjects in need of WLL during 24-weeks of treatment		X	X				
Table	14.2.2.38	Life table summary of time from randomization to need of WLL		X					
Figure	14.2.2.39	Kaplan – Meier plot for time to need of WLL during 24-weeks of treatment		X					
Table	14.2.2.40	Analysis of difference between treatment groups (logrank test) in time to WLL during 24 weeks of treatment		X					
Table	14.2.2.41	Sensitivity analysis of difference between treatment groups (Cox regression) in time to WLL during 24 weeks of treatment		X					
Table	14.2.2.42	Sensitivity analysis of difference between treatment groups (logrank test) in time to WLL during 24 weeks of treatment – alternative censoring		X					
Table	14.2.2.43	Sensitivity analysis of difference between treatment groups (logrank test) in time to WLL during 24 weeks of treatment			X				
Table	14.2.2.44	Summary of number of subjects in need of WLL during 24-weeks of treatment by Age Group		X					
Table	14.2.2.45	Summary of number of subjects in need of WLL during 24-weeks of treatment by Sex		X					
Table	14.2.2.46	Summary of number of subjects in need of WLL during 24-weeks of treatment by aPAP severity		X					
Table	14.2.2.47	Summary of number of subjects in need of WLL during 24-weeks of treatment by aPAP diagnosis		X					
Table	14.2.2.48	Summary of number of subjects in need of WLL during 24-weeks of treatment by strata and by visit		X					
Figure	14.2.2.49	Kaplan – Meier plot for time to need of WLL during 24-weeks of treatment by strata		X					
Table	14.2.2.50	Summary of number of subjects in need of WLL during 24-weeks of treatment by region		X					
Figure	14.2.2.51	Kaplan – Meier plot for time to need of WLL during 24-weeks of treatment by region		X					
Figure	14.2.2.52	Forest Plot for analyses of time to WLL , Molgramostim OD vs Placebo by subgroups		X					
Figure	14.2.2.53	Forest Plot for analyses of time to WLL, Molgramostim Int. vs Placebo by subgroups		X					
Table	14.2.2.54	Overview of statistical hypothesis testing for primary and key secondary efficacy endpoints according to testing procedure		X					
		Further Secondary Efficacy Endpoints							
Table	14.2.3.1	Summary of VC (% predicted) and change from baseline in VC (% predicted) by visit during 24 weeks treatment		X					
Table	14.2.3.2	Analysis of change from baseline in VC (% predicted) after 24 weeks (ANCOVA)		X					
Table	14.2.3.3	Summary of DLCO (% predicted) and change from baseline in DLCO (% predicted) by visit during 24 weeks treatment		X					
Table	14.2.3.4	Analysis of change from baseline in DLCO (% predicted) after 24 weeks (ANCOVA)		X					
Figure	14.2.3.5	Summary of DLCO (% predicted) and change from baseline in DLCO (% predicted) by visit during 24 weeks treatment by aPAP severity		X					
Figure	14.2.3.6	Forest Plot for analyses change from baseline in DLCO (% predicted) by subgroups		X					
Table	14.2.3.7	Summary of FEV1(% predicted) and change from baseline in FEV1(% predicted by visit during 24 weeks treatment		X					
Table	14.2.3.8	Analysis of change from baseline in FEV1(% predicted) after 24 weeks (ANCOVA)		X					
Table	14.2.3.9	Summary of FVC (% predicted) and change from baseline in FVC (% predicted) by visit during 24 weeks treatment		X					
Table	14.2.3.10	Analysis of change from baseline in FVC (% predicted) after 24 weeks (ANCOVA)		X					
Figure	14.2.3.11	Summary of FVC (% predicted) and change from baseline in FVC (% predicted) by visit during 24 weeks treatment by aPAP severity		X					
Figure	14.2.3.12	Forest Plot for analyses of change from baseline in FVC (% predicted) by subgroups		X					

