NCT02004691

STATISTICAL ANALYSIS PLAN

A Phase 2/3, multicenter, randomized, double-blinded, placebo-controlled, repeat-dose study to evaluate the efficacy, safety, pharmacodynamics, and pharmacokinetics of olipudase alfa in patients with acid sphingomyelinase deficiency

GZ402665-DFI12712

STATISTICIAN: [redacted]

DATE OF ISSUE: 4-Nov-2019

Total number of pages: 76

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

ACE:	angiotensin-converting enzyme
ADA:	anti-drug antibody
AE:	adverse event
ALT:	alanine aminotransferase
AP:	alkaline phosphatase
APR:	acute phase reactions
ASMD:	acid sphingomyelinase deficiency
AST:	aspartate aminotransferase
ATC:	anatomic category
$AUC_{0\text{-last}}$:	area under the curve extrapolate from time 0 to the time corresponding to the last quantifiable concentration
AUC0-τ:	area under the plasma concentration curve during a dosing interval τ
BMB:	bone marrow burden
BMD:	bone marrow density
BMI:	body mass index
BQL:	below the quantification limit
CL:	clearance
C_{max}:	maximum plasma concentration observed
CRS:	cytokine release syndrome
CSR:	clinical study report
CV:	coefficient of variation
DBP:	diastolic blood pressure
DBS:	dry blood spot
DLco:	diffusing capacity of lungs for carbon monoxide
DLT:	dose-limiting toxicity
DUP24:	Duplication 24
DXA:	dual-energy X-ray absorption
ECG:	electrocardiogram
eCRF:	electronic case report form
EQ-5D:	EuroQol –5 dimensions quality of life questionnaire
ETP:	extension treatment period
FACIT:	Functional Assessment of Chronic Illness Therapy
FEV_1:	forced expiratory volume in the first 1 second
FVC:	forced vital capacity
GGT:	gamma glutamyl-transferase
Hb:	hemoglobin
HDL:	high-density lipoprotein
HLGT:	high-level group term
HLT:	high-level term
HR:	heart rate
HRCT:	high-resolution computed tomography

IAR:	infusion associated reactions
ICF:	informed consent form
ILD:	infiltrative lung disease
IMP:	investigational medicinal product
INR:	international normalized ratio
ISS:	integrated summary of safety
IXRS:	interactive voice/web response system
LDL:	low-density lipoprotein
LFT:	liver function test
LLOQ:	lower limit of quantification
LLT:	lower-level term
MAR:	missing-at-random
MedDRA:	Medical Dictionary for Regulatory Activities
mITT:	modified intent to treat
MMRM:	mixed model for repeated measures
MN:	multiples of normal
MRI:	magnetic resonance imaging
NMR:	nuclear magnetic resonance
NPB-HAQ:	Niemann-Pick B Health Assessment Questionnaire
PAP:	primary analysis period
PCSA:	potentially clinically significant abnormality
PFT:	pulmonary function test
PGIC:	Patient Global Impression of Change
PGIS:	Patient Global Impression of Symptom Severity
PKDM:	pharmacokinetics, dynamics and metabolism
PRO:	Patient reported outcome
PT:	preferred term
ROW:	rest of world
SAE:	serious adverse event
SAP:	statistical analysis plan
SBP:	systolic blood pressure
SD:	standard deviation
SEM:	standard error of the mean
SF-36:	Short Form-36 Health Survey
SMQs:	standard MeDRA queries
SOC:	system organ class
SRS:	splenomegaly-related score
TEAE:	treatment-emergent adverse event
TLC:	total lung capacity
ULN:	upper limit of normal
US:	United States
VLDL:	very low-density lipoprotein
V_{ss}:	volume of distribution at steady state
WBC:	white blood cells
WHO-DD:	World Health Organization-Drug Dictionary
WMW:	Wilcoxon-Mann-Whitney

WT:	Wildtype
β-HCG:	β-human chorionic gonadotropin

1 OVERVIEW AND INVESTIGATIONAL PLAN

This statistical analysis plan (SAP) focuses on descriptions of strategy and statistical methodology for the analysis of clinical data from DFI12712 trial. It has been developed based on DFI12712 Amended Clinical Trial Protocol 11 dated 05 August 2019. This version replaces the previous versions of the SAP.

