Title: Statistical Analysis Plan for Protocol SAV006-05 A randomized, double-blind, placebo-controlled clinical trial of once-daily inhaled molgramostim nebulizer solution in adult subjects with autoimmune pulmonary alveolar proteinosis (aPAP)

Compound Name: Molgramostim 300 µg nebulizer solution

Effective Date: 24APR2024

Subject: Autoimmune Pulmonary Alveolar Proteinosis (aPAP) Disease

Author's Name, Title, and Functional Area: [redacted]

SIGNATURE PAGES

The signatories on the following pages have all read and approved this present version of the document.

This document has been prepared and/or reviewed by:

NAME:	[redacted]
TITLE:	[redacted]
SIGNATURE AND DATE:	[redacted]

This document has been reviewed and accepted by:

NAME:	[redacted]
TITLE:	[redacted]
SIGNATURE AND DATE:	[redacted]

NAME:	[redacted]
TITLE:	[redacted]
SIGNATURE AND DATE:	[redacted]

TABLE OF ABBREVIATIONS

A-aDO_2	Alveolar-arterial oxygen difference
ADR	Adverse drug reaction
AE	Adverse event
AESI	Adverse event of special interest
Anti-GM-CSF	Antibodies to granulocyte macrophage colony stimulating factor
aPAP	Autoimmune pulmonary alveolar proteinosis
ATC	Anatomical Therapeutic Chemical
ATS	American Thoracic Society
BAL	Bronchoalveolar lavage
CEA	Carcinoembryonic antigen
CGIC	Clinician's Global Impression of Change
CGIS	Clinician's Global Impression of Severity
CI	Confidence interval
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
CS	Clinically significant
CSP	Clinical Study Protocol
CT	Computed tomography
CV%	Coefficient of variation
CYFRA-21-1	Cytokeratin fragment
DLCO	Diffusing capacity of the lungs for carbon monoxide
$DLCO_{adj}$	Hemoglobin-adjusted DLCO
DMC	Data Monitoring Committee
DSS	Disease severity score
EC	Exercise capacity
ECG	Electrocardiogram
eCRF	Electronic case report form

EDC	Electronic Data Capture
eDiary	Electronic Diary
EQ-5D-5L	EuroQoL 5 Dimensions, 5 Levels
ERS	European Respiratory Society
FAS	Full analysis set
FEV_1	Forced expiratory volume in one second
FDA	Food and Drug Administration
FiO_2	Fraction of inspired oxygen
FVC	Forced vital capacity
GGO	Ground glass opacification/opacities
GM-CSF	Granulocyte macrophage colony stimulating factor
Hgb	Hemoglobin
Hct	Hematocrit
HRCT	High resolution computed tomography
ICF	Informed Consent Form
IMP	Investigational medicinal product
IRT	Interactive Response Technology
LDH	Lactate dehydrogenase
LK-6	Krebs von den Lungen-6
LL	Lung lavage
LSMean	Least squares mean
MAR	Missing at random
MedDRA	Medical Dictionary for Regulatory Activities
MET	Metabolic equivalent
MMRM	Mixed model for repeated measures
MNAR	Missing not at random
MOL	Molgramostim 300 µg nebulizer solution
NC	Not calculable

NCA	Non-compartmental analysis
NCS	Not clinically significant
NQ	Not quantifiable
PaO_2	Arterial partial pressure of oxygen
$PaCO_2$	Arterial partial pressure of carbon dioxide
PAP	Pulmonary alveolar proteinosis
PBO	Placebo (treatment group)
PGIC	Patient's Global Impression of Change
PGIS	Patient's Global Impression of Severity
PPS	Per-protocol set
PT	Preferred term
Q1	Quartile 1
Q3	Quartile 3
QTcB	QT interval corrected by Bazett
QTcF	QT interval corrected by Fridericia
rhGM-CSF	Recombinant human granulocyte macrophage colony stimulating factor
RPE	Rating of Perceived Effort
SAE	Serious adverse event
SAP	Statistical analysis plan
SAF	Safety analysis set
SD	Standard deviation
SDTM	Study Data Tabulation Model
SGRQ	Saint George's Respiratory Questionnaire
SoA	Schedule of Activities
SOC	System organ class
SpO_2	Oxygen saturation
US	United States
VAS	Visual Analogue Scale

VO_2	Maximum oxygen consumption
WHO	World Health Organization

TRADEMARK INFORMATION

SAS SAS (Statistical Analysis Software) is a registered trademark of SAS Institute Inc.

REVISION HISTORY

Version	Date	Summary of revisions
1.0	24APR2024	Not applicable

1. INTRODUCTION

Molgramostim nebulizer solution is being developed by Savara for the treatment of autoimmune pulmonary alveolar proteinosis (aPAP).

Data from a completed phase 2/3 trial, which was Savara's first trial in aPAP subjects, MOL-PAP-002, suggested that molgramostim nebulizer solution improves lung pathology, pathophysiology, and health status in a dose-frequency dependent fashion (Trapnell, et al. 2020). The present phase 3 trial, SAV006-05, is being conducted to further investigate the efficacy and safety of molgramostim nebulizer solution in subjects with aPAP. The overall trial design is described in Section 3.1, and the Schedules of Activities (SoAs) are provided in Appendix B. Refer to the SoAs for the planned time points for all assessments, including those for efficacy and safety.

This statistical analysis plan (SAP) provides details of the summaries and analyses to be performed to report the findings of the trial through the end of the original open-label treatment period at Visit 20/Week 144 and the 4-week safety follow-up (Visit 22/Week 148). This SAP should be read in conjunction with the Clinical Study Protocol (CSP), SAV006-05 version 9.1 (21 NOV 2023). A protocol amendment, version 9.2 (27 MAR 2024) has been put in place to extend the open-label treatment period for subjects enrolled in clinical sites in Japan. A SAP amendment applicable to Japan only will be produced separately.

2. TRIAL OBJECTIVES AND ENDPOINTS

2.1 PRIMARY EFFICACY OBJECTIVE AND ENDPOINT

Objective	Endpoint
Primary Efficacy	
Investigate the efficacy of molgramostim compared to placebo with respect to the following endpoint:	• Change in percent (%) predicted diffusing capacity of the lungs for carbon monoxide (DLCO)* from baseline to Week 24

*Using hemoglobin (Hgb)-adjusted DLCO ($DLCO_{adj}$)

The single-breath DLCO test is performed in accordance with American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines for DLCO testing (Graham et al. 2017). Further standardization across sites is obtained by using standardized equipment (EasyOne Pro®, ndd Medical Technologies) and implementing central overread by a team of external independent respiratory experts. At least two acceptable and repeatable maneuvers according to ATS/ERS criteria are required. Up to five maneuvers may be conducted, if needed, during a session. For the final DLCO result, the average of all acceptable efforts, as determined by the overreader, is used. Central overread is conducted in real time (within 1 hour) at Baseline, Week 24, and Week 48.

From the absolute DLCO value, the % predicted DLCO is calculated by the EasyOne Pro device. The absolute DLCO value is then adjusted for the hemoglobin (Hgb) value obtained on the same day as the DLCO test and available from the central laboratory. Using this absolute Hgb-adjusted DLCO ($DLCO_{adj}$), the % predicted $DLCO_{adj}$ is calculated. Change in % predicted $DLCO_{adj}$ are the values of interest for the primary and secondary DLCO-related efficacy endpoints.

After best test review, the DLCO test set at a visit is assigned a grade of Acceptable, Borderline, or Unacceptable by the overreader. Only DLCO results from a set assigned a grade of either Acceptable or Borderline will be included in the analysis of DLCO results, including the primary endpoint. All best test results, regardless of the grade assigned by the overreader, will be provided in a listing. All DLCO efforts are available in the Study Data Tabulation Model (SDTM) data package.

2.2 SECONDARY EFFICACY OBJECTIVES AND ENDPOINTS

Objectives	Endpoints
Secondary Efficacy	
Investigate the efficacy of molgramostim compared to placebo with respect to the following endpoint:	Secondary (For regulatory authorities outside of Japan and Korea)
	• Change in % predicted DLCO from baseline to Week 48
	• Change in Saint George's Respiratory Questionnaire (SGRQ) Total score from baseline to Week 24
	• Change in SGRQ Activity score from baseline to Week 24
	• Change in Exercise Capacity (EC), expressed as peak metabolic equivalents (METs) from baseline to Week 24
	• Change in SGRQ Total from baseline to Week 48
	• Change in SGRQ Activity from baseline to Week 48
	• Change in EC (expressed as peak METs) from baseline to Week 48
	Secondary (Specifically for Japan and South Korea)
	• *Change in SGRQ Total from baseline to Week 24*
	• *Change in SGRQ Activity from baseline to Week 24*
	• *Change in EC (expressed as peak METs) from baseline to Week 24*
	• *Change in alveolar-arterial oxygen difference (A-aDO_2) from baseline to Week 24*

Further details on SGRQ, EC, and A-aDO_2 are provided in the sections that follow.

2.2.1 SAINT GEORGE'S RESPIRATORY QUESTIONNAIRE

The SGRQ includes questions related to three components: Activity (activities that cause or are limited by breathlessness); Impact (social functioning and psychological disturbances resulting from

airway disease); and Symptoms (effect of respiratory symptoms, their frequency and severity). A component score is calculated for each component, and a total score is also calculated that summarizes the impact of the disease on overall health status.

The Activity component assesses the subject's current state in terms of disturbance to the subject's daily physical activity. The two questions used to calculate the Activity component score are disease non-specific and measure functional aspects of respiration.

The Impact component assesses the subject's current state and measures the impact of cough and breathlessness on a physical, psychosocial, and daily activity perspective. The questions used to calculate the Impact component score are disease non-specific.

The questions in the Symptoms component cover 4 symptoms (cough, sputum, shortness of breath, and attacks of wheezing) and 4 additional symptom-related questions (number of severe or very unpleasant respiratory attacks, duration of worst respiratory attack, number of good days with few respiratory problems, and morning wheeze).

SGRQ Total and component scores are scaled from 0 to 100, with higher scores indicating worse quality of life. Refer to Appendix D for further details on deriving the total and component scores.

Patient reported outcome data, including the SGRQ, are collected at sites during visits via the use of an eDiary, a smartphone device provided to each subject. The paper version of a questionnaire is used at a visit only if the eDiary cannot be used due to technical issues or if a subject forgets to bring their eDiary to their visit.

2.2.2 EXERCISE CAPACITY

As a functional measure of exertional limitation related to dyspnea, EC is assessed by an exercise treadmill test. EC is expressed in peak METs (1 MET=3.5 ml/O_2/kg/min).

A conservative ramp-up treadmill protocol, employing minimal adjustments in speed and grade from one stage to the next, is used. The highest treadmill speed and grade achieved are used to calculate peak METs. The treadmill test is conducted by qualified staff experienced in the conduct of clinical exercise testing.

Central overread by an external, independent expert is implemented. The overreader reviews all data collected in the test and makes the final confirmation that the exercise test is valid according to pre-specified criteria. Only results from valid exercise tests will be included in the analysis of data from the tests, including the secondary endpoint of peak METs. All results, whether valid or invalid, will be provided in a listing. The final confirmation of which tests are valid will be made by the overreader in a blinded manner and documented before the double-blind part of the trial database is locked and randomized treatment assignments are unblinded (after Visit 11/Week 48).

The exercise test consists of 3 phases:

- Pre-exercise phase monitoring (i.e., heart rate, ECG, blood pressure, pulse oximetry, and symptoms) with collection of data at rest. Duration 5 minutes.

- Exercise phase with controlled ramp-up exercise, increasing speed and gradient on the treadmill every 30 seconds. The initial stage is at 1 mph with no incline. For each stage, the speed is increased by 0.1 mph and the gradient by 0.5%. The full protocol comprises 31 stages (total duration of exercise phase is 15 minutes and 30 seconds). Subjects are encouraged to exercise until they achieve maximal effort. The subject may request to stop the test at any time. The exercise test is terminated prior to maximal effort if absolute termination criteria (e.g., drop in systolic blood pressure, ventricular arrythmia, and ST segment depression) develop.
- Recovery phase: If the test is not terminated prematurely, the first two minutes of recovery consist of walking at 1 mph and 0% gradient followed by passive recovery for 4 minutes. If no abnormal signs/symptoms develop, total duration of the monitored (i.e., heart rate, ECG, blood pressure, pulse oximetry, and symptoms) recovery phase is 6 minutes. Monitoring may continue past 6 minutes if abnormal signs/symptoms persist.

Based on the data collected during the test, the following parameters are calculated:

- Sufficient effort exhibited during exercise testing (Yes or No, according to the criteria listed in Section 8.2.3 of the protocol) – as confirmed by the central overreader.
- Peak METs (using an established equation based on the speed and grade of the last stage the subject was able to complete for at least 15 seconds). The following validated equation to calculate peak METs will be used:

  Peak METs = (speed X (0.17 + fractional grade X 0.79) + 3.5)/3.5
  Note: Speed in meters/minute (Kokkinos et al. 2017)

- Distance walked.
- Duration of exercise.

The scores from the Borg CR Scale® are used to investigate changes in dyspnea during the exercise treadmill test.

2.2.3 ALVEOLAR-ARTERIAL OXYGEN DIFFERENCE

The following variables are assessed from an arterial blood gas sample collected on room air:

- PaO_2 (mmHg/kPa) – arterial partial pressure of oxygen
- $PaCO_2$ (mmHg/kPa) – arterial partial pressure of carbon dioxide

As a measure of gas exchange, A-aDO_2 will be calculated using the formula in Section 9.5.2.2.

2.3 SAFETY OBJECTIVES AND ENDPOINTS

Objectives	Endpoints
Safety	
Investigate the safety of molgramostim compared	• Frequencies of (serious) adverse events [(S)AEs], (serious) treatment-related AEs*, AEs of special interest (AESIs), deaths and AEs leading to withdrawal from trial and/or permanent discontinuation from

to placebo with respect to the following endpoints:	treatment in the period between baseline and Week 24, and between baseline and Week 48
	• Development of on-treatment** antibodies to granulocyte macrophage colony stimulating factor (anti-GM-CSF), with antibody titers determined throughout 24 weeks and at 48 weeks of treatment
	• Changes in forced vital capacity (FVC), forced expiratory volume in one second (FEV_1), and FEV_1/FVC from baseline to Week 24 and Week 48
	• Changes in QT interval corrected by Fridericia (QTcF) and Bazett (QTcB) from baseline to Weeks 4 and 24

*The term "adverse drug reactions (ADRs)" that is used in the protocol has been replaced with the term "treatment-related" in this SAP.
**The term "treatment-boosted" that is used in the protocol has been replaced with the term "on-treatment" in this SAP.

Further details on AEs, on-treatment anti-GM-CSF antibody titers, spirometry, and ECGs are provided in the sections that follow.

2.3.1 ADVERSE EVENTS

AEs and SAEs are collected from the signing of the Informed Consent Form (ICF) until the Week 148 visit for subjects who consent or re-consent to protocol version 9.1 and the Week 100 visit for subjects who do not. Treatment-emergent AEs (TEAEs), defined in Section 9.5.4.1, will be presented in the summary tables. Medical Dictionary for Regulatory Activities (MedDRA) is used to code the AEs on an ongoing basis as data are collected, with the version used being updated every 6 months. The latest version to be used prior to the partial database lock and treatment unblinding (after all randomized subjects have either completed or discontinued during the double-blind treatment period) will be version 27.0 (i.e., the March 2024 version).

2.3.2 ON-TREATMENT ANTI-GM-CSF ANTIBODY TITERS

Blood sampling for assessment of diagnostic and on-treatment anti-GM-CSF antibodies is performed at Screening visit 1 and at the timepoints shown in the SoAs (Appendix B), respectively. Assays for diagnostic anti-GM-CSF antibodies are performed at three regional laboratories (in the U.S., Europe, and Japan). The cut-off values for diagnostic levels of anti-GM-CSF autoantibodies are according to clinical standards at the regional central laboratories.

The titer of anti-GM-CSF antibodies over the course of treatment until 4 weeks after end of treatment is analyzed at a central specialized laboratory.

2.3.3 SPIROMETRY

Spirometry is conducted prior to the DLCO, using the same equipment as for DLCO. FEV_1 and FVC are performed in accordance with ATS/ERS guidelines for spirometry testing (Miller et al. 2005). At least three acceptable and repeatable maneuvers according to ATS/ERS criteria are required, with up to eight maneuvers conducted during a session as needed. The largest FEV_1 and FVC, as determined by the overreader and not necessarily from the same maneuvers, will be used in the analysis. Standardization across sites is obtained by using the same equipment and implementing central

overread by a team of external, independent respiratory experts. Predicted values are calculated by the centrally provided equipment according to the Global Lung Function Initiative prediction equations (Quanjer et al. 2012). A manual is provided to all trial sites that describes the spirometry test in more detail.

2.3.4 ELECTROCARDIOGRAMS

Triplicate, resting, 12-lead ECGs are obtained using centrally provided ECG equipment at the visits shown in the SoAs (Appendix B). ECGs collected at Visit 3/Baseline, Visit 4/Week 4, and Visit 9/Week 24 are overread by a central ECG laboratory overseen by a cardiologist. RR, PR, QRS, and QT intervals are determined on 3 consecutive beats in Lead II, or an alternative lead if Lead II is not acceptable for measurement. Mean RR, PR, QRS, QT, and QTcF and QT corrected by Bazett (QTcB), over the 3 consecutive beats, are calculated by the central ECG laboratory. The cardiologist conducts standard interpretation of the ECGs as normal/abnormal and type of abnormality.

ECGs obtained at all other visits are not overread by a central ECG laboratory. For these ECGS, RR, PR, QRS, and QT intervals are computer-determined from a single beat by the supplied ECG equipment.

The triplicate ECGs at each visit are collectively assessed by the Investigator as normal; abnormal, not clinically significant (NCS); or abnormal, clinically significant (CS).

2.4 EXPLORATORY OBJECTIVES AND ENDPOINTS

Objectives	Endpoints
Efficacy	
Investigate the efficacy of molgramostim compared to placebo with respect to the following endpoints:	• Frequency of lung lavages (LLs) in the period from baseline to Weeks 24 and 48
	• Changes in SGRQ Impact score from baseline to Weeks 24 and 48
	• Changes in SGRQ Symptoms score from baseline to Weeks 24 and 48
	• Changes in distance walked during treadmill test from baseline to Weeks 24 and 48
	• Changes in duration of exercise during treadmill test from baseline to Weeks 24 and 48
	• Change in A-aDO_2 from baseline to Week 24 (included as Secondary endpoint for Japan and South Korea) and Week 48
	• Changes in arterial partial pressure of oxygen (PaO_2) from baseline to Weeks 24 and 48
	• Changes in disease severity score (DSS) from baseline to Weeks 24 and 48

	• Changes in ground glass opacification/opacities (GGO) from baseline to Week 24
	• Changes in Clinician's Global Impression of Severity (CGIS) from baseline to Weeks 24 and 48
	• Clinician's Global Impression of Change (CGIC) at Weeks 24 and 48
	• Changes in Patient's Global Impression of Severity (PGIS) from baseline to Weeks 24 and 48
	• Patient's Global Impression of Change (PGIC) at Weeks 24 and 48
	• Changes in supplemental oxygen use from baseline to Weeks 24 and 48
	• Changes in biomarker levels from baseline to Weeks 24 and 48
	• Changes in EuroQol 5 Dimensions, 5 Levels (EQ-5D-5L) from baseline to Weeks 24 and 48
	• Changes in dyspnea from baseline to Weeks 24 and 48
	Exploratory (Specifically for Japan and South Korea)
	• *Change in % predicted DLCO from baseline to Week 48*
	• *Change in SGRQ Total score from baseline to Week 48*
	• *Change in SGRQ Activity score from baseline to Week 48*
	• *Change in EC (expressed as peak METs) from baseline to Week 48*
Further Exploratory Longer-Term Objectives	
Efficacy	
Investigate the efficacy of molgramostim during open-label treatment with respect to the following endpoints:	• Frequencies of LLs in the period from baseline to the end of open-label treatment
	• Changes in % predicted DLCO, SGRQ (Total, Activity, Impact and Symptoms), A-aDO_2, PaO_2, DSS, CGIS, PGIS, supplemental oxygen use, biomarker levels and EQ-5D-5L from baseline to the end of open-label treatment
	• CGIC and PGIC at the end of open-label treatment
Safety	

Investigate the safety of molgramostim after 96-week treatment with respect to the following endpoints:	• Frequencies of (S)AEs, serious treatment-related AEs, AESIs, deaths and AEs leading to withdrawal from the trial and/or permanent discontinuation from treatment in the period between baseline and the end of open-label treatment
	• Development of on-treatment anti-GM-CSF antibody titers during double-blind and open-label treatment and 4 weeks' post-treatment
	• Changes in FVC, FEV_1, FEV_1/FVC from baseline to the end of open-label treatment

The following are exploratory endpoints that are not specified in the protocol but are of interest, in addition to the ones listed above:

- Changes from baseline in absolute $DLCO_{adj}$
- Changes from baseline in oxygen saturation (SpO_2)

Further details on DSS, GGO, CGIS, CGIC, PGIS, PGIC, supplemental oxygen use, biomarkers, EQ-5D-5L, and dyspnea are provided in the sections that follow.