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Table	14.2.3.13	Summary of PaO2 and relative change from baseline in PaO2 by visit during 24 weeks treatment		X					
Table	14.2.3.14	Analysis of relative change from baseline in PaO2 after 24 weeks (ANCOVA)		X					
Table	14.2.3.15	Number and % of subjects with >5 mmHg/>0.67 kPa improvement in (A-a)DO2 compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.16	Analysis of number of subjects with >5 mmHg/>0.67 kPa improvement in (A-a)DO2 compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.17	Number and % of subjects with >10 mmHg/>1.33 kPa improvement in (A-a)DO2 compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.18	Analysis of number subjects with >10 mmHg/>1.33 kPa improvement in (A-a)DO2 compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.19	Number and % of subjects with >5 percentage points improvement in VC (% predicted) compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.20	Analysis of number of subjects with >5 percentage points improvement in VC (% predicted) compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.21	Number and % of subjects with >10 percentage points improvement in VC (% predicted) compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.22	Analysis of number of subjects with >10 percentage points improvement in VC (% predicted) compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.23	Number and % of subjects with >10 percentage points improvement in DLCO (% predicted) compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.24	Analysis of number of subjects with >10 percentage points improvement in DLCO (% predicted) compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.25	Number and % of subjects with >10 percentage points improvement in FEV1(% predicted) compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.26	Analysis of number of subjects with >10 percentage points improvement in FEV1(% predicted) compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.27	Number and % of subjects with >10 percentage points improvement in FVC (% predicted) compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.28	Analysis of number of subjects with >10 percentage points improvement in FVC (% predicted) compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.29	Number and % of subjects with >10% relative improvement in PaO2 compared to baseline by visit during 24 weeks treatment		X					
Table	14.2.3.30	Analysis of number of subjects with >10% relative improvement in PaO2 compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.31	Summary of change from baseline in SpO2 at start and the worst SpO2 oxygen desaturation and distance of walk during 6 minutes walk test (6MWT) by visit during 24 weeks treatment		X					

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Table	14.2.3.32	Number of subjects with improved tolerance to exercise (increase from baseline in distance walked ≥50 m) by visit during 24 weeks treatment		X					
Table	14.2.3.33	Analysis of number of subjects with improved tolerance to exercise (increase from baseline in distance walked ≥50 m) after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.34	Number of subjects with improved tolerance to exercise (desaturation <4 percentage points on the 6MWT) by visit during 24 weeks treatment full analysis set		X					
Table	14.2.3.35	Analysis of number of subjects with improved tolerance to exercise (desaturation <4 percentage points on the 6MWT) after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.36	Summary of Borg CR10 pre post and corresponding difference in walk dyspnea score and change from baseline in scores by visit during 24 weeks treatment		X					
Table	14.2.3.37	Analysis of change from baseline in dyspnea score after 24 weeks (ANCOVA)		X					
Table	14.2.3.38	Summary of cough (VAS) scores and change from baseline in cough scores by visit during 24 weeks treatment		X					
Table	14.2.3.39	Analysis of change from baseline in cough (VAS) scores after 24 weeks (ANCOVA)		X					
Table	14.2.3.40	Summary of number of subjects with Improved Worsened or No change in CT score during 24 weeks of treatment		X					
Table	14.2.3.41	Summary of CT Ground-Glass Opacification (GGO) score during 24 weeks of treatment		X					
Table	14.2.3.42	Analysis of number of subjects with Improved CT score compared to baseline after 24 weeks treatment (Logistic regression)		X					
Table	14.2.3.43	Summary of number of WLLs over time		X					
Table	14.2.3.44	Summary of details related to WLLs including reasons for, one or both lungs, and saline volume used.		X					
Table	14.2.3.45	Analysis of number of WLLs over time (negative binomial regression)		X					
		Exploratory Efficacy Endpoints Double-Blind Treatment Period							
Table	14.2.4.1	Summary of time during which the (A-a)DO2 level is maintained below baseline (A-a)DO2 level -10mmHg		X					
Table	14.2.4.2	Summary of time period during which the improvement in tolerance to exercise is maintained		X					
Table	14.2.4.3	Summary of SGRQ dimension scores and corresponding changes from baseline by visit during 24 weeks of treatment		X					
Table	14.2.4.4	Summary of number of subject with improved QoL score (improvement of at least 4 units on SGRQ total score) during 24 weeks of treatment		X					
Table	14.2.4.5	Summary of EQ-5D-5L dimension scores and corresponding shifts from baseline by visit during 24 weeks of treatment		X					
Table	14.2.4.6	Summary of Krebs von den Lungen-6 (KL-6) and change from baseline in Krebs von den Lungen-6 (KL-6) by visit during 24 weeks of treatment		X					
Table	14.2.4.7	Summary of Krebs von den Lungen-6 (KL-6) and change from baseline in Krebs von den Lungen-6 (KL-6) by visit during 24 weeks of treatment by aPAP severity		X					
Table	14.2.4.8	Summary of Carcinoembryonic antigen (CEA) and change from baseline in Carcinoembryonic antigen (CEA) by visit during 24 weeks of treatment		X					
Table	14.2.4.9	Summary of Surfactant Protein A (SP-A) and change from baseline in Surfactant Protein A (SP-A) by visit during 24 weeks of treatment		X					
Table	14.2.4.10	Summary of Surfactant Protein A (SP-B) and change from baseline in Surfactant Protein A (SP-B) by visit during 24 weeks of		X					