A clinical study report (CSR) will be prepared when the 52-week primary analysis period (PAP) data become available. At study completion, ie, at the end of the extension treatment period (ETP), a CSR for the entire study including ETP will be prepared.

1.1 STUDY DESIGN AND RANDOMIZATION

This Phase 2/3, multicenter, repeat-dose, clinical trial will be divided into 2 consecutive major periods: 1) a randomized placebo-controlled, double-blind PAP from Day -60 to Week 52 to be followed by 2) an ETP. Initially, the ETP will be double-blind as patients in the placebo arm cross over to active treatment. Study drug will be defined during the PAP as olipudase alfa or placebo, and during the ETP as olipudase alfa, and will be administered intravenously once every 2 weeks for 52 weeks in PAP, then for ETP. Day -60 to -1 is part of screening period with no study drug; some of the screening assessments which are not repeated right before the first infusion of study drug on Day 1 serve as baseline value for the study.

After completing the PAP, patients will enter the ETP of the study. Those who were randomized to the active arm will continue receiving the dose they were receiving at the completion of the PAP. Patients randomized to the placebo arm in the PAP will cross over to active treatment in the first infusion in ETP and will undergo dose escalation to a target dose of 3.0 mg/kg olipudase alfa.

For treatment periods during the PAP and the ETP, dose-limiting toxicity (DLT) criteria (protocol Section 8.1.4) and study stopping rules (protocol Section 8.1.5) will apply. If a patient satisfies protocol-specified criteria of rescue therapy (protocol Section 8.1.6), the individual treatment blind will be broken and the patient if on placebo will be switched to olipudase alfa; for patients already on olipudase alfa, the treatment will continue unless the patient drops out of study.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of this Phase 2/3 study is to evaluate the efficacy of olipudase alfa (recombinant human acid sphingomyelinase) administered intravenously once every 2 weeks for 52 weeks in adult patients with acid sphingomyelinase deficiency (ASMD) by assessing changes in: 1) spleen volume as measured by abdominal magnetic resonance imaging (MRI) (and for the United States [US] only, in association with patient perception related to spleen volume as measured by splenomegaly-related score [SRS]); and, 2) infiltrative lung disease as measured by the pulmonary function test DLco.

1.2.2 Secondary objectives

- To confirm the safety of olipudase alfa administered intravenously once every 2 weeks for 52 weeks
- To characterize the effect of olipudase alfa on the patient perception related to spleen volume as measured by SRS after 52 weeks of study drug administration (For the United States, the effect of olipudase alfa on SRS is part of the primary objective)
- To characterize the effect of olipudase alfa on the following endpoints assessed sequentially:
    1. The effect of olipudase alfa on liver volume after 52 weeks of study drug administration
    2. The effect of olipudase alfa on platelet count after 52 weeks of study drug administration
    3. The effect of olipudase alfa after 52 weeks of study drug administration on fatigue
    4. The effect of olipudase alfa after 52 weeks of study drug administration on pain
    5. The effect of olipudase alfa after 52 weeks of study drug administration on dyspnea

1.2.3 Additional objectives

- To characterize the effect of olipudase alfa on liver function tests (LFT)
- To characterize the effect of olipudase alfa on infiltrative lung disease via pulmonary imaging
- To characterize the effect of olipudase alfa on pulmonary functioning
- To characterize the effect of olipudase alfa on the fasting lipid profile
- To characterize the effect of olipudase alfa on bone disease
- To characterize the effect of olipudase alfa on cardiopulmonary functioning
- To characterize the effect of olipudase alfa on biomarkers
- To characterize the effect of olipudase alfa on hematology parameters
- To characterize the effect of olipudase alfa on selected health-related quality of life questionnaires and questions therein
- To characterize the pharmacodynamic effect of olipudase alfa on clearing sphingomyelin accumulation in liver and/or blood
- To characterize the multiple-dose plasma pharmacokinetic profile of olipudase alfa
- To explore the effect of olipudase alfa on Physician's global assessment of change
- To explore the effect of olipudase alfa on the nuclear magnetic resonance (NMR) profile of high-density lipoprotein (HDL)
- To explore the effect of olipudase alfa on liver function via echo-Doppler ultrasound
- To explore the effect of olipudase alfa on inflammatory and vascular biomarkers