2.4.1 DISEASE SEVERITY SCORE

As defined by Inoue et al (Inoue et al. 2008), and elsewhere, the DSS score ranges from 1 (least severe) to 5 (most severe) and is based on subject symptoms and PaO_2 (assessed at rest on room air) as follows:

- DSS 1: PaO_2 ≥70 mmHg without symptoms
- DSS 2: PaO_2 ≥70 mmHg with symptoms
- DSS 3: 70 mmHg > PaO_2 ≥60 mmHg
- DSS 4: 60 mmHg > PaO_2 ≥50 mmHg
- DSS 5: 50 mmHg > PaO_2

2.4.2 GROUND GLASS OPACITY

As a measure of surfactant accumulation, GGO is assessed. For this purpose, a high-resolution computed tomography (HRCT) scan is collected at Screening visit 2 and Week 24 (and at early withdrawal if that occurs during the double-blind treatment period). The extent of GGO is scored in 3 zones (upper, middle, and lower) on a scale from 0-5:

0: No GGO
1: <5% GGO
2: 5-24% GGO
3: 25-49% GGO
4: 50-74% GGO
5: ≥75% GGO

The GGO scores are assessed by two independent, central readers with expertise in radiological diagnosis of PAP who are blinded to subjects' treatment assignment and sequence of the assessed scans. Furthermore, based on the local radiologic assessment, the HRCT scans are classified by the Investigator as normal, having a non-clinically significant abnormality, or having a clinically significant abnormality.

2.4.3 CLINICIAN'S GLOBAL IMPRESSION OF SEVERITY AND CHANGE

As measures of overall clinician-rated disease severity and treatment response, the Investigator assesses CGIS and CGIC. The current severity of aPAP (CGIS) is assessed on a five-point scale ranging from none, mild, moderate, severe, to very severe with none = 1 and very severe = 5. The change from baseline in aPAP severity (CGIC) is assessed on a five-point scale ranging from much improved, to somewhat improved, no change, somewhat worse, and much worse, with much improved = 1 and much worse = 5.

2.4.4 PATIENT'S GLOBAL IMPRESSION OF SEVERITY AND CHANGE

PGIS and PGIC are assessed in the eDiary in relation to and immediately after the SGRQ and exercise treadmill test, respectively. PGIS assesses current breathing problems and the impact of these on daily physical activity (in relation to the SGRQ) or the current exercise ability (in relation to the exercise treadmill test). PGIC assesses the same issues as PGIS but as changes from baseline, i.e., change from baseline in breathing problems and impact of these on daily physical activity or the change from baseline in exercise ability. In addition, at Week 12, Week 24, and Week 48, subjects reporting worsening or improvement from baseline in PGIC are asked if the change was important to them or not.

For the PGIS and PGIC completed in relation to and immediately after the SGRQ, the items to be rated include the following:

- Please choose the response that best describes the severity of your breathing problems (none, mild, moderate, severe, very severe)
- Please choose the response that best describes the overall change in your breathing problems since you started taking the study medication (much better, a little better, no change, a little worse, much worse)
- Was this change in your breathing problems important for you (Yes/No)
- Please choose the response that best describes how much your breathing problems limit your daily physical activity level (not at all, slightly, moderately, strongly, very strongly)
- Please choose the response that best describes the overall change in your daily physical activity level since you started taking the study medication (much better, a little better, no changes, a little worse, much worse)
- Was this change in your daily physical activity level important for you (Yes/No)

For the PGIS and PGIC completed in relation to and immediately after the exercise treadmill test, the items to be rated include the following:

- Please choose the response that best describes how much your breathing problems limit your ability to exercise (not at all, slightly, moderately, strongly, very strongly)
- Please choose the response that best describes the overall change in your ability to exercise since you started taking the study medication (much better, a little better, no change, a little worse, much worse)
- Was this change in your ability to exercise important for you (Yes/No)

2.4.5 SUPPLEMENTAL OXYGEN USE

Subjects on supplemental oxygen are asked to complete a daily oxygen diary in the eDiary for the 14-day period prior to a visit. The eDiary captures information on oxygen flow at rest, during sleep, and during exertion and hours of oxygen use during exertion. The data from the eDiary are to be combined into an oxygen index approximating the average use in liters per minute, assuming that oxygen use during rest equates to 24 hours per day, oxygen use during sleep to 8 hours per day, and oxygen use during exertion to the reported number of hours.

At each visit, the Investigator enters the following information in the electronic Case Report Form (eCRF), based on the period since the last visit:

- Oxygen use during exertion, sleep, and rest
- Flow rates

The need for supplemental oxygen is evaluated by the Investigator at each visit; criteria for using oxygen are entered in the eCRF:

- Resting PaO_2 ≤55 mmHg (7.3 kPa)
- Resting PaO_2 ≤60 mmHg (8kPa) and hematocrit (Hct) ≥55%
- Other

2.4.6 BIOMARKER LEVELS

The following aPAP-related biomarkers are assessed:

- Krebs von den Lungen-6 (LK-6)
- Cytokeratin 19 fragments (CYFRA 21-1)
- Carcinoembryonic antigen (CEA)
- Lactate dehydrogenase (LDH)
- Hgb and Hct

Assays of these blood-based biomarkers are performed by a central laboratory using validated methods. LDH is analyzed from the standard biochemistry safety laboratory sample, and Hgb and Hct from the standard hematology safety laboratory sample.

2.4.7 EUROQOL 5 DIMENSIONS 5 LEVELS

Subjects should complete the EQ-5D-5L questionnaire in the eDiary, after the SGRQ and before other trial assessments.

The EQ-5D-5L is a generic, multidimensional, health-related quality-of-life instrument that comprises a Visual Analogue Scale (VAS) and a short descriptive system questionnaire.

The VAS records the subject's overall current health on a vertical VAS where the endpoints are labelled "The best health you can imagine" and "The worst health you can imagine". The VAS provides a quantitative measure of the subject's self-perception of their overall health.

The descriptive system allows subjects to rate their health in 5 domains: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression using 5-level scales indicating: 1. No problem, 2. Slight problem, 3. Moderate problem, 4. Severe problem, or 5. Unable to/extreme problems. The perceived problem levels for each domain are combined into a 5-digit health state that can be converted to an index value that reflects how good or bad a health state is according to the preferences of the general population, with higher scores indicating better quality of life.

2.4.8 DYSPNEA

Dyspnea is assessed by the Borg CR Scale® at multiple points during the treadmill test. Only the Borg CR Scale (dyspnea) scores at the rest stage (pre-test) and at the last stage of the treadmill test (post-test) are entered in the eCRF. In addition to the score at each of the two stages. the difference in the post-test Borg CR Scale® (dyspnea) score minus the pre-test Borg CR Scale® (dyspnea) score will be used for endpoint analysis. Dyspnea score ranges from 0 (=nothing at all) to 10 (=absolute maximum).

3. TRIAL DESIGN

3.1 OVERALL DESIGN

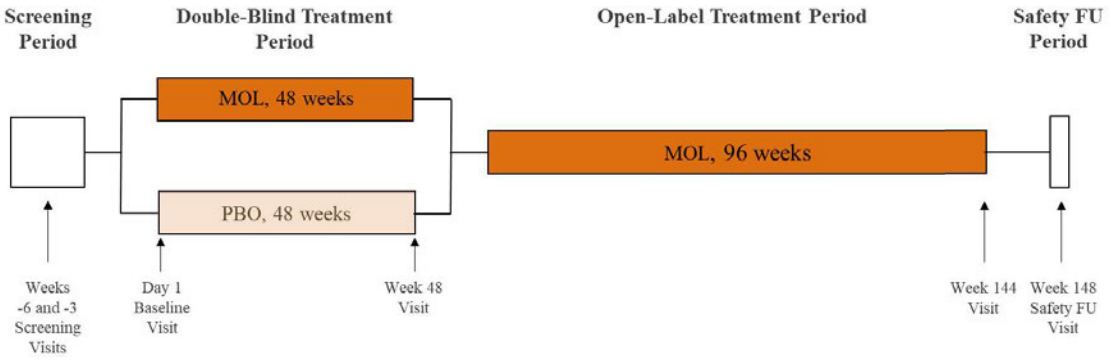
- This is an interventional, randomized, double-blind, 2-arm, parallel, placebo-controlled, multi-center, phase 3 trial in adult subjects who are diagnosed with aPAP.
- 160 subjects are planned to be randomized, and the randomization is intended to be stratified by baseline diffusing capacity of the lungs for carbon monoxide ($DLCO_{adj}$; >50% or ≤50% predicted) and by region (Asia and Australia, Europe including Turkey, or North America).
- aPAP diagnosis is confirmed by an anti-GM-CSF autoantibody test result, and history of PAP based on either high-resolution computed tomography, lung biopsy, or bronchoalveolar lavage cytology.
- The $DLCO_{adj}$ must be ≤70% predicted and the absolute change in the % predicted $DLCO_{adj}$ should be <15% points during the screening period. The subject should have a stable resting SpO_2)>85% without use of supplemental oxygen.
- The trial consists of a 6-week screening period, a 48-week randomized, double-blind treatment period, a 96-week open-label treatment period, and a 4-week safety follow-up period. Taking the scheduled visit windows of ± 7 days into account, the maximum treatment duration is 145 weeks, and the maximum trial duration is 156 weeks.
- Two screening visits are conducted at 6 and 3 weeks prior to the Baseline visit. At the Baseline visit, eligible subjects are centrally randomized through an Interactive Response Technology (IRT) to 48-week double-blind once-daily treatment with either molgramostim

300 µg nebulizer solution or placebo nebulizer solution. The treatment assignment is stratified according to the baseline % predicted $DLCO_{adj}$ and region.

- Subjects who complete the double-blind treatment period continue into the open-label treatment period where they receive open-label once-daily treatment with molgramostim.
- During the trial, LLs are allowed as rescue treatment in case of worsening of aPAP findings.

The overall trial design is presented in Figure 1 below.

Figure 1: Trial Design Schema



MOL = Molgramostim 300 µg nebulizer solution; PBO = Placebo nebulizer solution; FU = Follow-up

3.2 SAMPLE SIZE DETERMINATION

Considering $DLCO_{adj}$ data from MOL-PAP-002, the sample size was estimated for the primary endpoint assuming a conservative estimate of treatment effect for change from baseline to Week 24 in $DLCO_{adj}$ of 5.7 percentage points, a standard deviation (SD) of 11 percentage points and a significance level of 5%. With a total of 160 randomized subjects, assigned at a 1:1 ratio to double-blind treatment per treatment group, the power to reject the null hypothesis of no treatment effect on $DLCO_{adj}$ is 90%.

A blinded sample size re-assessment was conducted by Parexel in conjunction with a scheduled Data Monitoring Committee (DMC) review after the first 80 subjects completed the Week 24 visit. The re-assessment did not result in a change in sample size estimate. The Sponsor remained blinded to the treatment allocations throughout the process of sample size re-estimation. The blinded sample size re-estimation procedure is outside the scope of this SAP and is covered in the SAP for the analysis of data for the DMC. A report that documents the sample size re-assessment process and results is included in the trial master file.

4. ANALYSIS POPULATIONS

Analysis populations or sets include the Full analysis set (FAS), Per-protocol set (PPS), Safety analysis set (SAFS), and 24-Week Completer Analysis Set. The FAS is considered the primary efficacy analysis population. All efficacy summaries and endpoint analyses will be performed on the FAS. Analysis of primary and secondary endpoints will also be performed on the PPS. Safety summaries and safety endpoint analyses will be performed on the SAFS.

4.1 FULL ANALYSIS SET

The full analysis set (FAS) will include all randomized subjects, with treatment group assigned in accordance with randomization, regardless of treatment received. Subjects who are randomized but do not subsequently receive treatment (molgramostim or placebo) are included in the FAS.

4.2 SAFETY ANALYSIS SET

The safety analysis set (SAFS) will consist of all subjects who receive at least one dose of investigational medicinal product (IMP) (molgramostim or placebo). Safety data will be summarized and analyzed according to the treatment received, that is, erroneously treated subjects (e.g., those randomized to one arm but given the treatment of the alternate arm) will be summarized/analyzed according to the treatment they actually receive.

4.3 PER-PROTOCOL SET

The per-protocol set (PPS) will include all randomized subjects who complete 24 weeks of double-blind treatment for Visit 9/Week 24 endpoints or complete 48 weeks of double-blind treatment for Visit 11/Week 48 endpoints and are deemed to have no protocol deviations that could interfere with the primary and/or secondary efficacy objectives of this trial. Note that simply missing an assessment at a visit will not cause a subject to be excluded from the PPS.

A protocol deviation is defined as any change, divergence, or departure from the trial design or procedures defined in the protocol. Prospective approval of protocol deviations to inclusion or exclusion criteria, also known as protocol waivers or exemptions, was not permitted. Major protocol deviations are a subset of protocol deviations that are likely to have an impact on the subject's rights, safety, well-being, and/or on the validity of the data for analysis. A list of potential protocol deviations in the trial by category of deviation and their classifications as major or minor are provided in a separate document called the *Protocol Deviation Specification*. The SDTM datasets will include SDTM.DV containing required variables to summarize protocol deviations.

The identification of deviations from the protocol that will lead to exclusion of a subject or data points from the PPS will be made by the Sponsor in a blinded manner and documented before the double-blind part of the trial database is locked and randomized treatment assignments are unblinded (after Visit 11/Week 48).

4.4 24-WEEK COMPLETER ANALYSIS SET

The 24-week Completer Analysis Set is defined as all randomized subjects who complete 24 weeks of double-blind treatment (through Visit 9/Week 24), do not permanently discontinue treatment prior to Visit 9/Week 24, and have non-missing data for the primary endpoint.

The 24-week Completer Analysis Set will be used to perform sensitivity analyses for the primary endpoint analysis, see Section 6.1 for further details on analyses conducted on this analysis set.

5. STATISTICAL HYPOTHESES

5.1 PRIMARY HYPOTHESIS

The primary null hypothesis is that there is no difference between molgramostim and placebo in the mean change from baseline to Week 24 in % predicted $DLCO_{adj}$.

The alternative hypothesis is that there is a difference between molgramostim and placebo in the mean change from baseline to Week 24 in % predicted $DLCO_{adj}$.

5.2 SECONDARY HYPOTHESIS

For all secondary endpoints in which statistical tests are to be conducted (Section 10.2.2), 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.

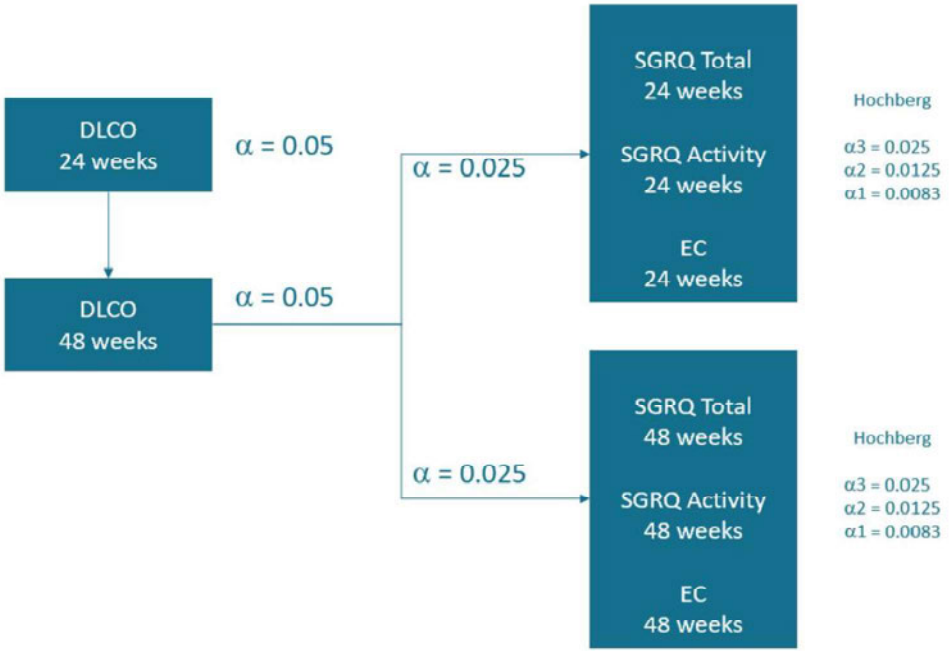
5.3 MULTIPLE TESTING STRATEGY

Due to the differences in the secondary endpoints between Japan and Korea (combined) and all other countries participating in the trial (Australia, countries in Europe including Turkey, and countries in North America), the multiple testing strategy differs between the two geographic groups.

5.3.1 MULTIPLE TESTING STRATEGY (FOR REGULATORY AUTHORITIES OUTSIDE OF JAPAN AND SOUTH KOREA)

There is one primary efficacy endpoint and seven secondary efficacy endpoints, which are intended to support conclusions based on the primary endpoint. A type I error-control procedure that uses a combination of sequential testing and alpha splitting will be used for analysis of these efficacy endpoints to maintain the overall type I error rate at 5%. The procedure is shown schematically in Figure 2 and described beneath the figure.

Figure 2. Multiple Testing Strategy (for regulatory authorities outside of Japan and South Korea)



1. DLCO at Week 24 is tested at a two-sided *alpha* = 0.05 ("Family 1"). If DLCO at Week 24 is not statistically significant, then the procedure stops.
2. If DLCO at Week 24 is statistically significant, then DLCO at Week 48 ("Family 2") is tested at two-sided *alpha* = 0.05.
3. If DLCO at Week 48 is statistically significant, then the 5% alpha is split equally and passed to "Family 3" and "Family 4", which consist of the SGRQ Total, SGRQ Activity, and EC endpoints for Week 24 and Week 48, respectively.
4. The Hochberg method (Hochberg 1988) is applied to each of "Family 3" and "Family 4". For the Hochberg method, each p-value result from the family of tests is ranked from largest to smallest ($i = m, m-1,...1$). The corresponding critical value for comparison of hypothesis $H_{(i)}$ is computed as $\alpha_{(i)} = \alpha/(m - i + 1)$ where m = total number of tests within the family (3 in each case), $i = m,m-1...,1$, and $\alpha = 0.025$ as shown in the figure. When the first $p_{(i)} \leq \alpha_{(i)}$ for hypothesis $H_{(i)}$ $[i=m,m-1...,1]$, the comparison stops and then concludes that this hypothesis plus the remaining hypotheses will be rejected at significance level α.

5.3.2 MULTIPLE TESTING STRATEGY (FOR REGULATORY AUTHORITIES IN JAPAN AND SOUTH KOREA)

There is one primary efficacy endpoint and four secondary efficacy endpoints which are intended to support conclusions based on the primary parameter. A type I error-control procedure that uses sequential testing will be used for analysis of these four secondary endpoints. The type I error control procedure for these authorities are shown schematically in Figure 3 and described beneath the schematic. The multiple testing strategy for primary and secondary endpoints as detailed in Figure 3 will use all subjects belonging to the FAS, per the primary and secondary estimands (Section 6).

Figure 3. Multiple Testing Strategy for regulatory authorities in Japan and South Korea

1. DLCO at Week 24 is tested at two-sided alpha = 0.05 ("Family 1"). If DLCO at week 24 is not statistically significant, then the procedure stops.
2. If DLCO at Week 24 is statistically significant, then the 0.05 alpha is passed to "Family 2" which consists of SGRQ Total, SGRQ Activity, EC, and A-aDO_2 endpoints for Week 24.
3. The Hochberg procedure (Hochberg 1988) is applied to Family 2 as described in Section 5.3.1 but with *m*=4 and α=0.05.

6. TRIAL ESTIMANDS

The following sections describe the attributes of the estimands that will be used for evaluation of the primary efficacy endpoint and secondary efficacy endpoints.

6.1 ESTIMANDS FOR THE PRIMARY EFFICACY ENDPOINT

The primary analysis for the primary efficacy endpoint of change from baseline to Week 24 in % predicted $DLCO_{adj}$ will be performed on the FAS using treatment policy strategy to handle intercurrent events, control-based imputation to handle missing data, and a mixed model for repeated measurements (MMRM) to estimate the difference between the two treatment groups in the mean change from baseline to Week 24 in % predicted $DLCO_{adj}$. Refer to the estimand-to-analysis table below for details on this primary analysis and for descriptions of sensitivity and supplemental analyses that will also be performed.