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
		treatment							
Table	14.2.4.11	Summary of Surfactant Protein A (SP-C) and change from baseline in Surfactant Protein A (SP-C) by visit during 24 weeks of treatment		X					
Table	14.2.4.12	Summary of Surfactant Protein A (SP-D) and change from baseline in Surfactant Protein A (SP-D) by visit during 24 weeks of treatment		X					
Table	14.2.4.13	Summary of Cytokeratin 19 Fragment (Cyfra 21-1) and change from baseline in Cytokeratin 19 Fragment (Cyfra 21-1) by visit during 24 weeks of treatment		X					
Table	14.2.4.14	Summary of Lactate Dehydrogenase (LDH) and change from baseline in Lactate Dehydrogenase (LDH) by visit during 24 weeks of treatment		X					
Table	14.2.4.15	Summary of anti-drug antibodies (anti-GM-CSF, anti-Polyethylene glycol (anti-PEG) and anti-recombinant Human Albumin (anti-rHA) by visit during 24 weeks of treatment		X					
Table	14.2.4.16	Summary of anti-drug antibodies (France only) (anti-GM-CSF, anti-Polyethylene glycol (anti-PEG) and anti-recombinant Human Albumin (anti-rHA) by visit during 24 weeks of treatment		X					
Table	14.2.4.17	Summary of GM-CSF pre- and post-dose at baseline and after 4-weeks of treatment and change from pre-dose at baseline in GM-CSF during 24 weeks of treatment		X					
Table	14.2.4.18	Summary of GM-CSF at baseline and after 4-weeks of treatment in pre-dose and repeated post dose (in Japan only) during 24 weeks of treatment		X					
Table	14.2.4.19	Summary of GM-CSF pre- and post-dose at baseline, after 4, 12 and 24 weeks of treatment and change from pre-dose at baseline in GM-CSF (in US only) during 24 weeks of treatment		X					
Table	14.2.4.20	Summary of number and percentage of subjects in need for oxygen supplement therapy during 24-weeks treatment		X					
Table	14.2.4.21	Analysis of number of subjects in need for oxygen supplement therapy during 24-weeks treatment (Logistic regression)		X					
Table	14.2.4.22	Summary of DSS at Screening and at Week 24		X					
Table	14.2.4.23	Shift tables for DSS at Screening and at Week 24		X					
Table	14.2.4.24	Summary of number and percentage of subjects with DSS 1 or 2 at Screening and at Week 24		X					
Table	14.2.4.25	Shift tables for subjects with DSS 1 or 2 at Screening and at Week 24		X					
		Exploratory Efficacy Endpoints Follow-Up Period							
Table	14.2.5.1	Summary of number of subjects in need of WLL or other treatment during 48-weeks follow-up					X	X	
Table	14.2.5.2	Summary of number of subjects in need of WLL or other treatment during 48-weeks follow-up by visit and type of treatment					X	X	
Table	14.2.5.3	Life table summary of time from start of Follow-up need for WLL or other treatment during Follow-up					X	X	
Figure	14.2.5.4	Kaplan – Meier plot for time from start of Follow-up need for WLL or other treatment during Follow-up					X	X	
Table	14.2.5.5	Summary of and change from baseline in (A-a)DO2 during 48-weeks follow-up by visit					X	X	
Table	14.2.5.6	Summary and change from baseline in (A-a)DO2 by visit (including only samples for arterial blood gas) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.7	Summary of and change from baseline in VC (% predicted) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.8	Summary of and change from baseline in DLCO (% predicted) during 48-weeks follow-up by visit					X	X	