1.3 DETERMINATION OF SAMPLE SIZE

The sample size calculations are based on the 2 primary efficacy endpoints:

1. Percentage change in DLco (expressed as % predicted of normal) from baseline to Week 52

2. a) For Europe and the rest of the world (ROW) outside of USA: the percentage change in spleen volume (expressed in multiples of normal [MN]) from baseline to Week 52

   b) For US: the combination spleen endpoint which consists of two components: the percentage change in spleen volume (expressed in MN) from baseline to Week 52 and change in splenomegaly-related score from baseline to Week 52.

Although the multiplicity adjustments in statistical testing of hypotheses ensures the study to be positive if only one of the two primary endpoints is significant (Section 2.4.5.3), each endpoint has been powered adequately in order to choose the sample size for the study as 36 randomized patients in 1:1 ratio in placebo and olipudase alfa treatment arms.

Endpoint	Null hypothesis	Assumptions			Signif. level	Sample Size N	Power
		Trt Diff	SD	Non-evaluable			
DLco	There is no difference between olipudase alfa group and placebo group in percentage change from baseline in percent predicted DLco from baseline to Week 52.	25%	20%	11%	0.05	36	0.93
Spleen Volume	There is no difference between olipudase alfa group and placebo group in percentage change from baseline in spleen volume in multiples of normal from baseline to Week 52.	30%	11.8%	11%	0.05	36	0.99
SRS	There is no difference between olipudase alfa group and placebo group in change from baseline in SRS from baseline to Week 52.	8	9.4	11%	0.15	36	0.82

The following assumptions for the DLco endpoint are based on results obtained from the olipudase alfa Phase 1b study DFI13412:

- A 20% common standard deviation
- A 25% mean difference from baseline to Week 52 between olipudase alfa and placebo in percentage change in DLco (in % predicted)
- An expected exclusion rate from the primary analysis due to non-availability of results at Week 52 is 11%

Based on these assumptions, a comparison between the olipudase alfa arm and the placebo arm will have 93% power using a t-test at a 2-sided 5% significance level and with 36 patients randomized 1:1 to placebo and olipudase alfa.

The assumptions for the spleen volume endpoint are the following:

- An 11.8% common standard deviation based on data from previous ASMD and Gaucher disease Type 1 studies
- A 30% mean difference from baseline to Week 52 between olipudase alfa and placebo in percentage change in spleen volume (MN)
- An expected exclusion rate from the primary analysis due to non-availability of results at Week 52 is 11%

Based on the above assumptions, a comparison between olipudase alfa and placebo will have over 99% power using a t-test at a 2-sided 5% significance level with 36 patients randomized 1:1 to placebo and olipudase alfa. At 2.5% significance level, the power will still be over 99%.

The assumptions for the splenomegaly-related score endpoint came from the Sanofi-sponsored clinical trial in myelofibrosis (JAKARTA, NCT# 01437787). The assumptions for the power calculations are the following:

- A common standard deviation of 9.4
- A mean difference of 8.0 from baseline to Week 52 between olipudase alfa and placebo in the SRS score
- An expected exclusion rate from the primary analysis due to non-availability of results at Week 52 is 11%

Based on the above assumptions, a comparison between olipudase alfa and placebo will have 82% power using a t-test to detect a statistical trend, defined as 2-sided p-value $\leq$0.15, with 36 patients randomized 1:1 to placebo and olipudase alfa.

Using the simplifying assumption that the 2 components of the combination spleen endpoint are independent, the likelihood that statistical significance will be declared for the combination spleen endpoint is greater than 80%.