Primary Objective: To investigate the efficacy of molgramostim compared to placebo with respect to change in % predicted $DLCO_{adj}$ from Baseline to Week 24	
Estimand: The effect of inhaled molgramostim nebulizer solution on % predicted $DLCO_{adj}$ as an index of pulmonary gas transport in subjects with aPAP	
Treatment: molgramostim nebulizer solution 300 µg/1.2 mL administered once daily	
Estimand	**Analysis**
Target Population: Subjects with % predicted $DLCO_{adj}$ ≤70% with diagnosed aPAP and positive anti-GM-CSF antibody titers	**Analysis Set:** FAS. Subjects assigned to molgramostim through randomization will be the active treatment group. Subjects assigned to placebo will be the comparator group. Sensitivity Analysis Sets: • PP. • 24-week Completer Analysis Set (Note: Analysis will be performed only if ≤90% of the FAS belongs to the 24-week Completer Analysis Set).
Variable	**Outcome Measure**
% predicted $DLCO_{adj}$	Change from Baseline to Week 24 in % predicted $DLCO_{adj}$.
Intercurrent Event Handling	**Missing Data Strategy**
Treatment policy strategy will be used for all intercurrent events, including but not limited to the following: • Premature treatment discontinuation: All data after treatment discontinuation will still be included. • Post-baseline LL or use of supplemental oxygen that starts after baseline: All data irrespective of these events will be included. • Occurrence of coronavirus disease 2019 (COVID-19)-related TEAE: All data irrespective of COVID-19-related TEAE will be included. • Study treatment non-compliance: All data will be used irrespective of compliance. Sensitivity Analyses (on the FAS only): • Composite strategy will be applied for premature treatment discontinuation, post-baseline LL, or use of supplemental oxygen that starts after baseline. Subjects with these events will be considered as	Control-based imputation, as described in Section 9.3, will be used to handle missing % predicted $DLCO_{adj}$ during the double-blind treatment period including but not limited to the following: • Missing data due to subject discontinuing study treatment or discontinuing from the study. • Missing data for a subject who misses a visit or is present at a visit but the DLCO cannot be adequately conducted or the DLCO is outside the analysis window for the visit. Sensitivity Analyses (on the FAS only and using treatment policy strategy to handle intercurrent events): • No imputation of missing data. • Conservative imputation under a missing not at random (MNAR) assumption to facilitate a tipping point analysis, as described in Section 9.4, will be used to handle missing % predicted $DLCO_{adj}$ during the double-blind treatment period.

"treatment failures" only for the purpose of this sensitivity analysis. All non-missing data collected during the double-blind treatment period and after the intercurrent event for the subject will be replaced with the worst value for that subject up to the end of the double-blind treatment period. For all other intercurrent events, the treatment policy strategy will be used. • An alternative composite strategy will also be applied that is similar to the one above, except data after post-baseline LL for any subject will be replaced with a single "poor value", namely, the worst value observed for all randomized subjects at any time from baseline up to the end of the double-blind treatment period. Only observed values within ± 3 x SD of the overall mean (of all the observed values) will be considered in determining the single "poor value". For all other intercurrent events, the treatment policy strategy will be used. • Hypothetical strategy will be applied for premature treatment discontinuation, post-baseline LL, or use of supplemental oxygen that starts after baseline. All non-missing data collected during the double-blind treatment period and after the intercurrent event will be censored, i.e., considered missing at random (MAR) and, therefore, imputed via multiple imputation. For all other intercurrent events, the treatment policy strategy will be used. Supplemental Analysis (on the FAS only): • Hypothetical strategy will be applied for occurrence of COVID-19-related TEAE and for premature treatment discontinuation due to COVID-19-related TEAE. All non-missing data, during the double-blind treatment period, collected from the onset of the TEAE until its resolution will be considered MAR. For all other intercurrent events, the treatment policy strategy will be used.	Supplemental Analysis (on the FAS only and using hypothetical strategy to handle COVID-19-related TEAEs): • Missing data due to subject withdrawal resulting from a COVID- 19-related TEAE will be imputed via multiple imputation under a MAR assumption as described in Section 9.3. All other missing data will be handled via control-based multiple imputation as described in Section 9.3.
Population-level Summary Measure	**Analysis Approach**
Difference between the two treatment groups in mean change from Baseline to Week 24 in % predicted $DLCO_{adj}$	The difference will be estimated using least squares mean (LSMean) changes in % predicted $DLCO_{adj}$ from

	Baseline to Week 24 based on a MMRM, as described in Section 10.2.1. Sensitivity Analyses (on the FAS only, using treatment policy strategy to handle intercurrent events, and using control-based imputation to handle missing data): • An analysis of covariance (ANCOVA) model will be used, as described in Section 10.2.1.2. • A nonparametric approach, van Elteren test, will be used, as described in Section 10.2.1.2, if data appear to be non-normally distributed. • The primary and sensitivity analysis approaches will be repeated using DLCO severity stratification at randomization based on the Hgb values at Visit 3/Baseline rather than those at Visit 2/Screening visit 2, if more than 5% of the randomized subjects would have been assigned to a different stratum.

For a subject already using supplemental oxygen at baseline, an increase in the volume of oxygen use after baseline will not be considered an intercurrent event.

COVID 19-related TEAEs will be identified using the list of COVID-19-related terms provided by MedDRA (COVID-19-related New Terms MedDRA 23.0 – 25.1 Spreadsheet), accessible via the MedDRA website.

6.2 ESTIMANDS FOR THE SECONDARY EFFICACY ENDPOINTS

The primary analyses for the secondary efficacy endpoints will be performed on the FAS using treatment policy strategy to handle intercurrent events, control-based imputation to handle missing data, and a mixed model for repeated measurements (MMRM) to estimate the difference between the two treatment groups in the mean change from baseline to Week 24 or Week 48 in variables of interest. Refer to the estimand-to-analysis tables below for details on these primary analyses and for descriptions of sensitivity analyses that will also be performed.

6.2.1 CHANGE IN % PREDICTED $DLCO_{ADJ}$ FROM BASELINE TO WEEK 48 (OUTSIDE OF JAPAN AND SOUTH KOREA)

The estimand-to-analysis table for the secondary efficacy endpoint of change in % predicted $DLCO_{adj}$ from baseline to Week 48 is the same as the one for the primary efficacy endpoint but with Week 24 replaced with Week 48 and excluding the sensitivity analysis based on the 24-week completer analysis set. Note that this secondary efficacy endpoint applies only to regulatory authorities outside of Japan and South Korea.

6.2.2 CHANGES IN SAINT GEORGE'S RESPIRATORY QUESTIONNAIRE TOTAL SCORE AND ACTIVITY SCORE FROM BASELINE TO WEEK 24 AND WEEK 48

Secondary Objectives: To investigate the efficacy of molgramostim compared to placebo with respect to changes in SGRQ Total Score and SGRQ Activity Score from Baseline to Week 24 and Week 48 *Note: The secondary objectives corresponding to change from Baseline to Week 48 in SGRQ Total Score and SGRQ Activity Score apply only to regulatory authorities outside of Japan and South Korea.*	
Estimand: The effect of inhaled molgramostim nebulizer solution on the SGRQ Total Score and Activity Score in subjects with aPAP	
Treatment: molgramostim nebulizer solution 300 µg/1.2 mL administered once daily	
Estimand	**Analysis**
Target Population: Subjects with % predicted $DLCO_{adj}$ ≤70% with diagnosed aPAP and positive anti-GM-CSF antibody titers	**Analysis Set:** FAS. Subjects assigned to molgramostim through randomization will be the active treatment group. Subjects assigned to placebo will be the comparator group. Sensitivity Analysis Set: • PP
Variable	**Outcome Measure**
• SGRQ Total Score • SGRQ Activity Score	• Change from Baseline to Week 24 in SGRQ Total Score • Change from Baseline to Week 24 in SGRQ Activity Score • Change from Baseline to Week 48 in SGRQ Total Score • Change from Baseline to Week 48 in SGRQ Activity Score
Intercurrent Event Handling	**Missing Data Strategy**
Treatment policy strategy will be used for all intercurrent events, including but not limited to the following: • Premature treatment discontinuation: All data after treatment discontinuation will still be included. • Post-baseline LL or use of supplemental oxygen that starts after baseline: All data irrespective of these events will be included. • Study treatment non-compliance: All data will be used irrespective of compliance. Sensitivity Analysis (on the FAS only): • Composite strategy will be applied for premature treatment discontinuation, post-	Control-based imputation, as described in Section 9.3, will be used to handle missing SGRQ Total Score and SGRQ Activity Score during the double-blind treatment period, including but not limited to the following: • Missing data due to subject discontinuing study treatment or discontinuing from the study. • Missing data for a subject who misses a visit or is present at a visit, but the SGRQ is not conducted or the date of the SGRQ is outside the analysis window for the visit.

baseline LL, or use of supplemental oxygen that starts after baseline. Subjects with these events will be considered as "treatment failures" only for the purpose of this sensitivity analysis. All non-missing data collected during the double-blind treatment period and after the intercurrent event for the subject will be replaced with the worst value for that subject up to the end of the double-blind treatment period. For all other intercurrent events, the treatment policy strategy will be used. • An alternative composite strategy will also be applied that is similar to the one above, except data after post-baseline LL for any subject will be replaced with a single "poor value", namely, the worst value observed for all randomized subjects at any time from baseline up to the end of the double-blind treatment period. Only observed values within ± 3 x SD of the overall mean (of all the observed values) will be considered in determining the single "poor value". For all other intercurrent events, the treatment policy strategy will be used.	Sensitivity Analyses (on the FAS only and using treatment policy strategy to handle intercurrent events): • No imputation of missing data. • Conservative imputation under a MNAR assumption to facilitate a tipping point analysis, as described in Section 9.4, will be used to handle missing SGRQ Total Score and SGRQ Activity Score during the double-blind treatment period.
Population-level Summary Measure	**Analysis Approach**
• Difference between the two treatment groups in mean change from Baseline to Week 24 in SGRQ Total Score • Difference between the two treatment groups in mean change from Baseline to Week 24 in SGRQ Activity Score • Difference between the two treatment groups in mean change from Baseline to Week 48 in SGRQ Total Score • Difference between the two treatment groups in mean change from Baseline to Week 48 in SGRQ Activity Score	The difference will be estimated using LSMean changes in SGRQ Total Score and SGRQ Activity Score from Baseline to Week 24 and Week 48 based on a MMRM, as described in Section 10.2.1 and Section 10.2.2.2. Sensitivity Analyses (on the FAS only, using treatment policy strategy to handle intercurrent events, and using control-based imputation to handle missing data): • ANCOVA model will be used, as described in Section 10.2.1.110.2.1 and Section 10.2.2.2. • A nonparametric approach, van Elteren test, will be used as described in Section 10.2.1.1 and Section 10.2.2.2, if data appear to be non-normally distributed.

6.2.3 CHANGES IN EXERCISE CAPACITY (EXPRESSED AS PEAK METS) FROM BASELINE TO WEEK 24 AND WEEK 48

Secondary Objectives: To investigate the efficacy of molgramostim compared to placebo with respect to changes in EC (expressed as peak METs) from Baseline to Week 24 and Week 48 *Note: **The secondary objective corresponding to change from Baseline to Week 48 applies only to regulatory authorities outside of Japan and South Korea.***
Estimand: The effect of inhaled molgramostim nebulizer solution on peak METs, a measure of the energy cost of activities and the functional capacity or exercise tolerance of subjects with aPAP
Treatment: molgramostim nebulizer solution 300 µg/1.2 mL administered once daily

Estimand	Analysis
Target Population: Subjects with % predicted $DLCO_{adj}$ ≤70% with diagnosed aPAP and positive anti-GM-CSF antibody titers	**Analysis Set:** FAS. Subjects assigned to molgramostim through randomization will be the active treatment group. Subjects assigned to placebo will be the comparator group. Sensitivity Analysis Set: • PP
Variable	**Outcome Measure**
EC (expressed as peak METs)	• Change from Baseline to Week 24 in peak METs • Change from Baseline to Week 48 in peak METs
Intercurrent Event Handling	**Missing Data Strategy**
Treatment policy strategy will be used for all intercurrent events, including but not limited to the following: • Premature treatment discontinuation: All data after treatment discontinuation will still be included. • Post-baseline LL or use of supplemental oxygen that starts after baseline: All data irrespective of these events will be included. • Study treatment non-compliance: All data will be used irrespective of compliance. Sensitivity Analysis (on the FAS only): • Composite strategy will be applied for premature treatment discontinuation, post-baseline LL, or use of supplemental oxygen that starts after baseline. Subjects with these events will be considered as "treatment failures" only for the purpose of	Control-based imputation, as described in Section 9.3, will be used to handle missing peak METs during the double-blind treatment period, including but not limited to the following: • Missing data due to subject discontinuing study treatment or discontinuing from the study • Missing data for a subject who misses a visit or is present at a visit, but the exercise treadmill test is not conducted, or the date of the test is outside the analysis window for the visit Sensitivity Analyses (on the FAS only): • No imputation of missing data • Conservative imputation under a MNAR assumption to facilitate a tipping point analysis, as described in Section 9.4, will be

this sensitivity analysis. All non-missing data collected during the double-blind treatment period and after the intercurrent event for the subject will be replaced with the worst value for that subject up to the end of the double-blind treatment period. For all other intercurrent events, the treatment policy strategy will be used. • An alternative composite strategy will also be applied that is similar to the one above, except data after post-baseline LL for any subject will be replaced with a single "poor value", namely, the worst value observed for all randomized subjects at any time from baseline up to the end of the double-blind treatment period. Only observed values within ± 3 x SD of the overall mean (of all the observed values) will be considered in determining the single "poor value". For all other intercurrent events, the treatment policy strategy will be used.	used to handle missing peak METs during the double-blind treatment period
Population-level Summary Measure	**Analysis Approach**
• Difference between the two treatment groups in mean change from Baseline to Week 24 in peak METs • Difference between the two treatment groups in mean change from Baseline to Week 48 in peak METs	The difference will be estimated using LSMean changes in peak METs from Baseline to Week 24 and Week 48 based on a MMRM, as described in Section 10.2.1 and Section 10.2.2.3. Sensitivity Analyses (on the FAS only and using control-based imputation to handle missing data): • ANCOVA model will be used, as described in in Section 10.2.1.1 and Section 10.2.2.3. • A nonparametric approach, van Elteren test, will be used as described in Section 10.2.1.1 and Section 10.2.2.3, if data appear to be non-normally distributed.

6.2.4 CHANGE IN ALVEOLAR-ARTERIAL OXYGEN DIFFERENCE FROM BASELINE TO WEEK 24 (FOR JAPAN AND SOUTH KOREA ONLY)

Secondary Objective: To investigate the efficacy of molgramostim compared to placebo with respect to change A-aDO_2 from Baseline to Week 24 *Note: The secondary objective applies only to regulatory authorities in Japan and South Korea.*
Estimand: The effect of inhaled molgramostim nebulizer solution on A-aDO_2, a measure of gas exchange in subjects with aPAP
Treatment: molgramostim nebulizer solution 300 µg/1.2 mL administered once daily

Estimand	Analysis
Target Population: Subjects with % predicted $DLCO_{adj}$ ≤70% with diagnosed aPAP and positive anti-GM-CSF antibody titers	**Analysis Set:** FAS. Subjects assigned to molgramostim through randomization will be the active treatment group. Subjects assigned to placebo will be the comparator group. Sensitivity Analysis Set: • PP
Variable	**Outcome Measure**
A-aDO_2	Change from Baseline to Week 24 in A-aDO_2
Intercurrent Event Handling	**Missing Data Strategy**
Treatment policy strategy will be used for all intercurrent events, including but not limited to the following: • Premature treatment discontinuation: All data after treatment discontinuation will still be included. • Post-baseline LL or use of supplemental oxygen that starts after baseline: All data irrespective of these events will be included. • Study treatment non-compliance: All data will be used irrespective of compliance. • Non-physiologic values, e.g., values from venous samples collected while on supplemental O_2. Sensitivity Analysis (on the FAS only): • Hypothetical strategy will be applied to the occurrence of non-physiologic values. All such values will be treated as MAR. For all other intercurrent events, the treatment policy strategy will be used. (Note: A blinded review of A-aDO_2 values will be performed to identify non-physiologic values).	Control-based imputation, as described in Section 9.3, will be used to handle missing peak METs during the double-blind treatment period including but not limited to the following: • Missing data due to subject discontinuing study treatment or discontinuing from the study. • Missing data for a subject who misses a visit or is present at a visit arterial blood gas assessment is not conducted or it is conducted but the sample is not analyzable, or the date of sample collection is outside the analysis window for the visit. Sensitivity Analyses (on the FAS only): • No imputation of missing data. • Conservative imputation under a MNAR assumption to facilitate a tipping point analysis, as described in Section 9.4, will be used to handle missing A-aDO_2 during the double-blind treatment period.
Population-level Summary Measure	**Analysis Approach**
Difference between the two treatment groups in mean change from Baseline to Week 24 in A-aDO_2	The difference will be estimated using LSMean changes in A-aDO_2 from Baseline to Week 24 based on a MMRM, as described in Section 10.2.1 and Section 10.2.2.4. Sensitivity Analyses (on the FAS only and using control-based imputation to handle missing data): • ANCOVA model will be used, as described in in Section 10.2.1.1 and Section 10.2.2.4.

	• A nonparametric approach, van Elteren test, will be used as described in Section 10.2.1.1 and Section 10.2.2.4, if data appear to be non-normally distributed.

7. TIMING OF PLANNED ANALYSES

7.1 INTERIM ANALYSIS

No interim analysis is planned, and thus, no early stopping rule for efficacy is needed.

An independent DMC was established to perform a safety evaluation during the trial. A DMC charter defines the primary responsibilities of the DMC, its membership, purpose and timing of the meetings, and procedures, including those for restricted access to unblinded data.

7.2 FINAL ANALYSES

The trial will be unblinded, and the final analysis of all data up to and including the end of the 48-week double-blind treatment period of the trial will be conducted after all randomized subjects have either completed or discontinued prior to the end of this period. This 48-week analysis (SAV006-05 Double-blind) will include the analysis of all primary and secondary efficacy endpoints. Prior to the database lock and treatment unblinding for this 48-week analysis, the data up to the end of the double-blind treatment period will be cleaned and reviewed in a blinded manner to resolve data queries, the major protocol deviations will be identified, including those that will lead to the exclusion of subjects from the PPS, and compositions of the analysis populations will be determined.

The final analysis of the data from the open-label treatment period (SAV006-05 Open-label) and the safety follow-up four weeks after the end of the open-label treatment period will be conducted after all subjects have either completed the study or discontinued during the open-label treatment period.

Certain data up to the end of the double-blind treatment period (Week-48) will be partially locked for the 48-week analysis and will remain locked for the remainder of the trial. Exceptions to data locking will be made for adverse events, hospitalizations and concomitant medications that are ongoing as of the end of the double-blind treatment period. A final database lock following the end of the open-label treatment period and safety follow-up will encompass all trial data.

8. GENERAL CONSIDERATIONS FOR DATA ANALYSES

8.1 STANDARD SUMMARY STATISTICS AND LISTINGS

The following general analysis principles will apply:

Descriptive statistics will be used for all variables, as appropriate, and will be presented by treatment group. Continuous variables will be summarized by the number of observations, mean, standard deviation, minimum, median, and maximum value. Categorical variables will be summarized by frequency counts and percentages for each category. Unless otherwise stated, percentages will be calculated derived from the population total for the corresponding treatment group. All descriptive summaries will be based on observed values only, i.e., imputed values will not be included in the descriptive summaries.

To describe incidence of TEAEs, exposure-adjusted incidence rates (per 100 patient-years) of first TEAE occurrence will be calculated as the number of subjects exposed to the randomized treatment and experiencing a certain TEAE divided by the total exposure time (years) of all subjects who are at risk for the event (multiplied by 100 years). Specifically, for subjects with no event, the exposure time is the time from the first IMP dose to the last follow-up assessment; for subjects with at least one event, the exposure time is the time from the first IMP dose to first event.

For continuous data, the mean, median, and their associated confidence intervals (CIs) will be rounded to 1 additional decimal place compared to the original data. The standard deviation will be rounded to 2 additional decimal places compared to the original data. Minimum and maximum will be displayed with the same number of decimal places as the original data.

For categorical data, percentages will be rounded to 1 decimal place.

LSMeans and associated CIs will be rounded to 1 additional decimal place compared to the original data. Odds ratios, hazard ratios, and associated CIs will be rounded to 1 decimal place.

For all significance tests, p-values will be displayed to 4 decimal places.

Data collected in the trial database (including the trial eCRF and data from third party vendors used in the summaries and analyses) will generally be listed. Listings will primarily be sorted by the treatment group, site, subject-number, visit, and any other natural ordering related to the assessment.

If a date is recorded on the eCRF, the date and relative trial day (in relation to the date of the first dose of IMP) will be printed in the corresponding listing. If both a start and stop date are recorded, a duration will be included in the listings.

8.2 STRATA AND COVARIATES

The treatment randomization is intended to be stratified by % predicted $DLCO_{adj}$ at Visit 3/Baseline and by region. Thus, subjects will be randomized in a 1:1 ratio to treatment arms and stratified according to whether they have a $DLCO_{adj}$ of >50% predicted or ≤50% predicted. At the time of Visit 3/Baseline, the Visit 3 Hgb value is not yet known, so for the purposes of randomization, the Visit 2/Screening 2 Hgb value is used to determine the subject's Hgb-adjusted DLCO, and therefore, the subject's DLCO stratum. Furthermore, subjects will be regionally stratified into 3 regions based on the locations of the sites: [Asia and Australia], [Europe including Turkey], or North America.