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Table	14.2.5.9	Summary of and change from baseline in FEV1 (% predicted) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.10	Summary of and change from baseline in FVC (% predicted) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.11	Summary of and relative change from baseline of PaO2 during 48-weeks follow-up by visit					X	X	
Table	14.2.5.12	Summary of number of subjects with improved tolerance to exercise (increase in distance walked ≥50 m) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.13	Summary of number of subjects with improved tolerance to exercise (desaturation <4 percentage points on the 6MWT) during 48-weeks follow-up by visit					X	X	
Table	14.2.5.14	Summary of Borg CR10 pre post and corresponding difference in walk dyspnoea score and change from baseline in scores during 48-weeks follow-up by visit					X	X	
Table	14.2.5.15	Summary of cough (VAS) scores and change from baseline in cough scores during 48-weeks follow-up by visit					X	X	
Table	14.2.5.16	Summary of SGRQ dimension scores and corresponding changes from baseline during 48-weeks follow-up by visit					X	X	
Table	14.2.5.17	Summary of EQ-5D-5L dimension scores and corresponding shifts from baseline during 48-weeks follow-up by visit					X	X	
Table	14.2.5.18	Summary of levels of and change from baseline in anti-GM-CSF after 12 and 24 weeks of follow-up by visit					X	X	
Table	14.2.5.19	Summary of GM-CSF pre- and post-dose during Follow-up (US only)					X	X	
Table	14.2.5.20	Summary of DSS at during 48-weeks follow-up by visit					X	X	
Table	14.2.5.21	Shift tables for DSS during 48-weeks follow-up by visit					X	X	
Table	14.2.5.22	Summary of number and percentage of subjects with DSS 1 or 2 during 48-weeks follow-up by visit					X	X	
Table	14.2.5.23	Shift tables for subjects with DSS 1 or 2 during 48-weeks follow-up by visit					X	X	
Table	14.2.5.24	Summary of details related to WLLs including reasons for, one or both lungs, and saline volume used during 48-weeks follow-up.					X	X	
		Exposure							
Table	14.2.6.1	Extent of Exposure		X	X	X	X	X	X
Table	14.2.6.2	Extent of Exposure by Age Group				X			
Table	14.2.6.3	Extent of Exposure by Sex				X			
Table	14.2.6.4	Extent of Exposure by by aPAP severity				X			
Table	14.2.6.5	Extent of Exposure by aPAP diagnosis				X			
Table	14.2.6.6	Extent of Exposure by previous WLL				X			
Table	14.2.6.7	Compliance		X	X	X	X	X	X
Table	14.2.6.8	Concomitant Medication				X			X
		14.3 SAFETY DATA							
		Displays of Adverse Events							
Table	14.3.1.1	Summary of adverse events during 24 weeks of treatment				X			
Table	14.3.1.2	Adverse events by system organ class and preferred term during 24 weeks of treatment				X			