1.4 STUDY PLAN

This Phase 2/3, multicenter, repeat-dose, clinical trial will be divided into 2 consecutive major periods: 1) a randomized placebo-controlled, double-blind PAP from Day -60 to Week 52, to be followed by 2) ETP for up to 4 years. Initially, the ETP will be double-blind as patients in the placebo arm cross over to active treatment, and the last patient completes PAP and when this patient is going through dose escalation the study is unblinded for analysis of PAP; ETP after unblinding for PAP analysis is open-label where all patients receive olipudase alfa.

The PAP will include a screening/baseline period and a treatment period. Approximately 36 patients who meet the eligibility criteria will be randomized in a 1:1 ratio across sites into 1of 2 arms, placebo or 3.0 mg/kg olipudase alfa target dose. Patients will undergo a dose-escalation period to achieve the target dose level.

After completing the PAP, patients will enter the ETP. Patients randomized to the placebo arm in the PAP will cross over to active treatment in the ETP and will undergo dose escalation to a target dose of 3.0 mg/kg olipudase alfa. The patients who were randomized to the active arm will continue receiving olipudase alfa at the dose they complete PAP (with a mock dose escalation period to maintain blinding).

Study drug will be administered intravenously once every 2 weeks. In-patient hospitalization is required during dose escalation period and specific visits. Other infusions are administered at the clinic.

All patients will have an end of study/early discontinuation visit within 2 weeks after the last infusion of study treatment. In addition, all patients will receive a follow-up phone call for safety approximately 30 to 37 days after the last infusion of study drug.

A rescue strategy for patients who experience significant clinical decline (specific guidelines in the protocol, Section 8.1.6) will be in place: these individual patients may be unblinded and if receiving placebo, will be switched to olipudase alfa with proper dose escalation to a target dose of 3.0 mg/kg. These patients will continue to be followed to the end of study.

1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

This section summarizes discrepancies between the statistical analysis described in the amended DFI12712 protocol 11, dated 12 August 2019, and this SAP.

- The following language about the summary of immune response assessments in the protocol Section 11.5.5.4 will not be provided by the dose level of the infusion before or during the evaluation, since the infusion information will be included in the data listing.

  The presence, absence, and titers of serum anti-olipudase alfa IgG antibodies and routine levels of acute phase reactants will be summarized over time overall and by the dose level of the infusion before or during the evaluation.

- The following language about the exploratory pharmacodynamic analyses in the protocol Section 11.5.6.2 is not planned in this SAP but may be done ad hoc if deemed necessary.

  Exploratory correlation analyses will be attempted to evaluate relationships between different PD markers and between sphingomyelin and /or lysosphingomyelin levels from different sources.

- Multiplex assay of inflammatory and vascular biomarkers are analyzed as both exploratory efficacy and safety endpoints in the protocol but will be treated only as safety endpoints in this SAP.

- The ETP population mentioned in the protocol Section 11.4 was removed from the SAP, since data for PAP+ETP will be analyzed in a cumulative way.

- The annualized rate of TEAEs and separate analyses of AEs by demographic characteristics (eg, age and sex) mentioned in the protocol Section 11.5.5.1 will be presented in the integrated summary of safety and will not be done in the study SAP.

- Protocol Section 11.5.6.3 for pharmacokinetic-pharmacodynamic analyses will not be done at the study level, but may be conducted using pooled data from different clinical studies and the results may be included in a separate report.
- For the summary of exposure, the protocol mentioned using the safety population. However, the population was changed to mITT population in the SAP in order to accurately summarize the duration of exposure prior to rescue therapy for patients who were randomized to placebo at the beginning.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

An updated full statistical analysis plan dated 29 August 2019 was submitted to FDA for their review and agreement. This document builds on that analysis plan. The main modifications made in this statistical analysis plan are listed as follows.