8.3 STANDARD COMPARISON METHODS

Comparisons methods will be detailed under the appropriate subsections of Section 10.

8.4 STATISTICAL SIGNIFICANCE

Comparisons between molgramostim and placebo treatments will use placebo as the reference group. Change from baseline $DLCO_{adj}$ at Week 24 will be tested at a two-sided *alpha* = 0.05. A multiple testing strategy is defined in Section 5.3 for the secondary endpoints.

8.5 EXAMINATION OF SUBGROUPS

Subgroup analyses may be conducted for the primary endpoint of change in % predicted $DLCO_{adj}$ from baseline to Week 24, the secondary endpoints of change in % predicted $DLCO_{adj}$ from baseline to Week 48 and changes in SGRQ Total score, SGRQ Activity score, EC as measured in peak METs, A-aDO_2 (for regulatory authorities in Japan and South Korea) from baseline to Week 24 and from

baseline to Week 48 (except for A-aDO_2), and the safety endpoint of incidence of TEAEs during the first 24 weeks of the double-blind treatment period and during the entire 48-week double-blind treatment period. Plausible subgroups have some predictive biological rationale for an interaction with treatment and have been included in the trial stratification factors.

Plausible subgroups:

- Geographic region – Asia and Australia, Europe including Turkey, and North America
    - A subgroup of the subjects from Japan and South Korea only will also be analyzed.
- Visit 3/Baseline $DLCO_{adj}$ – $DLCO_{adj}$ >50% predicted, $DLCO_{adj}$ ≤50%, where the Visit 2/Screening 2 Hgb value was used to calculate $DLCO_{adj}$
    - Subjects will also be categorized based on their actual baseline $DLCO_{adj}$, that is, the baseline Hgb value, rather than the Visit 2/Screening 2 Hgb is used to calculate $DLCO_{adj}$. If more than 5% of subjects in either treatment group ends up being assigned to a stratum different from their randomization stratum assignment, then subgroup analysis based on their actual baseline $DLCO_{adj}$ category will also be performed.

Exploratory subgroups:

- Sex (male vs. female)
- Age at informed consent (≥18<40, 40 to <65, and ≥65 years)
- Race (Asian, White, All Other Race Categories)
- Smoking status at screening (current smoker, previous smoker and never smoked)

Subgroups may also be formed based on other baseline variables if there is a clinical justification, or an imbalance is observed between the treatment groups. If a baseline imbalance is observed between treatment groups, ad-hoc subgroup analysis may be used to investigate any potential for impact on the main results. Subgroup analyses will be limited to subgroups with a size of at least 25% of the FAS (for efficacy) or SAFS (for TEAEs), with the exception of subgroups based on geographic region, sex, and age.

No adjustment to the significance level for testing of the subgroup analyses will be made, since all these analyses will be considered supportive of the analyses on the overall population.

9. DATA HANDLING CONVENTIONS

9.1 BASELINE, FIRST DOSE DATE, AND LAST DOSE DATE

There are two treatment periods in the study: the double-blind treatment period (main period of interest) and the open-label treatment period. Within the double-blind treatment period, the 24-week timepoint is of primary interest and the 48-week timepoint is of secondary interest; however, both time points will be important to establish efficacy and durability of response. For subjects randomized to molgramostim and who receive molgramostim in both periods, the two periods combined (i.e., the entire molgramostim treatment period) is of interest. Baseline visit, Baseline value, first dose date, and last dose date are defined for each of the two periods, or the two periods combined, as described below:

- Double-blind Treatment Period
  - Baseline visit is Visit 3/Baseline.
  - Baseline value is the pre-dose value at Visit 3/Baseline; if this value is scheduled to be collected at this visit but is missing or inadvertently collected post-dose, then the last non-missing value prior to Visit 3/Baseline, if there is any, will be used as the Baseline value. For procedures that are not performed at Visit 3/Baseline but at an earlier visit (e.g., ETT and HCRT at Visit 3/Screening 2), then the baseline value is the last non-missing value prior to Visit 3/Baseline.
  - First dose date (and the start of the treatment period) is the date of the first dose of double-blind IMP. This first dose is scheduled to be taken during Visit 3/Baseline, but there may be a few instances where the first dose is taken at an unscheduled visit a few days after Visit 3/Baseline.
  - Last dose date is the date of completion/discontinuation of IMP in the double-blind treatment period (as entered in the End of Double-blind Treatment eCRF).
- Open-label Treatment Period
  - Baseline visit is Visit 11/Week 48.
  - Baseline value is the pre-dose value at Visit 11/Week 48. If this value is missing or inadvertently collected post-dose, then the last non-missing value prior to Visit 11/Week 48 will be used as the Baseline value.
  - First dose date (and the start of the treatment period) is the date of the first dose of open-label IMP. This first dose is scheduled to be taken during Visit 11/Week 48.
  - Last dose date is the date of completion/discontinuation of IMP in the open-label treatment period (as entered in the End of Open-label Treatment eCRF).
- Two Treatment Periods Combined (applicable only to subjects randomized to molgramostim)
  - Baseline visit, Baseline value, and First dose date are the same as the one for the double-blind treatment period above.
  - Last dose date is the same as the one for the double-blind treatment period above for subjects who did not participate in the open-label treatment period.
  - Last dose date is the same as the one for the open-label treatment period above for subjects who participated in the open-label treatment period.

9.2 PREMATURE WITHDRAWAL AND MISSING DATA

For subjects who withdraw early from the trial, the Investigator is instructed to attempt to collect the assessments as shown on the SoA (Appendix B). These assessments will be allocated to visits according to the analysis visit windows for that assessment (Appendix C).

For the primary and secondary endpoint analyses (Section 10.2.1 and Section 10.2.2), missing endpoint data will be imputed using control-based imputation, a method based on the assumption that subjects that discontinue IMP will have a similar response profile to subjects in the control group. Details of control-based imputation methods are described in Section 9.3.

There will be no missing data imputation for safety analyses. Missing data for categorical variables will be included in the summary table as a category. Since missing data are not imputed, continuous variables with missing data will have analysis performed on observed data only.

9.3 CONTROL-BASED IMPUTATION FOR EFFICACY ENDPOINTS

Control-based imputation will be performed for missing data as specified in Section 9.2. Data imputation will be carried out in 2 stages, which are described in detail below. Stage 1 will impute all arbitrary missing data under a MAR assumption within the treatment group. Stage 2 will impute all monotone missing data using a control-based imputation method. The imputation process will use pattern mixture models to create predictive posterior distributions which will generate imputed values for the missing data and will be facilitated through the SAS procedure PROC MI.

The imputation models will include conditional variables per the primary analysis model.

Stage 1 – imputing arbitrary missing data patterns

When imputing arbitrary missing data under a MAR assumption, each treatment group will be imputed separately within the SAS MI procedure. In order to impute all missing data across trial visits, the input data must be in wide format; one record per subject with trial visits (both observed and missing) represented as columns, for each subject. Example MI procedure SAS code for the primary endpoint is provided in Appendix F.

Stage 2 – Applying control-based imputation

Following stage 1, the output dataset will consist of 50 complete datasets of imputed data. To re-impute data for visits post-trial withdrawal/intercurrent event (where applicable) using control-based imputation, the missing assessments must be set back to missing in the read-in dataset, for each subject. Example code for control-based imputation at stage 2 for the primary endpoint is provided in Appendix F. For the supplemental estimand strategy to handle missing data following the occurrence of a COVID-19-related TEAE, the imputed data from Stage 1 study withdrawal due to a COVID-19-related TEAE should not be reset to missing, in order to preserve the MAR assumption for these particular visits.

The MI procedure defined for the control-based imputation (stage 2) is repeated on each of the 50 datasets produced from stage 1 imputation. Due to uncertainty around the within-subject correlation across visits under a jump to reference framework, the % predicted $DLCO_{adj}$ changes from baseline at previous visits are excluded from the imputation model at stage 2.

Analysis of the imputed datasets

To analyze the 50 complete datasets resulting from imputation stages 1 and 2, the output datasets must be converted to long format (one record per subject and visit) to be analyzed. A separate analysis model will be fitted to each of the 50 complete datasets. Example SAS code for the primary endpoint used to analyze the complete data is provided in Appendix F.

Least squares mean estimates and associated standard errors of the treatment differences will be output from the analysis model procedure for each of the 50 imputed datasets. The analysis results are combined to produce an aggregate p-value using Rubin's rules (Rubin, 1976). The SAS procedure PROC MIANALYZE will be used for this process, example code is provided in Appendix F.

For Week 24 endpoints, imputation for missing values for all secondary efficacy endpoints will follow the same procedure as for the primary efficacy endpoint as described above. For Week 48

endpoints, imputation will follow the same principles as for the primary efficacy endpoint but will include data up to Week 48.

9.4 PENALTY ASSIGNMENT FOR TIPPING POINT ANALYSES

Per Section 10.2.1.1, a tipping point analysis will be conducted to assess the effect of missing data on the reliability of the efficacy results by determining the extent the missing data have to change for the results to tip from statistically significant to not. The workflow for this imputation will be carried out as described for the control-based imputation. The assigned penalty for the primary endpoint, defined as the arithmetic reduction in % predicted $DLCO_{adj}$ change from baseline compared to the observed data in subjects randomized to placebo, can be implemented in Stage 2 of the imputation approach for the primary endpoint, facilitated through the SHIFT option in the MNAR statement. Example code is provided in Appendix F.

Data will be imputed to produce a $p^2 \times 50$ complete datasets, where p is the number of unique penalty increments explored for a single treatment group.

Similar tipping point analyses will be conducted for the secondary efficacy endpoints.

9.5 DERIVED AND TRANSFORMED DATA

9.5.1 TRIAL POPULATION

9.5.1.1 DEMOGRAPHICS AND BASELINE CHARACTERISTICS

Age as of the date of informed consent is auto-calculated on the eCRF based on the subject's year of birth, imputing the month and day of birth as January 01. Besides a statistical summary of age as a continuous variable, frequency of age will be presented in the following categories: ≥18<40, 40 to <65 years and ≥65 years. BMI will be calculated from the last recorded weight prior to first dose of IMP (i.e., baseline weight) and height recorded at screening. BMI is calculated as follows: weight(kg)/[height (m)]^2. In the summaries, BMI will be summarized as a continuous variable and as a frequency of subjects grouped into the following categories: [<18.5 kg/m^2, ≥18.5 kg/m^2], [<25.0 kg/m^2, ≥25.0 kg/m^2] and [<30.0 kg/m^2, ≥30.0 kg/m^2].

9.5.1.2 DISEASE HISTORY

Disease history for aPAP is collected on the eCRF at Visit 1/Screening 1, which includes the date of aPAP diagnosis. The time (months) since aPAP diagnosis to the screening visit (Visit 1) will be calculated as the number of days (inclusive) between the date of aPAP diagnosis and the date of screening Visit 1 divided by 30.4375 to approximate in months. In the event of partial date of aPAP diagnosis, the earliest of the month/year will be used to facilitate the duration calculation. If the date of diagnosis is completely missing, a duration will not be calculated. For subjects who had anti-GM-CSF antibody test performed prior to Visit 1/Screening 1, the date of the first positive test is collected, and the time (months) since the date of the first positive test to the screening visit (Visit 1) will be calculated in a similar manner.

9.5.1.3 PRIOR AND CONCOMITANT MEDICATIONS

Medications received prior to, concomitantly, or post-treatment will be coded using World Health Organization (WHO) Drug Dictionary Anatomical Therapeutic Chemical (ATC) level 3 classification codes and preferred terms from the March 2024, B3 version.

Prior, concomitant, and post-treatment medications are defined based on start and stop dates as follows:

- Prior medications are those taken prior to IMP with a stop date prior to the first dose of IMP.
- Concomitant medications are those with a stop date on or after the first dose date of IMP (and could have started prior to or during treatment) or Ongoing.
- Post-IMP medications are those with a start date after the last dose date of IMP.

For partially or completely missing medication start dates, the following imputation rules will be applied:

a. Missing day - Impute the 1st of the month unless month is the same as month of the first dose of IMP, then impute first dose date.

b. Missing day and month - Impute 1st January unless year is the same as first dose date of IMP, then impute first dose date.

c. Completely missing - Impute first dose date unless the medication stop date suggests it could have started prior to this in which case impute 1st January of the same year as the medication stop date.

For partially missing medication stop dates for medications that are not ongoing, the following imputation rules will be applied:

a. Missing day - Impute the last day of the month unless month is same as month of last dose of IMP, then impute last dose date.

b. Missing day and month - Impute 31st December unless year is the same as last dose date of IMP, then impute last dose date.

For completely missing medication stop dates, the following imputation rules will be applied:

c. Check whether the medication is still ongoing and when it started in relation to study drug. If the ongoing flag is present, then assume that the medication is still being taken (i.e., do not impute the date). If the medication has stopped and its start date is prior to first dose date of IMP, then impute the first dose date; if it started on or after first dose date of IMP, then impute to the last date of study participation for the subject.

Flags will be retained in the database indicating where any programmatic imputation has been applied, and in such cases, any durations would not be calculated.

9.5.1.4 TREATMENT COMPLIANCE AND EXPOSURE

Subjects' self-administration of IMP at home and treatment compliance is assessed at each visit, including timing of dose (e.g., morning or evening). Compliance is assessed by checking unused and used vials during the site visits, and data are entered in the source documents and eCRF. Deviation(s) from the prescribed dosage regimen should be recorded in the eCRF. Compliance percentage is calculated at each visit and recorded in the eCRF.

Treatment Compliance

For the analysis, treatment compliance for a treatment period will be calculated as 100% times the total number of empty vials returned plus the number of in-clinic doses received during the treatment period, divided by the actual exposure duration (in days) in the treatment period for each subject. Treatment compliance will be calculated for the first 24 weeks of double-blind treatment, the full 48-week double-blind treatment period, and the 96-week open-label period. In addition, treatment compliance will be calculated over the two treatment periods combined for subjects randomized to molgramostim.

The eCRF completion instructions are for the sites to calculate treatment compliance at a visit as days medication used since last visit divided by the total number of days since the last visit. Compliance will **NOT** use the eCRF compliance calculation in any summary or analysis.

Treatment Exposure

Treatment exposure will be calculated for the first 24 weeks of the double-blind treatment period, the full 48-week double-blind treatment period, the 96-week open-label treatment period, and over the two treatment periods combined for subjects randomized to molgramostim, as shown below.

<u>Exposure (24 weeks, double-blind):</u>

- Date of dose at the Week 24 visit – date of first dose + 1.

OR

- Date of discontinuation of IMP in the double-blind treatment period (as entered in the End of Double-blind Treatment eCRF) – date of first dose + 1 if subject withdraws prior to Week 24.

<u>Exposure (48 weeks, double-blind):</u>

- Date of completion/discontinuation of IMP in the double-blind treatment period (as entered in the End of Double-blind Treatment eCRF) – date of first dose + 1.

<u>Exposure (96 weeks, open-label):</u>

- Date of completion/discontinuation of IMP in the open-label treatment period (as entered in the End of Open-Label Treatment eCRF) – date of dose at the Week 48 visit + 1 if subject participated in the open-label treatment period.

OR

- Zero if subject did not participate in the open-label treatment period.

<u>Exposure (144 weeks total on molgramostim; only for subjects randomized to molgramostim):</u>

- Date of completion/discontinuation of IMP in the open-label treatment period (as entered in the End of Open-Label Treatment eCRF) – date of first dose + 1 if subject was randomized to molgramostim and participated in the open-label treatment period.

OR

- Date of completion/discontinuation of IMP in the double-blind treatment period (as entered in the End of Double-blind Treatment eCRF) – date of first dose + 1 if subject was randomized to molgramostim and did not participate in the open-label treatment period.

9.5.2 EFFICACY DERIVATIONS

9.5.2.1 PRIMARY ENDPOINT

The predicted DLCO value is calculated by the centrally provided equipment according to the Global Lung Function Initiative prediction equation (Stanojevic, et al. 2017). The calculation requires acceptable and repeatable maneuvers at baseline and Week 24, including the required discontinuation of supplemental oxygen for 15 minutes prior to the assessments.

The measured DLCO value is adjusted for the Hgb value obtained from the central laboratory, expressed in g/dL, using the following formula:

- Males: Predicted DLCO adjusted for Hgb = Predicted DLCO / (1.7Hgb/(10.22+Hgb))
- Females: Predicted DLCO adjusted for Hgb = Predicted DLCO / (1.7Hgb/(9.38+Hgb))

For each subject and visit, an adjusted value of predicted DLCO is derived based on the adjusted absolute value.

The Hgb value from the same day as the DLCO test is entered by the site into the DLCO device to be used for the adjustment. At Visit 3/Baseline, for decision on the inclusion/exclusion criteria and determination of the DLCO randomization stratum, the Hgb value obtained at Screening visit 2 is entered by the site and used for real time adjustment of % predicted DCLO results, because the Visit 3/Baseline Hgb is not available. After the Visit 3/Baseline Hgb value becomes available, the Screening visit 2 Hgb value that had been previously entered into the DLCO device will be replaced by the site with the Visit 3/Baseline Hgb, and the % predicted DLCO results at Visit 3/Baseline will therefore be adjusted for Hgb based on the Hgb value obtained at Visit 3/Baseline.

In instances where the Hgb value from the same day as the DLCO test at a visit is missing (due to the blood sample not being taken on the same day or the blood sample being taken on the same day but not deemed analyzable by the central laboratory), the non-missing Hgb value from the closest previous visit (which may be a scheduled visit or an unscheduled visit) is entered into the DLCO device for that visit. One exception to this rule is if the Hgb is measured at a later unscheduled visit that is closer in time to the date of the DLCO test than the closest previous visit, in which case the Hgb from this later unscheduled visit is used.

9.5.2.2 SECONDARY ENDPOINTS

SGRQ

The SGRQ Total score has 3 components: Activity, Impact, and Symptoms. The SGRQ is described in Section 2.2.1. Derivations of the SGRQ Total, Activity, Impact, and Symptoms Scores are described in Appendix D. Missing SGRQ Total scores and SGRQ Activity scores will be imputed using multiple imputation approaches as described for the primary endpoint in Section 9.3.

A range of endpoints based on the SGRQ Total score will be explored using the following categorizations:

- Subjects achieving a ≥ 4-point change in SGRQ Total score
  - Responder: ≥ 4-point change in SGRQ Total score
  - Non-responder: < 4-point change in SGRQ Total score
- Subjects achieving a ≥ 8-point change in SGRQ Total score
  - Responder: ≥ 8-point change in SGRQ Total score
  - Non-responder: < 8-point change in SGRQ Total score
- Subjects achieving a ≥ 12-point change in SGRQ Total score
  - Responder: ≥ 12-point change in SGRQ Total score
  - Non-responder: < 12-point change in SGRQ Total score

The responder status will be calculated based on change from baseline on all SGRQ Total scores, both from the observed data and for the imputed values following multiple imputation.

EC

Peak METs is the metric used to measure Exercise Capacity (using an established equation based on the speed and grade of the last stage the subject was able to complete for at least 15 seconds). The following validated equation to calculate peak METs will be used:

Peak METs = (speed X (0.17 + fractional grade X 0.79) + 3.5)/3.5
Note: Speed in meters/minute (Kokkinos et al. 2017)

Peak METs is derived and entered in the eCRF Exercise Overread form.

Peak METs will be explored using the following categorization:

- Responder: Change from baseline in peak METs ≥ 1
- Non-responder: Change from baseline in peak METs < 1

Responder status will be calculated at each scheduled visit where peak MET is collected.

A-aDO_2

For the calculation of A-aDO_2, the following variables will be assessed from an arterial blood gas sample collected on room air at the timepoints shown in the SoAs (Appendix B).

- PaO_2 (mmHg or kPa unit) – arterial partial pressure of oxygen
- $PaCO_2$ (mmHg or kPa unit) – arterial partial pressure of carbon dioxide.

As a measure of gas exchange, the A-a gradient (i.e., the A-aDO_2) will be calculated centrally using the following formula:

$$\text{Aa Gradient} = F_iO_2(P_{atm} - PH_2O) - (PaCO_2/0.8) - PaO_2$$

where P_{atm} (ambient atmospheric pressure) is measured at each visit in hPa unit as part of the DLCO assessment. Note that 1 hpa = 0.1 kPA = 0.75006 mmHg. The PH_2O (saturated vapor pressure of

water at body temperature) will be set to 47 mmHg or 6.266 kPa. The F_iO_2 (fraction of inspired oxygen) will be set to 0.21. For the analysis, the unit to use for A-aDO_2 is mmHg.