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Table	14.3.1.3	Adverse events: Number of subjects by severity and relationship to treatment during 24 weeks of treatment				X			
Table	14.3.1.4	Summary of serious adverse events during 24 weeks of treatment				X			
Table	14.3.1.5	Serious adverse events by system organ class and preferred term during 24 weeks of treatment				X			
Table	14.3.1.6	Serious adverse events: Number of subjects by severity and relationship to treatment during 24 weeks of treatment				X			
Table	14.3.1.7	Adverse events by system organ class and preferred term during 24 weeks of treatment by Age Groups				X			
Table	14.3.1.8	Adverse events by system organ class and preferred term during 24 weeks of treatment by Sex				X			
Table	14.3.1.9	Summary of adverse events in total and during 48 weeks of follow-up							X
Table	14.3.1.10	Adverse events by system organ class and preferred term in total and during 48 weeks of follow-up							X
Table	14.3.1.11	Adverse events: Number of subjects by severity and relationship to treatment total and during 48 weeks of follow-up							X
Table	14.3.1.12	Summary of serious adverse events during total and during 48 weeks of follow-up							X
Table	14.3.1.13	Serious adverse events by system organ class and preferred term total and during 48 weeks of follow-up							X
		Listings of Deaths Other Serious and Significant Adverse Events							
Listing	14.3.2.1	List of deaths				X			X
Listing	14.3.2.2	List of other serious adverse events				X			X
Listing	14.3.2.3	List of adverse events leading to treatment and/or trial discontinuation				X			X
		Laboratory Data							
Listing	14.3.4.1	Laboratory reference ranges	X						
Listing	14.3.4.2	Abnormal laboratory Value Listing				X			X
Table	14.3.5.1	Summary of haematology laboratory measurements and corresponding changes from baseline by visit during 24 weeks of treatment				X			
Table	14.3.5.2	Shift tables for haematology laboratory measurements from baseline by visit during 24 weeks of treatment				X			
Table	14.3.5.3	Summary of biochemistry laboratory measurements and corresponding changes from baseline by visit during 24 weeks of treatment				X			
Table	14.3.5.4	Shift tables for biochemistry laboratory measurements from baseline by visit during 24 weeks of treatment				X			
Table	14.2.5.5	Summary of urinalysis laboratory measurements (France only) by visit during 24 weeks of treatment				X			
Table	14.3.5.6	Summary of haematology laboratory measurements and corresponding changes from baseline by visit during 48 weeks of follow-up							X
Table	14.3.5.7	Shift tables for haematology laboratory measurements from baseline by visit during 48 weeks of follow-up							X
Table	14.3.5.8	Summary of biochemistry laboratory measurements and corresponding changes from baseline by visit during 48 weeks of follow-up							X
Table	14.3.5.9	Shift tables for biochemistry laboratory measurements from baseline by visit during 48 weeks of follow-up							X
Table	14.3.5.10	Summary of urinalysis laboratory measurements (France only) by visit during 48 weeks of treatment							X

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
		Other Safety Data							
Table	14.3.6.1	Summary of vital signs and change from baseline in vital signs by visit during 24 weeks of treatment				X			
Table	14.3.6.2	Shift tables for changes in vital signs from baseline by visit during 24 weeks of treatment				X			
Table	14.3.6.3	Shift tables for changes in Physical examination from baseline by visit during 24 weeks of treatment				X			
Table	14.3.6.4	Summary of ECG and change from baseline in ECG by visit during 24 weeks of treatment full analysis set				X			
Table	14.3.6.5	Shift tables for changes in ECG from baseline by visit during 24 weeks of treatment				X			
Table	14.3.6.6	Summary of adverse findings in CT at week 24 compared to baseline				X			
Table	14.3.6.7	Summary of vital signs and change from baseline in vital signs by visit during 48 weeks of follow-up							X
Table	14.3.6.8	Shift tables for changes in vital signs from baseline by visit during 48 weeks of follow-up							X
Table	14.3.6.9	Shift tables for changes in Physical examination from baseline by visit during 48 weeks of follow-up							X
Table	14.3.6.10	Summary of ECG and change from baseline in ECG by visit during 48 weeks of follow-up							X
Table	14.3.6.11	Shift tables for changes in ECG from baseline by visit during 48 weeks of follow-up							X
Listing	16.1.7	Randomization Scheme		X					
		16.2 PATIENT DATA LISTINGS							
		Discontinued Patients							
Listing	16.2.1.1	Discontinued Subjects Reason for Discontinuation				X			X
Listing	16.2.1.2	Visit Dates and Other Important Dates (informed consent rand date study termination)				X			X
Listing	16.2.1.3	Study Termination				X			X
Listing	16.2.2	Protocol Deviations	X						X
		Patients excluded from the efficacy analysis							
Listing	16.2.3.1	Subjects Excluded from the Efficacy Analysis (Evaluability Reason for Evaluability Classification)		X			X		
Listing	16.2.3.2	Treatment Allocation and Evaluability for All Subjects		X			X		
		Demographic data							
Listing	16.2.4.1	Demographic Data				X			X
Listing	16.2.4.2	Medical History				X			X
Listing	16.2.4.3	aPAP History				X			X
Listing	16.2.4.4	Inclusion Criteria Not Met and Exclusion Criteria Met				X			X
		Compliance and/or drug concentration data							
Listing	16.2.5.1	Exposure				X			X