- A subset of protocol-defined IARs was further defined programmatically by searching for preferred terms within Hypersensitivity reaction SMQs in Section 2.4.6.1.2.
- The responder analyses for BFI scale - Item 3, BPI scale - Item 3, and FACIT-Dyspnea symptom score were added in Section 2.4.5.5 based on the interactions with health authorities.
- Numerical thresholds for meaningful within-patient improvement on the SRS, BFI scale - Item 3, BPI scale - Item 3, and FACIT-Dyspnea symptom score were added in Section 2.4.5.5 based on the interactions with health authorities.
- Cumulative distribution functions (CDFs) for the SRS, BFI scale - Item 3, BPI scale - Item 3 and FACIT-Dyspnea symptom score by treatment group were added in Section 2.4.5.5 based on the interactions with health authorities.
- The definitions in Section 2.4.6.12 in the Immune assessment section were expanded and refined.
- The compliance of eDiary data was added in Section 2.4.5.6 based on the interactions with health authorities.

This SAP supersedes any previous SAPs.

2 STATISTICAL AND ANALYTICAL PROCEDURES

The PAP analysis will be performed after all patients have had the chance to complete the Week 52 assessments (ie, at the completion of the 52-week PAP treatment period) and will include a formal database lock, unblinding of the study treatment arms, and analyses of efficacy, safety, pharmacodynamic, and pharmacokinetic data. The study treatments will remain blinded to the Sponsor until completion of the PAP; the investigator and patients will remain blinded until the last patient completes the dose escalation period in the ETP period. A CSR will be prepared for the PAP analysis. The database will contain all visits, events or procedures that occurred up to a cutoff date - to be pre-defined as few days after the last patient finishes PAP, and the data will be cleaned to a pre-specified extent. This will include PAP evaluations for all patients randomized, but will also include evaluations in ETP for most patients.

A formal summary of data as a CSR during the ETP may be produced to support regulatory approval(s) and/or other submission/application requirement(s).

The final CSR will be prepared after all patients have completed all the study assessments at the completion of the ETP and will include a final database lock. Any difference in data from PAP CSR will be highlighted within the final CSR.

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value in general is defined as the last available value before first infusion. It is understood that these evaluations may be before or after the randomization of the patient in the interactive voice/web response system (IXRS). One exception to this baseline definition is the hematology parameters – these samples are collected twice at screening, and twice on first infusion visit (once before and once after the first infusion); the average of all measurements taken before the first infusion will serve as the baseline for hematology parameters. Another exception may be the ECG parameters – Day 1 (the first infusion visit) is to collect ECG in triplicate before the infusion; if these data qualify for the latest ECG before first infusion, the average of ECG parameter results from the triplicate ECG will be considered as the baseline value. Since all other time points including screening will collect a single ECG, if the Day 1 pre-infusion triplicate ECG does not qualify for baseline, the baseline value of ECG parameters will be taken from the single ECG at the qualifying time point.

All baseline safety and efficacy parameters (apart from those listed below) will be presented along with the respective safety and efficacy displays.

Demographic characteristics

Demographic variables are gender (Male, Female), race (American Indian or Alaska Native; Asian; Black; Native Hawaiian or Other Pacific Islander; White; other), age in years at screening,

ethnicity (Hispanic, non-Hispanic), ancestry (with special details of Jewish ancestry), country, human immunodeficiency virus (HIV) antibody testing, and Hepatitis tests (B surface antigen and C antibody tests).

Medical history

Complete medical history includes prior or existing condition(s) in different body systems as well as prior or existing medical conditions/surgical procedures of specific interest.

Disease characteristics at baseline

Baseline disease characteristics include age at ASMD diagnosis, number of years since ASMD diagnosis, acid sphingomyelinase (ASM) activity in peripheral leukocytes and dried blood spot, spleen status, genotyping of *SMPD1*, *CHIT1*, *UGT1A1*, severe splenomegaly and severe % predicted DLco adjusted for hemoglobin and ambient barometric pressure.

The SMPD1 variant type (ie, missense, frameshift, deletion, nonsense, etc) will be derived if possible from the SMPD1 amino-acid change.