9.5.3 EXPLORATORY ENDPOINT DERIVATIONS

9.5.3.1 LUNG LAVAGE

Lung lavage is a rescue therapy reflecting the underlying pathophysiology of aPAP. The procedure required a hospitalization to perform because of the requirement for intubation of a single lung and mechanical respiration while the lavage is being performed on the opposite lung. Lung lavage can be performed during the clinical trial when the Investigator deems it necessary to relieve dyspnea or hypoxia as part of clinical care for aPAP subjects. Because LL is a consequence of the underlying disease, hospitalizations for LLs will be counted separately from other SAEs.

The number of all reported post-baseline LLs for a subject, regardless of whether some of the LLs were performed during the same hospitalization visit, will be determined for each subject. In addition, because some subjects may undergo multiple LL procedures during the same hospitalization for a clinical deterioration, the number of hospitalizations for LL will also be evaluated. In this alternative approach, a blinded review of the LL data for a subject will be performed to determine if multiple reports of LLs for a subject were performed during the same hospitalization visit, and if so, the multiple LLs for a subject occurring within the same hospitalization visit will be counted as a single unique event only.

The number of LLs, including date, start time, end time, primary reason for performing LL, and the lung(s) the procedure is performed on (both, right only, left only, segmental/lobar lavage), are also captured in the eCRF. The cumulative number of all reported post-baseline LLs and number of hospitalizations for post-baseline LLs will be calculated for each subject from baseline up to the following post-baseline timepoints:

- the Week 24 scheduled visit
- the Week 48 scheduled visit
- the Week 144 scheduled visit, for subjects randomized to molgramostim

The number of post-baseline LLs and number of hospitalizations for post-baseline LLs will also be calculated for each subject during the 96-week open-label treatment period.

To explore the type of LL over 24, 48, and 144 weeks, subjects will also have cumulative totals of all reported post-baseline LLs performed on:

- Both lungs
- The left lung
- The right lung
- Segmental/lobar lavage

9.5.3.2 DISTANCE WALKED AND DURATION OF EXERCISE DURING TREADMILL TEST

Duration of exercise is calculated as follows:

> Duration = ((number of stages fully completed x 30 seconds) + seconds completed at last stage)/60 secs/min. Results in minutes.

Distance walked is calculated as follows:

> Distance walked = the sum at each test stage of (speed at test stage x seconds completed at that stage). Results in meters.

The change from baseline in distance walked will be categorized into: < 50 m vs. ≥ 50 m.

9.5.3.3 DYSPNEA

The dyspnea score used in the exploratory efficacy analysis comes from the treadmill test. The difference in the Borg CR Scale® (CR10) taken (post-test result – pre-test result) is the dyspnea score used for endpoint analysis.

9.5.3.4 GROUND GLASS OPACITY

As described in Section 2.4.2, GGO scores range from 0-15.

The total GGO score will be calculated by summing up zonal GGO scores (i.e., total GGO score ranges from 0-15). If a zonal GGO score is missing, then the GGO total score will be missing. The average total GGO score of the two readers will be used in the statistical analysis. If there is only one non-missing GGO total score, then the non-missing will be used in the statistical analysis.

9.5.3.5 SUPPLEMENTAL OXYGEN USE

As described in Section 2.4.5, subjects record use of supplemental oxygen use daily over the 14 days leading up to a scheduled visit via the eDiary.

For a given day of oxygen use recorded by the subject, the total daily use will be derived as follows:

$$Daily\ supplemental\ oxygen\ use\ \left(\frac{L}{min}\right) = (\ [(960 - \{60\ \times q_8\}) \times q_3] + [480 \times q_5] + [60 \times q_8 \times q_7])/1440$$

Where q_3, q_5, q_7 are the answers to the supplemental oxygen eDiary questions on supplemental oxygen use (L/min) at rest, sleep and exertion activities respectively. Variable q_8 is the amount of time (hours) spent carrying out exerting activities for the given day of recording. If a subject records no oxygen use for a given activity, the oxygen use will be derived as 0 L/min for that given activity. Daily supplemental oxygen use will be missing if either q_i is missing for $i = 3,5,7,8$. Refer to Appendix E – Sample Oxygen eDiary for further detail on the questions implemented.

To analyze the data, a mean daily supplemental oxygen use (L/min) will be derived by calculating the average over the 14-day period preceding a clinic visit. Only the non-missing daily supplemental oxygen use will be used in the average calculation and the denominator will be adjusted for non-missing days.

Baseline mean daily supplemental oxygen use (L/min) will be defined as the average of the 14 non-missing use as recorded by the subject in the eDiary preceding the baseline visit.

9.5.4 SAFETY DERIVATIONS

Missing safety data will generally not be imputed. However, safety assessment values of the form of "< x" (i.e., below the lower limit of quantification) or > x (i.e., above the upper limit of quantification) will be imputed as "x" in the calculation of summary statistics but displayed as "< x" or "> x" in the listings. Note that 0 should not be used as an imputed value.

For missing diagnostic dates, if the day and/or month are missing, use 01 and/or Jan. If year is missing, put the complete date to missing.

Trial day will be calculated as the number of days from the date of first dose of IMP as follows:

- For any event on or after dosing = date of event – date of first dose + 1
- For events prior to dosing = date of event – date of first dose

9.5.4.1 ADVERSE EVENTS

TEAEs will be presented in the summary tables. TEAEs will be defined as any AEs observed from first dose of IMP through up to 30 days after the last dose of IMP.

The time to first onset of each AE from the first dose date and time will be calculated for presentation in listings as:

- AE onset date/time – first dose date/time, if the AE onset time is reported
- AE onset date – first dose date + 1 day, if the AE onset time is not reported

The duration of each AE will be calculated for presentation in listings as:

- AE end date/time – AE onset date/time, if both onset and end times are reported
- AE end date – AE onset date + 1 day, if one or both of onset and end times are not reported

For partially or completely missing AE start dates, the following imputation rules will be applied:

- Missing day - Impute the 1st of the month unless month is the same as month of the first dose of IMP, then impute first dose date.
- Missing day and month - Impute 1st January unless year is the same as first dose date of IMP, then impute first dose date.
- Completely missing - Impute first dose date unless the AE end date suggests it could have started prior to this in which case impute the first of January of the same year as the end date

For partially missing AE end dates, the following imputation rules will be applied:

- Missing day - Impute the last day of the month unless month is same as month of last dose of IMP, then impute last dose date.
- Missing day and month - Impute 31st December unless year is the same as last dose date of IMP, then impute last dose date.
- Completely missing (and the AE outcome is recorded as either resolved or resolved with sequelae) – Impute the last date of study participation by the subject, except if start date was after this date (e.g., an SAE that was reported within 30 days after this last date), in which case, impute the 31st of December of the same year as the AE end date.

Adverse events with missing severity data will be considered severe. AEs with missing relationship data will be considered related to the IMP.

If a subject is known to have died where only a partial death date is available, then the date of death will be imputed as the latest of the last date known to be alive + 1 from the database and the death date using the available information provided:

- For missing day only, use the 1st of the month.
- For missing day and month, use the 1st of January.

The following AEs have been identified as AEs of Special Interest (AESIs) with need for additional data collection which may include additional investigation when required to further characterize and understand them.

- Hypersensitivity reaction
- Chest pain

The events may be serious or non-serious and must follow the standards for AE/SAE reporting.

In case the Sponsor identifies potentially missed AESIs through predefined review of available data, the Investigator will be asked to reconsider if this is an AESI.

9.5.4.2 SPIROMETRY

At least three acceptable and repeatable maneuvers according to ATS/ERS criteria are required. Up to eight maneuvers may be conducted, if needed, during a session. The largest FEV_1 and FVC values meeting acceptable quality per ATS/ERS criteria, as identified by the overreader, will be used in the analysis.

9.5.4.3 ELECTROCARDIOGRAMS

Triplicate ECGs are obtained at each visit. For each ECG measurement or interval, the average of the triplicate values will be used in the analysis. For the overall cardiologist interpretation

(normal/abnormal) at each of Visit 3/Baseline, Visit 4/Week 4, and Visit 9/Week 24, the worst interpretation will be used in the analysis.

9.6 ANALYSIS VISIT WINDOWS

Analysis visit windows will be used for the by-visit analysis of efficacy and safety endpoints. The analysis visit windows will be constructed in such a way that the upper limit of an interval falls halfway between the two scheduled visits with the exception that the lower limit of the first post-baseline visit is Trial Day 2. If an even number of days exists between two consecutive visits, then the upper limit will be the higher number. Each scheduled trial day is 1 for Day 1 and *a* x 7 days + 1 for Week *a*. The Week 52 telephone visit will not be included in the visit windows since assessments are not performed at this visit.

In general, the value chosen for analysis at a visit will be from the assessment done, whether scheduled or unscheduled, closest to the target day according to the schedule of assessments provided in Appendix B. If there are two values that are equally close to the target day, the average value will be selected. For qualitative outcome measures, the earliest collected result will be selected.

If a subject discontinues treatment/withdraws from the trial, the last assessments recorded will be assigned to the visit window corresponding to the relative trial day on which the assessments are done. These assessments will be considered for analysis based upon other assessments within that visit window.

See Appendix C for all safety and efficacy data visit windows.

9.7 HANDLING OUTLIERS

A blinded review of the data related to the key parameters such as % predicted DLCOadj, A-aDO2, and exercise capacity will occur prior to the database lock after the completion of the double-blind treatment period. Endpoint data at each visit will be reviewed to identify outliers, including improbable values. Endpoint data collected during treatment discontinuation, treatment interruption(s), and use of rescue therapies (LL or other) will also be identified. The decision of what data will be excluded in any analysis and the reason identified during the blinded review of data will be documented in a "Pre-analysis-review" document before the database lock. In general, all data will be included in the main analyses of endpoints for FAS. Values that are considered extreme outliers can be considered for exclusion in the analyses of endpoints for PPS and for sensitivity and other supportive analyses.

10. STATISTICAL ANALYSES AND METHODOLOGY

10.1 TRIAL POPULATION

10.1.1 DISPOSITION OF SUBJECTS

Subject counts by region, country, site, and study status within each period will be presented in a table for all screened subjects. Subject disposition will be summarized using frequencies and percentages for all screened subjects and include the following parameters:

- Subjects consented

- Subjects randomized
- Subjects not randomized with a break-out of whether eligibility criteria not fulfilled or other reason

Subject disposition will be summarized by treatment group using frequency and percentages for both the FAS and the PPS and include the following parameters:

- Subjects in each of the analysis sets (FAS, SAFS, PPS, 24-week completer analysis set)
- Subjects who completed the double-blind treatment period and those who terminated prior to completion and the reason for early termination
- Subjects who completed the open-label treatment period and those who terminated prior to completion and the reason for early termination

All subjects who discontinued IMP and/or withdrew from the trial will be included in a listing.

10.1.2 PROTOCOL DEVIATIONS

Major protocol deviations will be summarized for the double-blind treatment period and the open-label treatment period separately and for the two periods combined, by categories defined in the *Protocol Deviation Specification* and will be provided to PHASTAR via SDTM.DV. The protocol deviation verbatim will be contained in DVTERM, standardized protocol deviation terms/categories collected in DVDECOD, and final classification (Major/Minor) will be collected in DVCAT.

All protocol deviations will be listed.

10.1.3 DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Demographics and subject characteristics will be summarized by treatment group using frequency and percentages (for categorical variables) and descriptive statistics of mean, standard deviation, minimum, median, and maximum for continuous variables.

The following characteristics will be summarized for both the FAS and PPS:

- Demographics including age, age group (≥18 to <40, 40 to <65, and ≥65 years), sex, race, and region (stratification used for randomization)
- % Predicted $DLCO_{adj}$ stratification used in the randomization (≤50%, >50%)
- Subject characteristics at baseline (height (cm), weight (kg), BMI (kg/m^2)), and BMI group (<18.5, ≥18.5 and <25.0, ≥25.0 and <30.0, ≥30.0 kg/m^2)
- aPAP Medical History including time (months) from aPAP diagnosis to the screening visit (Visit 1), mode of diagnosis (HRCT chest, lung biopsy, BAL cytology, or other), anti-GM-CSF antibody test performed prior to the screening visit (Visit 1) (yes, no) and if yes, time (months) from first positive test to the screening visit (Visit 1), prior LL procedures (yes, no), prior plasmapheresis procedures (previous, current, never), GM-CSF treatment status (previous, current, never), Rituximab treatment status (previous, current, never), and use of supplemental oxygen for aPAP (previous, current, never)

The following characteristics will be summarized for the FAS:

- Smoking history (previous, current, or never) and occupational dust exposure (previous, current, or never)
- COVID-19 infection history as of Visit 1/Screening 1 (yes, no) and if no, hospitalization (yes, no), ICU admission (yes, no), and oxygen required due to their COVID-19 infection (yes, no)
- COVID-19 vaccination and/or booster as of Visit 1/Screening 1 (yes, no)

10.1.4 TREATMENT COMPLIANCE

Treatment compliance is derived as described in Section 9.5.1.4.

Compliance will be summarized by treatment group in the first 24 weeks of the double-blind treatment period, the full 48-week double-blind treatment period, and the 96-week open-label treatment period. In addition, compliance will be summarized over the two treatment periods combined for subjects randomized to molgramostim.

Compliance will be summarized by frequency and percentage of subject in the following intervals: ≥90%, ≥80% to <90%, ≥70% to <80%, ≥50% to <70% and <50%.

Compliance will be listed by site and subject, including information on early withdrawal from the trial or discontinuation of IMP where relevant.

The SAFS will be used for compliance summaries.

10.1.5 EXTENT OF EXPOSURE

Treatment exposure is calculated as described in Section 9.5.1.4.

Duration of exposure will be summarized by treatment group as a continuous variable. The summaries will be analogous to those described above for treatment compliance.

Duration of exposure will also be presented as frequency and percentage of subjects in each of the following categories: 1 day to <8 weeks, ≥8 weeks to <16 weeks, ≥16 weeks to <24 weeks, ≥24 weeks to <32 weeks, and continuing by 8-week intervals.

The SAFS will be used for the treatment exposure summaries.

10.1.6 PRIOR AND CONCOMITANT MEDICATIONS AND THERAPIES

Prior and concomitant medications will be summarized by treatment group separately. They will be summarized as frequency and percentage of subjects being treated with each type of medication/therapy classified according to Anatomical Therapeutic Chemical (ATC) level 3 and WHO Drug Global Dictionary preferred term. The SAFS will be used for these summaries.

Summaries of concomitant medications will be presented separately for each of the two treatment periods for all subjects, as well as for the two periods combined for subjects randomized to molgramostim.

History of treatments for aPAP will be summarized in a separate table as noted in Section 10.1.3 as part of baseline data.

All prior and concomitant medications will be listed.

10.1.7 MEDICAL HISTORY

Medical history terms will be summarized by treatment group, and preferred term within system organ class based on MedDRA.

10.2 EFFICACY ANALYSES

10.2.1 PRIMARY EFFICACY ANALYSIS

A general linear MMRM will be used to analyze the primary endpoint and will be fitted with treatment, a binary indicator for DLCO severity stratification at randomization, and a 3-level factor for region, and visit as categorical fixed effects, along with a treatment-by-visit interaction term, and baseline % predicted $DLCO_{adj}$, as a covariate. The estimated treatment effect will be the difference in LSMean change in % predicted $DLCO_{adj}$ from baseline to Week 24, taken from the treatment-by-visit interaction term at 24 weeks. Although data is also collected after the Week 24 visit, only data up to Week 24 will be used in the statistical model for the primary endpoint at Week 24 (i.e., up to Visit 9). The estimated treatment effect will be presented with a 95% CI and a p-value to test the null hypothesis that the effects of molgramostim and placebo at Week 24 are the same.

The analysis model is

$$Y_{ijk} = \beta_0 * y_{ij0} + T_i + R_l + S_m + V_k + TV_{ik} + s_{ij} + e_{ijk}$$

where

Y_{ijk} is the change from baseline in % predicted $DLCO_{adj}$ value for the j^{th} subject of treatment group i at visit k (where $k=4,\ldots,9$)

y_{ij0} is the baseline % predicted $DLCO_{adj}$ value for the j^{th} subject of treatment group i

β_0 is the unknown fixed slope for the baseline % predicted $DLCO_{adj}$

T_i is the unknown fixed effect of treatment i

R_l is the unknown fixed effect of regional stratification factors l (0 or 1 or 2)

S_m is the unknown fixed effect of DLCO severity stratification at randomization factor m (0 or 1)

V_k is the unknown fixed effect of visit k

TV_{ik} is the unknown fixed interaction effect between treatment i and visit k

s_{ij} is the subject effect associated with the j^{th} subject of treatment i

e_{ijk} is the error (residual) associated with the j^{th} subject of treatment i at visit k

s_{ij} and e_{ijk} are assumed to be independent from each other and follow a multivariate normal distribution. The covariance matrix for e will be the unstructured variance-covariance matrix, since it assumes pair-wise correlations are not constrained by the data. An unstructured covariance matrix will be applied to model within-subject errors. If this analysis

fails to converge, a compound symmetry matrix will be tested. Kenward-Roger approximation will be used to estimate denominator degrees of freedom.

The estimated treatment effect is taken from the TV_{ik} interaction term at Visit 9 (i.e., 24 weeks). Refer to Appendix G for example SAS code for fitting a repeated measures model to the primary endpoint.

Any subject with missing % predicted $DLCO_{adj}$ will have those missing values imputed using a multiple imputation method, using a conservative control-based rule as described in Section 9.3.

Under the primary estimand, the analysis will include all observed % predicted $DLCO_{adj}$ changes from baseline during the trial at the scheduled trial visits up to Week 24. Missing data will be handled in the primary analysis using multiple imputation methods as described in Section 9.3, where the results from the 50 complete datasets will be combined using Rubin's rules (Rubin, 1976).

The estimated LSMean changes from baseline in % predicted $DLCO_{adj}$ changes from baseline at each post-baseline visit up to Week 24 will be displayed in a separate output, along with the treatment group differences in LSMeans and 95% CIs.

All assessed % predicted $DLCO_{adj}$ measurements at each visit and the corresponding changes from baseline will be summarized descriptively by treatment group. Descriptive summaries of % predicted $DLCO_{adj}$ will be based on observed % predicted $DLCO_{adj}$ and missing data will not be imputed.

The primary analysis described will be run on the FAS. It will be the analysis used to assess the primary objective under the type I error-controlled testing strategy. As a supportive analysis to the primary, the analysis will be repeated on subjects in the PPS.

10.2.1.1 SENSITIVITY ANALYSES OF THE PRIMARY ENDPOINT

Sensitivity analyses for the primary efficacy endpoint are described within the estimand-to-analysis table in Section 6.1.

In addition to the primary analysis approach using general linear MMRM as described in Section 10.2.1, an ANCOVA will be used, as a sensitivity analysis, to analyze the change from baseline to Week 24 in % predicted $DLCO_{adj}$ and will be fitted with treatment, a binary indicator for DLCO severity stratification at randomization, a 3-level factor for region as categorical fixed effects, and baseline % predicted $DLCO_{adj}$ as a covariate.

If the data for change from baseline to Week 24 in % predicted $DLCO_{adj}$ appear to be non-normally distributed, van Elteren test will be used, as a sensitivity analysis, to analyze the change from baseline to Week 24 in % predicted $DLCO_{adj}$. The test will be stratified by a binary indicator for DLCO severity stratification at randomization, and a 3-level factor for region as categorical fixed effects. The Shapiro-Wilk test will be used to assess deviations from normality for change from baseline to Week 24 in % predicted $DLCO_{adj}$, where a resulting p-value ≤0.05 means that the van Elteren test should be conducted. The Rubin's rules for combining the results from the 50 imputed datasets cannot be applied to the van Elteren's test. The median for each descriptive statistic and the median p-value across the 50 imputed datasets will be reported.

In addition, responder and tipping point analyses will be performed, as described below. Sensitivity analyses will be performed on the FAS.

Responder Analyses

A responder analysis is an analysis or presentation of the proportion of participants who achieve a pre-defined level of improvement on one of the main outcome variables at a certain time point.

For the primary endpoint, there will be three responder analyses, with a responder defined as having ≥ 5 percentage-point improvement (i.e., increase), ≥ 7 percentage-point improvement, and a ≥ 10 percentage-point improvement in % predicted $DLCO_{adj}$.

For each responder threshold, a logistic regression model, fitted with the same covariates as the MMRM model from the primary analysis, will be used to analyze the responder endpoint. Odds ratios, 95% CLs, and 2-sided p-values will be presented. Refer to Appendix G for example SAS code to fit a logistic regression model to a binary response variable. As with the primary analysis, the responder analysis will use data from multiple imputation, and will fit 50 logistic regression models and their results will be combined using Rubin's rules (Rubin, 1976).

Tipping Point Analyses

Tipping point analysis will be performed as a sensitivity to the primary analysis to assess robustness of conclusions under varying, conservative assumptions on the missing primary endpoint data.