			SCR	DB-FAS	DB-PPS	DB-SAF	FU-FAS	FU-MOL INT	FU-SAF
Listing	16.2.5.2	Compliance				X			X
Listing	16.2.5.3	Rescue medication				X			X
		Individual efficacy response data							
Listing	16.2.6.1	Individual Efficacy Response Data (A-a)DO2 PaO2 and Lung function tests		X			X		
Listing	16.2.6.2	Individual Efficacy Response Data 6MWT		X			X		
Listing	16.2.6.3	Individual Efficacy Response Data GM-CSF		X			X		
Listing	16.2.6.4	Individual Efficacy Response Data Anti-Drug antibodies		X			X		
Listing	16.2.6.5	Individual Efficacy Response Data Biomarkers		X			X		
Listing	16.2.6.6	Individual Efficacy Response Data DSS		X			X		
Listing	16.2.6.7	Individual Efficacy Response Data Dyspnoea score		X			X		
Listing	16.2.6.8	Individual Efficacy Response Data Cough scores		X			X		
Listing	16.2.6.9	Individual Efficacy Response Data EQ-5D		X			X		
Listing	16.2.6.10	Individual Efficacy Response Data SGRQ		X			X		
Listing	16.2.6.11	Individual Efficacy Response Data CT		X			X		
		Adverse Event listings							
Listing	16.2.7.1	Adverse Event Listings by Treatment Subject Relative Day or Week				X			X
Listing	16.2.7.2	Adverse Event Listings by System Organ Class Preferred Term Treatment and Subject				X			X
		Listing of individual laboratory measurements by patient							
Listing	16.2.8.1	Listing of Reference Ranges	X						
Listing	16.2.8.2	Listing of Individual Laboratory Measurements by Subject				X			X
Listing	16.2.8.3	Listing of Relevant Comments Regarding Laboratory Values				X			X
Listing	16.2.9	Vital Signs				X			X
Listing	16.2.10	Physical Examination				X			X
Listing	16.2.11	ECG				X			X
Listing	16.2.12	Concomitant Medications and Therapy				X			X

It's all about trust

APPENDIX 2. SAMPLE SIZE RECALCULATION MOL-PAP-002

1.1 Introduction

This document describes the results from the pre-planned blinded sample size re-calculation as described in the Protocol, Final Version 5, dated 01 December 2017 (Section 12.2). The sample size calculation was planned to be in January 2018, when approximately 50 patients would have reached 24 weeks of treatment.

1.2 Methods

The blinded sample size recalculation was conducted for the following three Key secondary endpoints according to protocol

- Change from baseline to week 24 in 6MWD
- Change from baseline to week 24 in SGRQ Total Score
- Time to WLL

The change from baseline to week 24 in 6MWD [and SGRQ] were analysed in an ANCOVA model including WLL (stratification), and baseline 6MWD [SGRQ]. For the purposes of the sample size re-estimation process, only data on patients who had baseline and week 24 data was used and no imputations were made for missing values.

The minimal clinically important effect sizes for 6MWD and SGRQ in aPAP were considered to be in the order of 40-50 m and 7-10 points, respectively. From each ANCOVA model, the respective residual variance for 6MWD [and SGRQ] was used to calculate sample size based on the effect sizes of 50 m and 10 points, respectively.

The overall (pooled) event rate for WLL was determined including any patients who would meet the criteria for WLL any time prior to week 24, but excluding any patients who withdraw prior to week 24 where WLL has not been required. WLL rates of 5% (active) versus 20% (placebo) at 24 weeks was assumed, and compared to the observed (pooled) WLL rate.

All sample size calculations correspond to two group comparisons and were made based on a 2-sideded 5% significance level, and 80% and 90% power. The observed SD was increased/ decreased with 10% for sensitivity. All sample size calculations were treated as independent calculations without considering that the power to reject any hypothesis is conditional on first rejecting the null hypothesis for the primary endpoint in a hierarchical manner.

If the maximum of these sample size calculations were to be no more than 33 patients per group, then no change to the target sample size would be made. Otherwise, the sample size would be increased to attain 90% power for at least 2 of the 3 key secondary endpoints, subject to the total study size not exceeding 150 patients.