The tipping point analysis will be facilitated through multiple imputation, whereby a penalty will be assigned to the imputed values of change from baseline in % predicted $DLCO_{adj}$. The penalty can be interpreted as the arithmetic reduction in % predicted $DLCO_{adj}$ change from baseline compared to the observed data in subjects randomized to placebo. A penalty will be assigned to both treatment groups ranging from 0% (yielding results equal to the primary analysis) up to twice the observed LSMean treatment difference in % predicted $DLCO_{adj}$ change from baseline, obtained from the primary analysis repeated measures model. The range will be explored in increments of 1%. Please refer to Section 9.4 for details on implementing a penalty in the multiple imputation process.

The tipping point analysis will produce a 2-dimensional array of outcomes under the exhaustive combinations of penalty (at 1% increments) assigned to both treatment groups. For each penalty combination, results will be aggregated using Rubin's rules (Rubin, 1976). Outputs will present the LSmean % predicted $DLCO_{adj}$ changes from baseline, treatment difference in LSMeans and corresponding 95% CIs and aggregated 2-sided p-values.

10.2.1.2 SUPPLEMENTAL ANALYSIS OF THE PRIMARY ENDPOINT

As a supplemental analysis to the primary analysis, the analyses as described in Section 10.2.1 will be repeated for the % predicted $DLCO_{adj}$ changes from baseline under the COVID-19 estimand in the FAS population. Refer to Section 6.2.1 for details. The estimand will be clearly indicated in the output titles for these supplementary analyses.

10.2.2 SECONDARY EFFICACY ANALYSES

10.2.2.1 % PREDICTED $DLCO_{ADJ}$ AT WEEK 48 (OUTSIDE OF JAPAN AND SOUTH KOREA)

The main and sensitivity analyses for the change in % predicted $DLCO_{adj}$ from baseline to Week 48 are described within the estimand-to-analysis table in Section 6.2.1. For the main analysis, this secondary efficacy endpoint will be analyzed using a similar MMRM as described in Section 10.2.1 for change in % predicted $DLCO_{adj}$ from baseline to Week 24 but with the Week 48 timepoint used for inference and using all visits up through Week 48 in the model.

10.2.2.2 SAINT GEORGE'S RESPIRATORY QUESTIONNAIRE TOTAL SCORE AND ACTIVITY SCORE AT WEEKS 24 AND 48

The main and sensitivity analyses for the changes in SGRQ Total score from baseline to Week 24 and Week 48 are described within the estimand-to-analysis table in Section 6.2.2. For the main analyses, these secondary efficacy endpoints will be analyzed using a similar MMRM as described in Section 10.2.1 and Section 10.2.2.1 for change in % predicted $DLCO_{adj}$ from baseline to Week 24 and Week 48, respectively, except using baseline SGRQ Total score as the covariate (y_{ij0}) instead of baseline % predicted $DLCO_{adj}$. For the sensitivity analyses, these secondary efficacy endpoints will be analyzed using similar ANCOVA and van Elteren test as described in Section 10.2.1.1 for the primary efficacy endpoint, except using baseline SGRQ Total score as the covariate.

SGRQ Total scores at each visit and the corresponding changes from baseline will be summarized descriptively by treatment group. Descriptive summaries of SGRQ Total score will be based on observed data.

Responder analyses

For the SGRQ Total score change from baseline to Week 24 secondary endpoint, there will be three responder analyses with a responder defined as having ≥ 4-point improvement (i.e., decrease) in score, ≥ 8-point improvement in score, and a ≥ 12-point improvement in score for the SGRQ Total.

For each definition of responder (detailed in Section 9.5.2.2), a logistic regression model will be used to analyze the responder endpoint, fitted with the same covariates as the MMRM model from the secondary endpoint analysis. Odds ratios, 95% CIs, and 2-sided p-values will be presented. Refer to Appendix G for example SAS code in fitting a logistic regression model to a binary response variable.

Tipping point analyses

To assess the impact of missing SGRQ Total score changes from baseline, tipping point analyses will be implemented using a multiple imputation approach as described for the primary endpoint in Section 10.2.1.1. The multiple imputation process will impute the missing changes from baseline in SGRQ score (continuous variable) to produce 50 completed datasets. A penalty to the change from baseline in SGRQ Total score will be applied to both treatment groups, ranging from 0 to twice the difference between treatment groups in arithmetic mean change from baseline SGRQ Total score. The range will be explored in increments of 1 point score and will be independently applied to each treatment group. Should any imputed values correspond to change from baseline which is

impossible to obtain given the subject's observed baseline score and SQRQ absolute value range, the imputed value will be reset to the limit in which the imputed value PROC MI exceeded.

The tipping point analysis will produce a 2-dimensional array of outcomes under the exhaustive combinations of penalty assigned to both treatment groups. For each penalty combination, the changes from baseline in SGRQ Total score will be analyzed via MMRM fitted with the same covariates as in the secondary endpoint analysis. The estimates of the treatment difference, 95% CIs and 2-sided p-values from the tipping point analysis will be combined using Rubin's rules (Rubin, 1976).

Main and sensitivity analyses described above and in the estimand-to-analysis table in Section 6.2.2 will be repeated for the SGRQ Activity score.

10.2.2.3 EXERCISE CAPACITY AT WEEKS 24 AND 48

The main and sensitivity analyses for the changes in EC (expressed in peak METs) from baseline to Week 24 and Week 48 are described within the estimand-to-analysis table in Section 6.2.3. For the main analyses, these secondary efficacy endpoints will be analyzed using a similar MMRM as described in Section 10.2.1 and Section 10.2.2.1 for change in % predicted $DLCO_{adj}$ from baseline to Week 24 and Week 48, respectively, except using baseline peak METs as the covariate (y_{ij0}) instead of baseline % predicted $DLCO_{adj}$.

Peak METs at each visit and the corresponding changes from baseline will be summarized descriptively by treatment group. Any abnormalities and whether the abnormalities are clinically significant or not clinically significant will be summarized. Descriptive summaries of peak METs will be based on observed data.

Sensitivity analyses, consisting of tipping point (at increments of 0.2 METs) and responder analyses (using the categories for change in peak METs described in Section 9.5.2.2), will be handled using the same methods as described in Section 10.2.2.2**Error! Reference source not found.**.

10.2.2.4 A-aDO_2 AT WEEK 24 (SECONDARY ENDPOINT FOR JAPAN AND SOUTH KOREA ONLY)

The main and sensitivity analyses for the change in A-aDO_2 from baseline to Week 24 are described within the estimand-to-analysis table in Section 6.2.4. For the main analysis, this secondary efficacy endpoint will be analyzed using a similar MMRM as described in Section 10.2.1 for change in % predicted $DLCO_{adj}$ from baseline to Week 24, except using baseline A-aDO_2 as the covariate (y_{ij0}) instead of baseline % predicted $DLCO_{adj}$.

A-aDO_2 at each visit and the corresponding changes from baseline will be summarized descriptively by treatment group. Descriptive summaries of A-aDO_2 will be based on observed data.

Tipping point analysis (at increments of 1 mmHg) will be handled using the same method as described in Section 10.2.2.2.

In the event of change in A-aDO_2 from baseline to Week 24 appearing to be non-normally distributed, a van Elteren test will be used, as a sensitivity analysis, to analyze the change from baseline to Week 24 in A-aDO_2. Assessing deviations from normality and applying the van Elteren test will be conducted using the same methods as described in Section 10.2.1.1.

10.2.3 EXPLORATORY EFFICACY ANALYSES DURING THE DOUBLE-BLIND TREATMENT PERIOD

Exploratory endpoints during the double-blind treatment period will be summarized through descriptive statistics and/or frequency distributions. Statistical testing of exploratory endpoints will not be adjusted for multiplicity, as done for primary and secondary analyses as described in Section 5.3. Inferences will therefore be considered supportive. No imputation on missing data will be done. Analyses on these endpoints will be performed on the FAS.

10.2.3.1 LUNG LAVAGE

The cumulative numbers of all reported post-baseline LLs and hospitalizations for post-baseline LLs, as defined in Section 9.5.3.1, from baseline to Weeks 24 and 48 will be presented by treatment group.

Additionally, a breakdown of the number (%) of subjects requiring at least one LL procedure in the first 24 and 48 weeks of treatment will be presented.

To explore effects of randomized treatment on the overall incidence of LL, a generalized linear model, based on the negative binomial distribution will be used to analyze the endpoints of frequency of LL events in the period between baseline and 48 weeks. The cumulative number of LLs in the period (baseline to 48 weeks) will be the dependent variable. Model covariates will include treatment group, a binary indicator for DLCO stratification, a 3-level factor for region, and an indicator as to whether LL was performed at any time before randomization. The subject time at risk (weeks) will be included as an offset variable in the model. Time at risk (weeks) is defined as the number of days the subject remains in the trial, from baseline up to hospitalization for LL, divided by 7 days. LSmean estimates of LL event rates over 48 weeks and rate ratio of the treatment effect, along with 95% CIs will be estimated from the negative binomial model. Refer to Appendix G for example SAS code to fit a negative binomial regression model to recurrent event data.

Kaplan-Meier plots will be used to assess time to first LL (time from the date of first dose of double-blind IMP to the date of the first hospitalization for LL during the double-blind treatment period). Treatment comparison will be performed using the logrank test, adjusting for randomization stratifications based on DLCO and region. Subjects withdrawing from the trial during the double-blind treatment period will be censored at their times of discontinuation. Subjects who complete the double-blind treatment period without LL will be censored at the relevant analysis time point (either Week 24 or Week 48). Survival estimates (mean, median) for time to first LL will be tabulated for each treatment group with 95% CIs for the median time.

10.2.3.2 SAINT GEORGE'S RESPIRATORY QUESTIONNAIRE IMPACT AND SYMPTOMS

Similar to the SGRQ Total and Activity scores, the SGRQ Impact and Symptoms scores at each visit and the corresponding changes from baseline will be summarized by treatment group.

Changes from baseline to Week 24 and Week 48 in SGRQ Impact score and SGRQ Symptoms score will be analyzed using a similar model as for SGRQ Total score, except using baseline SGRQ Impact or SGRQ Activity as the covariate (y_{ij0}) instead of baseline SGRQ Total score. All available data will be used, but no imputation of missing data will be performed, as these are exploratory endpoints.

10.2.3.3 DISTANCE WALKED AND DURATION OF EXERCISE DURING TREADMILL TEST

The distance walked and duration of exercise during treadmill test (derived as described in Section 9.5.3.2), and other assessments made during the treadmill test will be summarized by treatment group at each visit the test was performed. These assessments include:

Pre-test: Current use of betablocker (yes/no), rating of perceived exertion (Borg RPE Scale®), dyspnea (Borg CR Scale®), angina scale, and SpO_2

Post-test: Rating of perceived exertion (Borg RPE Scale®), dyspnea (Borg CR Scale®), angina scale, SpO_2, and reason for stopping the test

Others: Change in dyspnea (post-test – pre-test); change in SpO2 (post-test SpO_2 – pre-test SpO_2), immediate post-test symptoms observed including chest discomfort, lightheadedness, leg fatigue, dyspnea, other; any other abnormalities reported during or after the test (yes/no); assessment of clinical significance for all abnormalities and symptoms.

The analysis of dyspnea is also described in Section 10.2.3.13. The analysis of SpO_2 is also described in Section 10.2.3.14.

Change from baseline at each visit the test was performed will also be summarized for the numeric measures. Changes from baseline to Week 24 and Week 48 in distance walked and duration of exercise will be analyzed using a similar MMRM model as for % predicted $DLCO_{adj}$, except using baseline distance walked and duration of exercise as the covariates (y_{ij0}) instead of baseline % predicted $DLCO_{adj}$. Responder analysis of distance walked (using the categories described in Section 9.5.3.2) will be performed using a logistic regression model. All available data will be used in the analyses, but no imputation of missing data will be performed, as these are exploratory endpoints. Refer to Appendix G for example SAS code in fitting a logistic regression model to a binary response variable.

10.2.3.4 ALVEOLAR-ARTERIAL OXYGEN DIFFERENCE

As stated in Section 10.2.2.4, A-aDO2 at each visit arterial blood gas sampling was performed and the corresponding changes from baseline will be summarized by treatment group.

Changes from baseline to Week 48 in A-aDO_2 will be analyzed using a similar MMRM model as for the main analysis for changes from baseline to Week 24 in A-aDO2. Note that change in A-aDO2 from baseline to Week 24 is a secondary endpoint for Japan and South Korea and will be analyzed as described in Section 10.2.2.

10.2.3.5 ARTERIAL PARTIAL PRESSURE OF OXYGEN

PaO_2 at each visit and the corresponding changes from baseline will be summarized by treatment group.

10.2.3.6 DISEASE SEVERITY SCORE

DSS at each visit arterial blood gas sampling was performed and the corresponding shifts from baseline will be summarized by treatment group.

10.2.3.7 GROUND GLASS OPACITY

Total GGO score will be summarized by treatment group over time in terms of absolute values and change from baseline at each scheduled assessment.

Additionally, the overall GGO interpretation by the Investigator (normal, abnormal non-clinically significant, and abnormal clinically significant) will be summarized by treatment group.

10.2.3.8 CLINICIAN'S GLOBAL IMPRESSION OF SEVERITY AND CHANGE

CGIS and CGIC will be summarized separately for each post-baseline visit. Frequency distributions of the responses will be presented, and treatment comparisons will be performed at Baseline (for CGIS only), Week 24, and Week 48 using Fisher's Exact test.

10.2.3.9 PATIENT'S GLOBAL IMPRESSION OF SEVERITY AND CHANGE

Frequency distributions of the responses to the PGIS and PGIC questions at each visit will be presented by treatment group, and treatment comparisons will be performed at Baseline (for PGIS only), Week 24, and Week 48 using Fisher's Exact test.

The responses to the severity of breathing problems (PGIS) will also be assigned a numeric value as follows: none = 1, mild = 2, moderate = 3, severe = 4, very severe = 5. The numeric responses at each visit and the corresponding changes from baseline will be summarized.

10.2.3.10 SUPPLEMENTAL OXYGEN

The mean daily supplemental oxygen use (L/min), as described in Section 9.5.3.5, and the changes from baseline will be summarized descriptively by treatment group at each visit where such data are collected according to the schedule of activities. The summaries will be presented on: 1) all subjects at the scheduled visit and 2) the subjects administered with supplemental oxygen in the 14 days prior to the scheduled visit; the composition of the subgroup may vary from visit to visit. For the summaries on all subjects at a visit, the subjects with no supplemental oxygen use reported (i.e., either 0 or missing) in the diary over the 14 days prior to the scheduled visit will be assigned a value of 0 for their mean daily supplemental oxygen use at that visit.

The shifts from baseline in subjects reporting use of supplemental oxygen (Yes/No response) since the last visit will be tabulated at each post-baseline visit.

For the subset of subjects who were not reported as being on supplemental oxygen at any time during the screening period, the time from the date of the first dose of double-blind IMP to the date of the first use of supplemental oxygen during the double-blind treatment period will be assessed using Kaplan-Meier plots, and treatment comparison will be performed using logrank test, adjusting for randomization stratifications based on DLCO and region. Subjects withdrawing from the trial during the double-blind treatment period will be censored at their times of discontinuation. Subjects who complete the double-blind treatment period without using supplemental oxygen will be censored at the relevant analysis time point (either Week 24 or Week 48). Survival estimates (mean, median) for time to first use of supplemental oxygen will be tabulated for each treatment group with 95% CIs for the median time.

10.2.3.11 BIOMARKER LEVELS

Biomarker levels at each visit with data collection and corresponding changes from baseline will be summarized by treatment group.

10.2.3.12 EUROQOL 5 DIMENSIONS 5 LEVELS

Frequency distributions to the responses for each of the 5 EQ-5D-5L domains will be presented by treatment group and scheduled visit. In addition, shifts in responses from baseline to each post-baseline visit will be tabulated. The 5 EQ-5D-5L domains at each of Week 24 and Week 48 will be modelled via an ordinal logistic regression model, with treatment, baseline EQ-5D-5L domain score, a binary indicator for DLCO severity stratification at randomization, and a 3-level factor for region as categorical fixed effects. Treatment group comparisons using odds ratios, 95% CIs and associated 2-sided p-values will be estimated from the model. Refer to Appendix G for example code in fitting an ordinal logistic regression model in SAS.

The EQ-5D-5L VAS score and changes from baseline will be summarized descriptively by treatment group and scheduled visit.

10.2.3.13 DYSPNEA

Dyspnea (pre-test, post-test, and difference between the two) at each visit that the exercise treadmill test was performed and the corresponding changes from baseline will be summarized by treatment group. The calculation of the difference in dyspnea scores used in the analysis is described in Section 9.5.3.3.

10.2.3.14 OTHER EXPLORATORY ENDPOINTS

Absolute $DLCO_{adj}$ and SpO_2 at each visit and the corresponding changes from baseline will be summarized by treatment group. The SpO_2 values are those from before and after the exercise treadmill test, and the changes from baseline will be summarized for the pre-test values and for the differences between the post-test and pre-test values.

10.2.4 EXPLORATORY EFFICACY ANALYSES DURING THE OPEN-LABEL TREATMENT PERIOD

Summaries by visit during the open-label treatment period will be presented by treatment group and for both groups combined for the efficacy endpoints % predicted $DLCO_{adj}$, SGRQ Total, Activity, Impact, and Symptom scores, PaO_2, CGIS, PGIS, supplemental oxygen use, biomarker levels, EQ-5D-5L, CGIC, and PGIC. For continuous endpoints, change from baseline for each post-baseline visit and change from Week 48 for each post-Week 48 visit during the open-label treatment period will also be summarized. The cumulative number of LLs during the open-label treatment period and from baseline to Week 144 will be summarized in a manner similar to that described for these endpoints during the double-blind treatment period in Section 10.2.3.1 (excluding the statistical modelling and testing described for LLs)**Error! Reference source not found.**.

10.3 SAFETY ANALYSES

10.3.1 ADVERSE EVENTS

As stated in Section 2.3.1, MedDRA will be used to code AEs. All AEs will be listed, and TEAEs will be summarized descriptively by frequency and percentage, and as exposure-adjusted incidence rates

per 100 patient-years (see Section 8.1 for details on deriving exposure-adjusted incidence rates). Descriptive summaries will be presented for each treatment group and overall (across both treatment groups). Non-TEAEs will be included in the AE listings but will not be included in the summary tables (unless otherwise stated). Separate summaries of TEAEs will be presented for the first 24 weeks of the double-blind treatment period, the entire double-blind treatment period, and the open-label treatment period. In addition, the summary of TEAEs over the two treatment periods combined will also be presented for subjects randomized to molgramostim.

An overall TEAE summary table will be presented with the frequency and percentage of subjects with at least one AE, subjects with at least one SAE, subjects with an AE with an outcome of death, subjects with at least one AE leading to IMP discontinuation, subjects with at least one AE leading to trial discontinuation, subjects with at least one severe AE, subjects with at least one AE of special interest (AESI), subjects with at least one serious AESI, subjects with at least one treatment-related AE (as assessed by the investigator), and subjects with at least one serious treatment-related AE.

Subject incidence summary tables, tabulated by system organ class (SOC) and preferred term (PT), will be presented for the following:

- All TEAEs
- Most frequent TEAEs (≥2% total incidence)
- All treatment-emergent SAEs

Multiple events per subject will not be accounted for apart from any episode level summaries which may be produced.

Subject incidence summaries tabulated by severity (mild, moderate, and severe) and relationship to IMP (related, not related), as assessed by the investigator, will be presented. Related includes the categories of “related” and “possibly related”; not related includes the categories of “unlikely related” and “not related”.

In the overall summary table of TEAEs and the summary of most frequent TEAEs (≥2% total incidence) during the double-blind treatment period, risk differences between treatment groups will be presented along with associated 95% CIs, per TEAE. Agresti-Caffo CIs will be computed for risk differences, which can accommodate zero events in one of the treatment groups. The purpose of including risk differences and CIs is to provide descriptive summaries of treatment differences for safety, rather than to perform formal hypothesis testing.

All SAEs and AEs leading to withdrawal from the trial and/or permanent discontinuation from IMP will be fully described in individual subject narratives.

In AE listings, the relative day of the start of the AE, counted from the first day of IMP (Day 1), will be presented together with the actual date.

10.3.2 DEATHS AND SERIOUS ADVERSE EVENTS

There will be listings for all AEs with an outcome of death and for all SAEs. SAEs will be tabulated by SOC and PT by treatment group.

10.3.3 ADVERSE EVENTS LEADING TO DISCONTINUATION OF INVESTIGATIONAL PRODUCT AND/OR WITHDRAWAL FROM THE TRIAL

Adverse events leading to the discontinuation of IMP and/or withdrawal from the trial will be listed.

10.3.4 OTHER ADVERSE EVENTS OF SPECIAL INTEREST

AEs considered as AESIs are hypersensitivity and chest pain as described in Section 9.5.4.1. AESIs will be tabulated by SOC and PT for each treatment group. AESIs will also be listed.

10.3.5 CLINICAL LABORATORY EVALUATIONS

Clinical chemistry and hematology parameters and urinalysis parameters with continuous values will be summarized at each visit by treatment group. Change from baseline will be summarized for each post-baseline visit. Change from Week 48 will also be summarized for each post-Week 48 visit during the open-label treatment period.