The sample size calculations were done using proc Power in SAS Enterprise Guide version 7.12.

1.3 Results from Sample Size Recalculation

The sample size for SGRQ Total Score was based on an analysis of 51 subjects that had both a baseline score and a score after 24 weeks of treatment. From the ANCOVA model for change from baseline, including adjustment for strata and baseline score, the Residual Mean Square Error (MSE) was 101.582, translating to an SD of 10.1. In addition to the calculation based on the observed SD, sample sizes were also calculated assuming a 10% increase/decrease in SD.

From Table 1, it can be seen that the number of subjects needed per treatment group needs to be at least 23 in order to detect a significant difference between groups of 10 based on the observed SD with a power of 90%.

Table 1 Recalculated Sample Size to detect a difference of 10 in Total SGRQ

	Estimated Standard Deviation		
Power	9.1	10.1	11.1
80%	15	18	21
90%	19	**23**	27
ANCOVA MSE = 101.582 based on 51 subjects with a result at baseline and after 24 weeks.			

As for SGRQ Total Score the sample size calculation for the 6MWD was based on 51 subjects having both a baseline measurement and a corresponding measurement after 24 weeks of treatment. From the ANCOVA including adjustment for baseline and strata the estimated residual MSE was 4928.203, corresponding to an SD=70.2. For sensitivity the SD was increased/decreased with 10%. As can be seen from Table 2, based on the observed SD at least 43 subjects would be needed per treatment group in order to detect a difference of 50m between any groups with a power of 90%.

Table 2 Recalculated Sample Size to detect a difference of 50 in 6MWD

	Estimated Standard Deviation		
Power	63.2	70.2	77.2
80%	27	32	39
90%	35	**43**	52
ANCOVA MSE = 4928.203 based on 51 subjects with a result at baseline and after 24 weeks.			

The observed event rate of WLL was low, only 4 events among 78 randomized subjects either having WLL and then discontinuing, or subjects still remaining in trial. For the calculations it was anticipated the sample size would be calculated assuming 5% of subjects on active treatment vs 20% on placebo would have an event. Given these assumed event rates, and assuming an exponential distribution of survival times, at least 47 subjects would be needed to detect a difference between treatment groups using a log-rank test, with 80% power, and 63 subjects with 90% power, if these event rates are true. Given the low event rate, even though probably slightly underestimated in observed subjects, these calculated sample sizes are likely too optimistic.

1.4 Conclusion

From these sample size calculations it is clear that the two endpoints with highest power to detect a clinically relevant difference between treatment groups based on the least number of subjects are SGRQ Total Score and 6MWD. Based on the pre-defined criterion that the sample size would be increased to attain 90% power for at least 2 of the 3 key secondary endpoints, the results indicate that the critical endpoint is the 6MWD, which would require that the sample size should be increased to at least 43x3=129 randomised subjects. After reviewing the results from the sample-size recalculation it was decided by Savara ApS to increase the sample size to 45x3=135 randomised subjects. Any subjects already in screening at the time of reaching 135 randomised subjects may also be randomised, if otherwise eligible for the trial.

APPENDIX 3. CONVERSION FACTORS

For parameters that were reported in the eCRF in more than one unit, the following conversion factors will be applied:

Parameter	Reported Unit:	Converted Unit:	Conversion Factor:
(A-a)DO2	kPa	mmHg	Multiply reported result with 7.5
Absolute DLCO/TLCO (corrected for haemoglobin) *	mmol/min/kPa	ml/min/mmHg	Multiply reported result with 2.987
Absolute DLCO/TLCO (uncorrected) *	mmol/min/kPa	ml/min/mmHg	Multiply reported result with 2.987
Haemoglobin value used for correction	g/L	g/dL	Multiply reported result with 0.1
Haemoglobin value used for correction	mmol/L	g/dL	Multiply reported result with 1.61
O2 flow rate	ml/h	L/min	Multiply reported result with 0.000017
PaCO2	kPa	mmHg	Multiply reported result with 7.5
PaO2	kPa	mmHg	Multiply reported result with 7.5
Patm	kPa	mmHg	Multiply reported result with 7.5

* https://www.thoracic.org/statements/resources/pft/DLCO.pdf