For clinical chemistry and hematology parameters, shift tables will summarize change from baseline at each visit using normal ranges provided by the central laboratory. Shift tables will also be produced for change from Week 48 to each visit after Week 48. For the first 24 weeks and for the entire double-blind treatment period, shift tables will also be presented for change from baseline to the minimum and maximum post-baseline results during this period; for these shift tables, all post-baseline results, not only those closest to the target day for an analysis visit window, will be considered in identifying the minimum and maximum post-baseline results. For the open-label treatment period, shift tables will also be presented for change from Week 48 to the minimum and maximum post-baseline results during this period. In addition, shift tables will be presented for change from baseline to the minimum and maximum post-baseline results over the two treatment periods combined for subjects randomized to molgramostim.

Urinalysis parameters with ordinal or categorical values will be summarized by treatment group at each visit through frequency distributions. Pregnancy test results will be listed only.

Data summaries and listings will be presented in preferred units provided by the Sponsor.

All laboratory data will be listed. Flags will identify values that fall outside of reference ranges. A separate listing of abnormal laboratory results will be presented and will include the investigator assessment of clinical significance at the laboratory panel-level only.

Box plots of absolute values by visit and treatment group may be presented for certain parameters if warranted after data review by the Sponsor.

10.3.6 VITAL SIGNS AND BODY WEIGHT

Vital signs (systolic blood pressure, diastolic blood pressure, heart rate, respiration rate, body temperature, and weight) will be summarized by treatment group at each visit in terms of absolute values and change from baseline at each scheduled measurement. Change from Week 48 will also be summarized for each post-Week 48 visit during the open-label treatment period.

10.3.7 ELECTROCARDIOGRAMS

ECG parameters (mean heart rate, RR interval, PR interval, QT interval [uncorrected], QRS duration, QTcB [Bazett's correction], and QTcF [Fridericia's correction]) will be summarized by treatment

group over time in terms of absolute values and change from baseline at each scheduled measurement. Change from Week 48 will also be summarized for each post-Week 48 visit during the open-label treatment period.

Additionally, the overall ECG interpretation by the cardiologist at the central ECG laboratory (normal, abnormal; at Visit 3/Baseline, Visit 4/Week 4, and Visit 9/Week 24 only) will be summarized by visit by treatment group.

The Investigator's assessment of an ECG as normal, abnormal NCS, or abnormal CS will not be summarized but will be presented in a listing.

10.3.8 SPIROMETRY

FVC, FEV_1, and FEV_1/FVC, both absolute values and % predicted values, will be summarized by treatment group at each visit along with change from baseline at each post-baseline visit and change from Week 48 at each post-Week 48 visit.

10.3.9 DEVELOPMENT OF ON-TREATMENT ANTI-GM-CSF ANTIBODY TITERS

On-treatment anti-GM-CSF antibody titers (numeric results and frequency of positive results) will be summarized by treatment group at each visit along with change from baseline at each post-baseline visit and change from Week 48 at each post-Week 48 visit.

10.3.10 LONGER-TERM SAFETY ANALYSES

Longer-term safety analyses refer to the safety summaries during the open-label treatment period by treatment group for all treated subjects and during the double-blind and open-label treatment periods combined for subjects randomized to molgramostim. These safety summaries are already described in the earlier subsections under Section 10.3.

10.4 PHARMACOKINETICS

PK samples are collected at pre-dose and 2 hours post-dose at each of Baseline, Week 4, Week 24, and Week 48. At each visit the GM-CSF concentrations at each of the two timepoints and the difference in levels between the two timepoints will be summarized by treatment group. The changes from baseline will be summarized as well.

The GM-CSF concentrations will be presented in a listing of individual values and aggregated in a summary table using the following descriptive statistics: sample size (n), arithmetic mean, SD, coefficient of variation (CV%), minimum and maximum values, median, geometric mean and associated 95% CI, and geometric mean CV%. GM-CSF concentrations will be reported to 3 significant figures.

If there are fewer than three values available for calculation of basic summary statistics, only the frequency (n), minimum, and maximum values will be reported.

10.5 BIOMARKERS

Analysis of biomarkers is described in Section 10.2.3.11.

11. CHANGES FROM THE PROTOCOL-SPECIFIED ANALYSES

- Per protocol Section 8.2.1, for the final DLCO results at a visit, the average of the two best efforts, as determined by the overreader, is used. According to the overread guidelines, however, the average is taken over all single acceptable DLCOs at that visit, and therefore, this is what is stated in the SAP (Section 2.1).
- Updated hypothesis testing schematic – schema was originally taken from the protocol, but alpha_1 alpha_2 and alpha_3 do not fully match the language in the description (nor the Hochberg step-up) process.
- EQ-5D-5L endpoint will not be analyzed/presented as a 5-digit value representing a subject health state, nor will it be converted into an index value, per protocol Section 8.11.1. Instead, EQ-5D-5L will be summarized within the 5 domains using the ordinal response values; the VAS score will be summarized as a continuous measure separately.
- The safety objective of frequency of serious ADRs is replaced with frequency of serious treatment-related AEs. ADRs for molgramostim will be assessed in the integrated summary of safety instead.
- Changes in QTcB from baseline to Weeks 4 and 24 included as safety endpoints, in addition to changes in QTcF.
- There is a change in nomenclature from Whole Lung Lavage (WLL) to Lung Lavage (LL) when describing the derivation and analysis of this exploratory endpoint.
- The exploratory efficacy endpoint of number of hospitalizations in the periods between baseline and Week 24 and between baseline and Week 48 was removed from the table of objectives and endpoints (Section 2.4).
- The equation provided in Section 9.5.3.5 to calculate daily supplemental oxygen use assumes that the amount of time in a day in a state of rest is 960 minutes minus the number of minutes spent in a state of exertion, rather than the assumption stated in Section 8.2.6 of the protocol that oxygen use during rest equates to 24 hours per day.

12. REFERENCES

Akira, M., G. Inoue, S. Yamamoto, and M. Sakatani. 2000. 'Non-specific interstitial pneumonia: findings on sequential CT scans of nine patients', *Thorax*, 55: 854-9.

Arena, R., J. Myers, and L. A. Kaminsky. 2016. 'Revisiting age-predicted maximal heart rate: Can it be used as a valid measure of effort?', *Am Heart J*, 173: 49-56.

Fletcher, G. F., P. A. Ades, P. Kligfield, R. Arena, G. J. Balady, V. A. Bittner, L. A. Coke, J. L. Fleg, D. E. Forman, T. C. Gerber, M. Gulati, K. Madan, J. Rhodes, P. D. Thompson, M. A. Williams, Cardiac Rehabilitation American Heart Association Exercise, Council on Nutrition Physical Activity Prevention Committee of the Council on Clinical Cardiology, Council on Cardiovascular Metabolism, Nursing Stroke, Epidemiology Council on, and Prevention. 2013. 'Exercise standards for testing and training: a scientific statement from the American Heart Association', *Circulation*, 128: 873-934.

Graham, B. L., V. Brusasco, F. Burgos, B. G. Cooper, R. Jensen, A. Kendrick, N. R. MacIntyre, B. R. Thompson, and J. Wanger. 2017. '2017 ERS/ATS standards for single-breath carbon monoxide uptake in the lung', *Eur Respir J*, 49.

Hochberg, Y. 1988. 'A Sharper Bonferroni Procedure for Multiple Tests of Significance', *Biometrika*, 75: 800-02.

Inoue, Y., B. C. Trapnell, R. Tazawa, T. Arai, T. Takada, N. Hizawa, Y. Kasahara, K. Tatsumi, M. Hojo, T. Ichiwata, N. Tanaka, E. Yamaguchi, R. Eda, K. Oishi, Y. Tsuchihashi, C. Kaneko, T. Nukiwa, M. Sakatani, J. P. Krischer, K. Nakata, and Consortium Japanese Center of the Rare Lung Diseases. 2008. 'Characteristics of a large cohort of patients with autoimmune pulmonary alveolar proteinosis in Japan', *Am J Respir Crit Care Med*, 177: 752-62.

Kang M, Kendall MA, Ribaudo H, Tierney C, Zheng L, Smeaton L, Lindsey JC. 2022. 'Incorporating estimands into clinical trial statistical analysis plans', Clin Trials, 3:285-291.

Kokkinos, P., L. A. Kaminsky, R. Arena, J. Zhang, and J. Myers. 2017. 'New Generalized Equation for Predicting Maximal Oxygen Uptake (from the Fitness Registry and the Importance of Exercise National Database)', *Am J Cardiol*, 120: 688-92.

Lee, K. N., D. L. Levin, W. R. Webb, D. Chen, M. L. Storto, and J. A. Golden. 1997. 'Pulmonary alveolar proteinosis: high-resolution CT, chest radiographic, and functional correlations', *Chest*, 111: 989-95.
Miller, M. R., J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, and ATS ERS Task Force. 2005. 'Standardisation of spirometry', *Eur Respir J*, 26: 319-38.

Morrell, M. J. 2015. 'One hundred years of pulmonary function testing: a perspective on 'The diffusion of gases through the lungs of man' by Marie Krogh', *J Physiol*, 593: 351-2.

Quanjer, P. H., S. Stanojevic, T. J. Cole, X. Baur, G. L. Hall, B. H. Culver, P. L. Enright, J. L. Hankinson, M. S. Ip, J. Zheng, J. Stocks, and E. R. S. Global Lung Function Initiative. 2012. 'Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations', *Eur Respir J*, 40: 1324-43.

Ross, R. M. 2003. 'ATS/ACCP statement on cardiopulmonary exercise testing', *Am J Respir Crit Care Med*, 167: 1451; author reply 51.

Rubin, D. B. (1976), 'Inference and Missing Data,' Biometrika, 63, 581–592.

Stanojevic, S., B. L. Graham, B. G. Cooper, B. R. Thompson, K. W. Carter, R. W. Francis, G. L. Hall, Tlco working group Global Lung Function Initiative, and Tlco Global Lung Function Initiative. 2017. 'Official ERS technical standards: Global Lung Function Initiative reference values for the carbon monoxide transfer factor for Caucasians', *Eur Respir J*, 50.

Tazawa, R., B. C. Trapnell, Y. Inoue, T. Arai, T. Takada, Y. Nasuhara, N. Hizawa, Y. Kasahara, K. Tatsumi, M. Hojo, H. Ishii, M. Yokoba, N. Tanaka, E. Yamaguchi, R. Eda, Y. Tsuchihashi, K. Morimoto, M. Akira, M. Terada, J. Otsuka, M. Ebina, C. Kaneko, T. Nukiwa, J. P. Krischer, K. Akazawa, and K. Nakata. 2010. 'Inhaled granulocyte/macrophage-colony stimulating factor as therapy for pulmonary alveolar proteinosis', *Am J Respir Crit Care Med*, 181: 1345-54.

Trapnell, B.C., Y. Inoue, F. Bonella, C. Morgan, S. Jouneau, E. Bendstrup, I. Campo, S.A. Papiris, E. Yamaguchi, E. Cetinkaya, M.M. Ilkovich, M.R. Kramer, M. Veltkam, M. Kreuter, T. Baba, C. Ganslandt, I. Tarnow, G. Waterer, and T. Jouhikainen, for the IMPALA Trial Investigators. 2020. 'Inhaled Molgramostim Therapy in Autoimmune Pulmonary Alveolar Proteinosis', *NEJM*, 383: 1635-44.

Trapnell, B. C., and C. McCarthy. 2019. 'The Alveolar Lipidome in Pulmonary Alveolar Proteinosis: A New Focus for Therapeutic Development?', *Am J Respir Crit Care Med*.

Van Hout, B., M.F. Janssen, Y-S Feng, T. Kohlmann, J. Busschbach, D. Golicki, A. Lloyd, L. Scalone, P. Kind, A.S. Pickard. 2012. 'Interim Scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L Value Sets', *Value in Health*, 15(5): 708-715.

Weinreich, U. M., L. P. Thomsen, A. Hansen, S. Kjaergaard, P. D. Wagner, and S. E. Rees. 2013. 'Time to steady state after changes in FIO(2) in patients with COPD', *COPD*, 10: 405-10.

13. APPENDIX A – LIST OF TABLES, LISTINGS AND FIGURES

Mock shells for tables, listings, and figures will be provided in a separate document.

14. APPENDIX B – SCHEDULE OF ACTIVITIES

Double-Blind Treatment Period

Visit Name	S1	S2	BL	W4	W8	W12	W16	W20	W24	W36	W48	Early withdrawal[a]	Unscheduled
Visit ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	EW	UN
Visit window (days)	± 7	± 7	-	± 7	± 7	± 7	± 7	± 7	± 7	± 7	± 7	NA	NA
Weeks for Visit 3	-6	-3	-	4	8	12	16	20	24	36	48	NA	NA
Informed consent	X												
Medical history (including aPAP history)	X												
Prior and concomitant therapy	X	X	X	X	X	X	X	X	X	X	X	X	X
Demographics	X												
Resting 12-lead ECG	X		X	X		X			X		X	X	(X)
Resting vital signs and body weight	X	X	X	X	X	X	X	X	X	X	X	X	X
Spirometry and DLCO	X	X	X	X	X	X	X	X	X	X	X	X	(X)
Physical examination	X	X[c]	X[c]	X	X[c]	X	X[c]	X[c]	X	X	X	X	(X)
Need for suppl. O_2 use or WLL	X	X	X	X	X	X	X	X	X	X	X	X	(X)
ABG sample and resting respiration rate			X			X			X		X	X[d]	(X)
Exercise treadmill test		X[d]				X[d]			X[d]		X[d]	X[d]	(X)
Handout of eDiary		X											
SGRQ[e]			X	X	X	X	X	X	X	X	X	X	(X)
PGIS & PGIC		X[f]	X[f]	X	X	X	X	X	X	X	X	X	(X)
EQ-5D-5L[e]			X			X			X		X	X	(X)
Oxygen diary[g]			X	X	X	X	X	X	X	X	X		
DSS			X			X			X		X	X	(X)
HRCT		X[d]							X[d]			(X[d])	(X)
CGIS & CGIC			X[h]	X	X	X	X	X	X	X	X	X	
Blood sample for diagnostic anti-GM-CSF autoantibodies	X												
Blood sample for pregnancy test and contraceptive check[i]	X	X	X[j]	X	X	X	X	X	X	X	X	X	(X)

Visit Name	S1	S2	BL	W4	W8	W12	W16	W20	W24	W36	W48	Early withdrawal[a]	Unscheduled
Visit ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	EW	UN
Visit window (days)	± 7	± 7	-	± 7	± 7	± 7	± 7	± 7	± 7	± 7	± 7	NA	NA
Weeks for Visit 3	-6	-3	-	4	8	12	16	20	24	36	48	NA	NA
Samples for hematology, biochemistry and urinalysis	X	X[k]	X	X[l]	X[k]	X[l]	X[k]	X[k]	X	X[l]	X	X	(X)
Blood sample for biomarkers, anti-GM-CSF antibodies and, optionally, biobank[m]			X	X		X			X		X	X	(X)
Blood samples for GM-CSF, pre-dose and 2 hrs (±30 min) post-dose			X	X					X		X		
Adverse events	X	X	X	X	X	X	X	X	X	X	X	X	X
Eligibility criteria[n]	X	X	X										
Randomization			X										
IMP administration training[o]			X	X	(X)	(X)	(X)	(X)	(X)	(X)	(X)		(X)
IMP dosing in clinic			X[p]	X	X	X	X	X	X	X	X[p]		
Dispense IMP and ancillaries			X	X	X	X	X	X	X	X	X		(X)
Return used and un-used IMP				X	X	X	X	X	X	X	X	X	
Return ancillaries												X	
Treatment compliance				X	X	X	X	X	X	X	X	X	(X)
Subject instruction and diary	X	X[q]	X[q]	X	X	X	X	X	X[r]	X[r]	X[r]	X	(X)
Exit interview									X[s]				

Abbreviations: ABG=Arterial blood gas; anti-GM-CSF=Anti-granulocyte macrophage colony stimulating factor antibodies; BL=Baseline visit; CGIC=Clinician's global impression of change; CGIS=Clinician's global impression of severity; DLCO=Diffusing capacity of the lungs for carbon monoxide; DSS=Disease severity score; ECG=Electrocardiogram; EQ-5D-5L=EuroQoL 5 Dimensions, 5 Levels; FU=Follow-up; GM-CSF= Granulocyte macrophage colony stimulating factor; hr=Hour; HRCT=High resolution-computed tomography; IMP=Investigational medicinal product; min=Minute; NA=Not applicable; PGIC=Patient's Global Impression of Change; PGIC= Patient's global impression of change; PGIS= Patient's global impression of severity; S1/S2=Screening Visit 1/2; SGRQ=Saint George's Respiratory Questionnaire; suppl. O_2=Supplementary oxygen; W=Week; WLL=Whole lung lavage (or LL=Lung lavage); X=Mandatory procedure; (X)=Optional procedure to be performed if judged necessary by the Investigator.

a. The Early Withdrawal visit should be conducted if a subject is withdrawn from the trial before completion of the Week 48 visit. (See Section 1.3.2 of the protocol for procedures to be conducted at the Early Withdrawal visit for the open-label treatment period.)
b. An unscheduled visit can be conducted if deemed necessary by the Investigator.
c. Symptom-oriented or brief physical examination as clinically indicated.
d. The exercise treadmill test and the HRCT scan can be performed up to 3 weeks after Screening Visit 2, but prior to the Baseline visit. At all other timepoints these assessments can be performed within 7 days after the scheduled visit. These procedures, as well as ABG sample, should occur at the Early Withdrawal Visit only if the withdrawal occurs prior to Week 48.
e. SGRQ and EQ-5D-5L should be performed before any other trial procedures.
f. Only PGIS will be assessed at Screening Visit 2 and the Baseline visit. PGIS and PGIC should be completed immediately after the SGRQ and exercise treadmill test.
g. The oxygen diary should be completed daily, starting from 14 days prior to and until the visit (NB. Only applicable for subjects on supplemental oxygen).
h. Only CGIS will be assessed at the Baseline visit.
i. For visits with 12 weeks intervals (i.e., visits after Week 24), women of childbearing potential should also check pregnancy at home with monthly urine dipstick pregnancy tests.
j. A urine pregnancy test must also be performed at the Baseline visit, prior to first dosing.
k. Samples only include hematology at these visits.
l. Samples only include hematology and biochemistry at these visits.
m. Blood samples must be obtained before IMP dosing.
n. The eligibility criteria will be assessed to the extent they are available at Screening visits 1 and 2. At the Baseline visit, all eligibility criteria must be assessable and complied with for the subject to be randomized.
o. Re-training, marked as (X), can take place at all visits during the treatment period, if needed.
p. The subject will be observed for 1 hour after the first dose.
q. The Patient Journey sheet will also be handed out at Screening Visit 2 and collected at the Baseline visit.
r. Urine pregnancy test kits will also be provided to females of child-bearing potential.
s. Applicable for prospectively selected sites in North America and/or Europe. Can be performed up to 14 days after the Week 24 visit.

Open-Label Treatment Period

Visit Name	W52[a]	W60	W72	W84	W96	W108 W120 W132 W144	W148 Safety FU[b]	Early withdrawal[c]	Unscheduled[d]
Visit ID	V12	V13	V14	V15	V16	V17-20	V21	EW	UN
Visit window (days)	± 7	± 7	± 7	± 7	± 7	± 14	± 7	NA	NA
Prior and concomitant therapy	X	X	X	X	X	X	X	X	X
Resting 12-lead ECG			X		X	X		X	(X)
Resting vital signs and body weight		X	X	X	X	X	X	X	X
Spirometry and DLCO		X	X	X	X	X		X	(X)
Resting respiration rate			X		X			X	(X)
Physical examination		X[e]	X[e]	X[e]	X[e]	X[e]		X	(X)
Need for suppl. O_2 use or WLL		X	X	X	X	X		X	(X)
Return of eDiary						X[f]			
SGRQ[g]		X	X	X	X	X		X	(X)
PGIS & PGIC		X	X	X	X	X		X	(X)
EQ-5D-5L[g]			X		X	X		X	(X)
Oxygen diary[h]		X	X	X	X	X			
CGIS & CGIC		X	X	X	X	X		X	
Blood sample for pregnancy test and contraceptive check[i]		X	X	X	X	X		X	(X)
Samples for hematology, biochemistry, and urinalysis		X[j]	X[j]	X[j]	X[j]	X[j]		X	(X)
Blood sample for biomarkers, anti-GM-CSF antibodies and, optionally, biobank[k]			X		X	X	X[b]	X	(X)
Adverse events	X	X	X	X	X	X	X	X	X
IMP administration training[l]		(X)	(X)	(X)	(X)	(X)[m]			(X)
IMP dosing in clinic		X	X	X	X	X[m]			
Dispense IMP and ancillaries		X	X	X	X	X[m]			(X)
Return used and un-used IMP		X	X	X	X	X		X	
Return ancillaries						X[f]		X	
Treatment compliance		X	X	X	X	X		X	(X)
Subject instruction and diary		X[n]	X[n]	X[n]	X[n]	X[m]		X	(X)

Abbreviations: anti-GM-CSF=Anti-granulocyte macrophage colony stimulating factor antibodies; CGIC=Clinician's global impression of change; CGIS=Clinician's global impression of severity; DLCO=Diffusing capacity of the lungs for carbon monoxide; DSS=Disease severity score; ECG=Electrocardiogram; EQ-5D-5L=EuroQoL 5 Dimensions, 5 Levels; FU=Followup; GM-CSF=Granulocyte macrophage colony stimulating factor; IMP=Investigational medicinal product; NA=Not applicable; PGIC=Patient's Global Impression of Change; PGIC= Patient's global impression of change; PGIS= Patient's global impression of severity;

SGRQ=Saint George's Respiratory Questionnaire; suppl. O_2=Supplementary oxygen; W=Week; WLL=Whole lung lavage (or LL=Lung lavage); X=Mandatory procedure; (X)=Optional procedure to be performed if judged necessary by the Investigator.

a. The Week 52 visit is a safety telephone visit.
b. At the Week 148 visit, a blood sample for anti-GM-CSF antibodies will be obtained and any ongoing AEs at the Week 144 visit will be followed up.
c. The Early Withdrawal visit should be conducted if a subject is withdrawn from the trial before completion of the Week 144 visit. (See Section 1.3.2 of the protocol for procedures to be conducted at the Early Withdrawal visit for the double-blind treatment period.)
d. An Unscheduled visit can be conducted if deemed necessary by the Investigator.
e. Symptom-oriented or brief physical examination as clinically indicated.
f. eDiary and ancillaries are to be returned at Week 144 only.
g. SGRQ and EQ-5D-5L should be performed before any other trial procedures.
h. The oxygen diary should be completed daily, starting from 14 days prior to and until the visit (NB. Only applicable for subjects on supplemental oxygen).
i. Women of childbearing potential should also check pregnancy at home with monthly urine dipstick pregnancy tests.
j. Samples only include hematology and biochemistry at these visits, except at Weeks 96 and 144 when urinalysis is to be performed as well.
k. Blood samples must be obtained before IMP dosing.
l. Re-training, marked as (X), can take place at all visits during the treatment period, if needed (NB. Does not apply at Week 144).
m. Does not apply at Week 144.
n. Urine pregnancy test kits will also be provided to females of child-bearing potential.

15. APPENDIX C - ANALYSIS VISIT WINDOWS

The following analysis visit windows will be used for assessments summarized by post-baseline visit.

Scheduled Visit	Scheduled Trial Day [a]	Window Interval		Upper Limit in weeks/days	Interval Length [b]
		Lower Limit	Upper Limit		
Week 4	29	2	43	6 wks, 1 day	42
Week 8	57	44	71	10 wks, 1 day	28
Week 12	85	72	99	14 wks, 1 day	28
Week 16	113	100	127	18 wks, 1 day	28
Week 20	141	128	155	22 wks, 1 day	28
Week 24	169	156	211	30 wks, 1 day	56
Week 36	253	212	295	42 wks, 1 day	84
Week 48 [c]	337	296	379	54 wks, 1 day	84
Week 60 [d]	421	380	463	66 wks, 1 day	84
Week 72	505	464	547	78 wks, 1 day	84
Week 84	589	548	631	90 wks, 1 day	84
The following scheduled visits apply to subjects who do not consent or re-consent to protocol version 9.0 (or 8.1 for France only) nor 9.1:					
Week 96 [e]	673	632	687	98 wks, 1 day	56
Week 100 [f]	701	688	715	102 wks, 1 day	28
The following scheduled visits apply to subjects who consent or re-consent to protocol version 9.0 (or 8.1 for France only) or 9.1:					
Week 96	673	632	715	102 wks, 1 day	84
Week 108	757	716	799	114 wks, 1 day	84
Week 120	841	800	883	126 wks, 1 day	84
Week 132	925	884	967	138 wks, 1 day	84
Week 144 [e]	1009	968	1023	146 wks, 1 day	56
Week 148 [f]	1037	1024	1051	150 wks, 1 day	28

NAP = Not applicable
[a] Trial Day = visit date minus date of first dose of IMP + 1 day. If visit date is on or after the date of first dose.
Trial Day = visit date minus the date of first dose of IMP, if visit date is before the date of first dose.
Trial Day = 1 on the date of the first dose of IMP. Trial Day = -1 on the day before the first dose of IMP.
Scheduled Trial Day = 1 for Baseline; Scheduled Trial Day = (a x 7) + 1 for Week a
[b] Including both lower and upper limits.
[c] Excludes data after the subject has started open-label IMP.
[d] Excludes data before the subject has started open-label IMP.
[e] The last dose of IMP is the day before the Week 96 visit for subjects who do not consent or re-consent to protocol version 9.0 (or 8.1 for France only) nor 9.1 and the Week 144 visit for subjects who consent or re-consent to protocol version 9.0 (or 8.1 for France only) or 9.1.
[f] For vital signs, body weight, and anti-GM-CSF antibodies only.

16. APPENDIX D – DERIVATION OF SAINT GEORGE'S RESPIRATORY QUESTIONNAIRE

There are three components of the SGRQ: Symptoms, Activity, and Impacts. One total score is also calculated. SGRQ Total and component scores are scaled from 0 to 100, with higher scores indicating worse quality of life.

Principle of Calculation

Each questionnaire response has a unique empirically derived 'weight'. The lowest possible weight is zero and the highest is 100.

Each component of the questionnaire is scored separately in three steps:

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

   Score = 100 x (Summed weights from positive items in that component)/(Sum of weights for all items in that component)

The Total score is calculated in a similar way:

Score = 100 x (Summed weights from positive items in the questionnaire)/(Sum of weights for all items in the questionnaire)

Sum of maximum possible weights for each component and Total:

Symptoms	662.5
Activity	1209.1
Impacts	2117.8
Total	3989.4

(Note: These are the maximum possible weights that could be obtained for the worst possible state of the subject).

It will be noted that the questionnaire requests a single response to questions 1-7, 9-10, and 17. If multiple responses are given to one of these questions, then averaging the weights for the positive responses for that question are acceptable.

Symptoms Component

This is calculated from the summed weights for the positive responses to questions 1-8.

Activity Component

This is calculated from the summed weights for the positive responses to questions 11 and 15 (a total of 16 items to be completed).

Impacts Component

This is calculated from the summed weights for the positive responses to questions 9-10, 12-14, and 16-17 (a total of 26 items to be completed).

Total Score

The Total score is calculated by summing all positive responses in the questionnaire and expressing the result as a percentage of the total weight for the questionnaire. There are 17 questions in the questionnaire, with a total of 50 items to be completed).

Handling Missed Items

Note: The scoring allows for up to 24% of missing items in the questionnaire.

Symptoms

The Symptoms component will tolerate a maximum of 2 missed items. The weight for each missed item is subtracted from the total possible weight for the Symptoms component (662.5) and from the Total weight (3989.4).

Activity

The Activity component will tolerate a maximum of 4 missed items. The weight for each missed item is subtracted from the total possible weight for the Activity component (1209.1) and from the Total weight (3989.4).

Impacts

The Impacts component will tolerate a maximum of 6 missed items. The weight for each missed item is subtracted from the total possible weight for the Impacts component (2117.8) and from the Total weight (3989.4).

Total score

The Total score will tolerate a maximum of 12 missed items. The weight for each missed item is subtracted from the Total weight (3989.4).

Item Weights:

Part 1		
1) Over the last year, I have coughed:		
	Most	80.6
	Several	63.2

	A few	29.3
	Only	28.1
	Not	0.0
2) Over the last year, I have brought up phlegm (sputum):		
	Most	76.8
	Several	60.0
	A few	34.0
	Only	30.2
	Not	0.0
3) Over the last year, I have had shortness of breath:		
	Most	87.2
	Several	71.4
	A few	43.7
	Only	35.7
	Not	0.0
4) Over the last year, I have had attacks of wheezing:		
	Most	86.2
	Several	71.0
	A few	45.6
	Only	36.4
	Not	0.0
5) During the last year, how many severe or very bad unpleasant attacks of chest trouble have you had?		
	More than three	86.7
	3 attacks	73.5
	2 attacks	60.3

	1 attack	44.2
	None	0.0
6) How long did the worst attack of chest trouble last?		
	A week or more	89.7
	3 or more days	73.5
	1 or 2 days	58.8
	Less than a day	41.9
7) Over the last year, in the average week, how many good days (with little chest trouble) have you had?		
	None	93.3
	1 or 2	76.6
	3 or 4	61.5
	Nearly every day	15.4
	Every day	0.0
8) If you have a wheeze, is it worse in the morning?		
	No	0.0
	Yes	62.0
Part 2		
9) How would you describe your chest condition?		
	The most important problem I have	83.2
	Causes me quite a lot of problems	82.5
	Causes me a few problems	34.6
	Causes no problem	0.0
10) If you have ever had paid employment?		

	My chest trouble made me stop work	88.9
	My chest trouble interferes with my work or made me change my work	77.6
	My chest trouble does not affect my work	0.0
11) Questions about what activities usually make you feel breathless		
	Sitting or lying still	90.6
	Getting washed or dressed	82.8
	Walking around the home	80.2
	Walking outside on the level	81.4
	Walking up a flight of stairs	76.1
	Walking up hills	75.1
	Playing sports or games	72.1
12) More questions about your cough and breathlessness		
	My cough hurts	81.1
	My cough makes me tired	79.1
	I get breathless when I talk	84.5
	I get breathless when I bend over	76.8
	My cough or breathing disturbs my sleep	87.9
	I get exhausted easily	84.0
13) Questions about other effects your chest trouble may have on you		
	My cough or breathing is embarrassing in public	74.1
	My chest trouble is a nuisance to my family, friends, or neighbors	79.1

	I get afraid or panic when I cannot get my breath	87.7
	I feel that I am not in control of my chest problem	90.1
	I do not expect my chest to get any better	82.3
	I have become frail or an invalid because of my chest	89.9
	Exercise is not safe for me	75.7
	Everything seems too much of an effort	84.5
14) Questions about your medication		
	My medication does not help me very much	88.2
	I get embarrassed using my medication in public	53.9
	I have unpleasant side effects from my medication	81.1
	My medication interferes with my life a lot	70.3
15) Questions about how activities may be affected by your breathing		
	I take a long time to get washed or dressed	74.2
	I cannot take a bath or shower, or I take a long time	81.0
	I walk more slowly than other people, or I stop for rests	71.7
	Jobs such as housework take a long time, or I have to stop for rests	70.6
	If I walk up one flight of stairs, I have to go slowly or stop	71.6
	If I hurry or walk fast, I have to stop or slow down	72.3

	My breathing makes it difficult to do things such as walk up hills, carry things upstairs, light gardening such as weeding, dance, play bowls or play golf	74.5
	My breathing makes it difficult to do things such as carry heavy loads, dig the garden or shovel snow, jog or walk at least 5 miles per hour, play tennis or swim	71.4
	My breathing makes it difficult to such things such as very heavy manual work, run, cycle, swim fast or play competitive sports	63.5
16) We would like to know how your chest trouble usually affects your daily life		
	I cannot play sports or games	64.8
	I cannot go out for entertainment or recreation	79.8
	I cannot go out of the house to do the shopping	81.0
	I cannot do housework	79.1
	I cannot move far from my bed or chair	94.0
17) Tick the statement that you think best describes how your chest affects you		
	It does not stop me doing anything I would like to do	0.0
	It stops me doing one or two things I would like to do	42.0
	It stops me doing most of the things I would like to do	84.2
	It stops me doing everything I would like to do	96.7

17. APPENDIX E – SAMPLE OXYGEN EDIARY

Subjects on supplemental oxygen will complete a daily electronic diary about their oxygen use for 14 days prior to the relevant visits.

1	Have you used supplemental oxygen during the last 24 hours?	Yes	No
	If no to question 1, skip the rest of the questionnaire.		
2	Did you use oxygen at rest?	Yes	No
	If no to question 2, go to question 4		
3	What oxygen flow was used most often at rest?	____	L/min
4	Did you use oxygen during sleep?	Yes	No
	If no to question 4, go to question 6		
5	What oxygen flow was used most often during sleep?	____	L/min
6	Did you use oxygen during exertion?	Yes	No
	If no to question 6, skip the rest of the questionnaire		
7	What oxygen flow was used most often during exertion?	____	L/min
8	For how many hours did you exert and use supplemental oxygen?	____	Hrs.

18. APPENDIX F – EXAMPLE SAS CODE FOR MULTIPLE IMPUTATION PROCESS

Example code to facilitate control-based imputation, as described in Section 9.3:

Stage 1 – impute arbitrary missing data patterns:

```
proc mi data=<<input dataset>> seed=3157 nimpute=50 out=<<stage 1 output dataset>>;
    * Impute each treatment group independently;
    by trtan;

     * Impute missing results at Week 4 using baseline factors;
    fcs reg(y4 = base region dlco_ind /details);

    * Impute missing results at Week 8 using baseline factors and previous visit DLCO;
    fcs reg(y5 = y4 base region dlco_ind /details);

    * Impute missing results at Week 12 using baseline factors and previous visit DLCO;
    fcs reg(y6 = y5 y4 base region dlco_ind /details);

    * Impute missing results at Week 16 using baseline factors and previous visit DLCO;
    fcs reg(y7 = y6 y5 y4 base region dlco_ind /details);

    * Impute missing results at Week 20 using baseline factors and previous visit DLCO;
    fcs reg(y8 = y7 y6 y5 y4 base region dlco_ind /details);

    * Impute missing results at Week 24 using baseline factors and previous visit DLCO;
    fcs reg(y9 = y8 y7 y6 y5 y4 base region dlco_ind /details);

    var y4 y5 y6 y7 y8 y9 base region dlco_ind;
run;
```

Abbreviations: y4-y9= % predicted $DLCO_{adj}$ changes from baseline; base = baseline % predicted $DLCO_{adj}$; region=Regional stratification factor; dlco_ind=DLCO stratification factor at randomization.

Stage 2 – imputing missing data post-study withdrawal or (where applicable) post-intercurrent event:

```
proc mi data=<<stage 1 output dataset>> seed=3157 nimpute=1 out=<<stage 2 output dataset>>;
   by impdata;

     * Impute missing results at Week 4 using baseline factors;
    fcs reg(y4 = base region dlco_ind /details);
    mnar model(y4 / modelobs= (trtp='placebo'));

    * Impute missing results at Week 8 using baseline factors;
    fcs reg(y5 = base region dlco ind /details);
    mnar model(y5 / modelobs= (trtp='placebo'));

    * Impute missing results at Week 12 using baseline factors;
    fcs reg(y6 = base region dlco_ind /details);
    mnar model(y6 / modelobs= (trtp='placebo'));

    * Impute missing results at Week 16 using baseline factors;
    fcs reg(y7 = base region dlco_ind /details);
    mnar model(y7 / modelobs= (trtp='placebo'));

    * Impute missing results at Week 20 using baseline factors;
    fcs reg(y8 = base region dlco_ind /details);
    mnar model(y8 / modelobs= (trtp='placebo'));

    * Impute missing results at Week 24 using baseline factors;
    fcs reg(y9 = base region dlco_ind /details);
    mnar model(y9 / modelobs= (trtp='placebo'));

    var y4 y5 y6 y7 y8 y9 base region dlco_ind;
run;
```

Abbreviations: y4-y9= % predicted $DLCO_{adj}$ changes from baseline; base = baseline % predicted $DLCO_{adj}$; region=Regional stratification factor; dlco_ind=DLCO severity stratification factor at randomization.

Analyzing the imputation datasets individually (example for primary endpoint analysis):

```
proc mixed data=<<input dataset>>;
     by impdata; * split analyses by imputed dataset;
     class avisit trtp region dlco_ind;
     model chg= avisit trtp avisit*trtp base region dlco_ind /ddfm=kr;
     repeated avisit /subject=usubjid type=un;
     lsmeans trtp trtp*avisit/diff cl;
run;
```

Abbreviations: chg= % predicted $DLCO_{adj}$ changes from baseline; base = baseline % predicted $DLCO_{adj}$; region=Regional stratification factor; dlco_ind=DLCO severity stratification factor at randomization; usubjid = unique subject id; trtp = randomized treatment group.

Aggregating the inferential statistics using Rubin's rules (Rubin, 1976):

```
proc mianalyze data = <<treatment difference dataset>>;

    modeleffects estimate;
    stderr StdErr;
    ods output parameterestimates = <<output dataset>>;
run;
```

Example code to facilitate multiple imputation with a penalty assignment, as described in Section 9.4:

```
proc mi data=<<stage 1 output dataset>> seed=3157 nimpute=1 out=<<stage 2 output dataset>>;

   by impdata;

     * Impute missing results at Week 4 using baseline factors;
    fcs reg(y4 = base region dlco_ind /details);
    mnar model(y4 / modelobs= (trtp='placebo')  SHIFT=XXX);

    * Impute missing results at Week 8 using baseline factors;
    fcs reg(y5 = base region dlco_ind /details);
    mnar model(y5 / modelobs= (trtp='placebo')  SHIFT=XXX);

    * Impute missing results at Week 12 using baseline factors;
    fcs reg(y6 = base region dlco_ind /details);
    mnar model(y6 / modelobs= (trtp='placebo')  SHIFT=XXX);

    * Impute missing results at Week 16 using baseline factors;
    fcs reg(y7 = base region dlco_ind /details);
    mnar model(y7 / modelobs= (trtp='placebo')  SHIFT=XXX);

    * Impute missing results at Week 20 using baseline factors;
    fcs reg(y8 = base region dlco ind /details);
    mnar model(y8 / modelobs= (trtp='placebo')  SHIFT=XXX);

    * Impute missing results at Week 24 using baseline factors;
    fcs reg(y9 = base region dlco_ind /details);
    mnar model(y9 / modelobs= (trtp='placebo')  SHIFT=XXX);

    var y4 y5 y6 y7 y8 y9 base region dlco_ind;
run;
```

Abbreviations: y4-y9= % predicted $DLCO_{adj}$ changes from baseline; base = baseline % predicted $DLCO_{adj}$; region=Regional stratification factor; dlco_ind=DLCO severity stratification factor at randomization.

19. APPENDIX G – EXAMPLE SAS CODE FOR MODELING PROCEDURES

Repeated measures model regression of a continuous endpoint

The following SAS code is provided to illustrate fitting a repeated measures model to a continuous response variable (`chg`), adjusted for baseline (`base`), randomized treatment (`trtp`), stratification factors of region (`region`) and DLCO severity at randomization (`dlco_ind`), visit (`avisitn`), and the interaction of randomized treatment and visit (`avisitn*trtp`).

```
proc mixed data=<<input dataset>>;
     class avisitn trtp region dlco ind;
     model chg= avisitn trtp avisitn*trtp base region dlco_ind /ddfm=kr;
     repeated avisitn /subject=usubjid type=un;
     lsmeans trtp trtp*avisitn/diff cl;
run;
```

Logistic regression of a binary outcome

The following SAS code is provided to illustrate fitting a logistic model to a binary outcome variable (`chgcat1`), adjusted for baseline (`base`), randomized treatment (`trtp`) and stratification factors of region (`region`) and DLCO severity at randomization (`dlco_ind`).

```
proc logistic data=<<input dataset>> plots=(none);
  class trtp (ref='Placebo') region dlco_ind / param = ref;
  model chgcat1 = trtp region dlco_ind base / link=logit orpvalue;
  oddsratio trtp;
  ods output OddsRatiosWald=ORR1;
run;
```

Ordinal logistic regression of an ordinal classification factor

The following SAS code is provided to illustrate fitting a cumulative logit model to an ordinal outcome variable (`avalcat1`), adjusted for baseline (`base`), randomized treatment (`trtp`), and stratification factors of region (`region`) and DLCO severity at randomization (`dlco_ind`).

```
proc logistic data=<<input dataset>>;
  class trtp (ref='Placebo') region dlco_ind / param = ref;
  model avalcat1 = trtp region dlco_ind base /link=logit orpvalue;
  oddsratio trtp;
  ods output OddsRatiosWald=ORR1;
run
```

Negative binomial regression model with an offset of follow-up time

The following SAS code is provided to illustrate fitting a negative binomial regression model to a recurrent event response (`aval`), adjusted for randomized treatment (`trtp`), stratification factors of region (`region`) and DLCO severity at randomization (`dlco_ind`), and an offset variable which corresponds to the log-transformed duration of observation for each patient (`logoffset`).

```
proc genmod data=<<input dataset>>;
               class trtp (ref='Placebo') region dlco_ind;
               model aval=trtp region dlco_ind /dist=negbin offset=logoffset;
               lsmeans trtp/ cl diff e;
               ods output lsmeans=lsmeans1 diffs=diffs1;
run;
```

SAV006-05_SAP_2024-04-24 FINAL

Final Audit Report 2024-04-25

Created:	2024-04-25
By:	[redacted]
Status:	
Transaction ID:	

"SAV006-05_SAP_2024-04-24 FINAL" History

[redacted]