For the classification of CHIT1 genotype, the following derivation will be applied:

- N= Wildtype (WT)/WT = Normal / 2 functional alleles
- H= WT/Duplication 24 (DUP24) = Heterozygous mutation / 1 functional allele
- M = DUP24/DUP24 = homozygous mutation / 2 non-functional alleles (no chito activity)

For the classification of UGT1A1 genotype, the following derivation will be applied:

- UGT1A1*1 = Fully functional allele = Not Gilbert variant
- UGT1A1*36 = Functional allele = Not Gilbert variant
- UGT1A1*28 = Gilbert syndrome allele = Gilbert variant
- UGT1A1*37 = Crigler-Najar syndrome type 2 allele (impaired UGT1A1) = Gilbert variant

Patients with 0 or 1 Gilbert variant alleles will be classified as “no Gilbert Syndrome”; patients with 2 Gilbert variants alleles will be classified as “Gilbert Syndrome”.

2.1.2 Prior or concomitant medications

All medications and therapeutic procedures received by the patient in the 30 days before the first scheduled infusion with study drug until the final safety follow-up phone call 30-37 days after the last infusion will be recorded on the Prior and Concomitant Medication eCRF.

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD).

- Prior medications are those the patient used prior to first intake of study drug, either placebo or olipudase alfa. Prior medications can be discontinued before first administration or can be ongoing during treatment phase.

- Concomitant medications are any treatments received by the patient concomitantly to the study drug (either placebo or olipudase alfa), as medications that are taken on and after the first infusion of study drug. A given medication can be classified both as a prior medication and as a concomitant medication.

Technical details related to computation, dates, imputation for missing dates are described in Section 2.6.6.

2.1.3 Smoking and alcohol drinking

Alcohol and tobacco usage data will be collected at screening and yearly visits.

2.1.4 Efficacy endpoints

2.1.4.1 Primary efficacy endpoint(s)

The primary efficacy endpoints for Europe and the ROW are:

- Percentage change in spleen volume (in MN) from baseline to Week 52
- Percentage change in DLco (in % predicted of normal) from baseline to Week 52

The primary efficacy endpoints for US are:

- Combination spleen endpoint which has two components:
  - Percentage change in spleen volume (in MN) from baseline to Week 52
  - Change in splenomegaly-related score from baseline to Week 52
- Percentage change in DLco (in % predicted of normal) from baseline to Week 52

DLco measurements

DLco is part of the pulmonary function tests (PFT), which will be performed at screening to measure, among other parameters, the diffusing capacity of carbon monoxide (DLco) as evidence of gas exchange across the alveolocapillary membrane for infiltrative lung disease (ILD). This parameter is measured at screening and approximately every 6 months after the first infusion in the study; for further details of timing, please refer to the protocol.

Observed values of DLco will be reported by the central vendor using the readings from the machine that the vendor has provided to each site. Percent predicted values will be computed based on following algorithm (height and hemoglobin values are from the available ones closest to the PFT evaluation date/time, and age is calculated at the PFT evaluation date):

Percent predicted DLco will be computed as follows (1, 2):

Predicted DLco (adults), mL CO/min/mmHg:
= 0.416*Height (cm) - 0.219*Age (years) - 26.34 for males
= 0.256*Height (cm) - 0.144*Age (years) - 8.36 for females

Unit conversion: mL CO/min/mmHg = mmol/min/kPa *2.987

Hb-adjustment factor (to be applied to DLco in unit of mL CO/min/mmHg, Hb in g/dL)
= 1 / [(1.7*Hb) / (10.22 + Hb)] for males
= 1 / [(1.7*Hb) / (9.38 + Hb)] for females

Ambient barometric pressure (Pamb) in unit of hPa is recorded in the vendor machine, which is multiplied by 0.75 to covert to the unit of mmHg. Water vapor partial pressure in lungs is assumed to be 47mmHg. The partial pressure of (inspired) oxygen (PiO2) in the lungs is calculated as:

PiO2 = (Pamb – 47)*0.20942

where 0.20942 is assumed to be the oxygen concentration in room air.

Altitude-adjustment factor = 1 + 0.0031*(PiO2-150)

DLco-adjustment factor = Hb-adjustment factor * Altitude-adjustment factor

DLco adjusted = Observed DLco*DLco-adjustment factor

Percent predicted adjusted DLco = 100*DLco adjusted/Predicted DLco

Note that percent predicted DLco adjusted for hemoglobin and ambient barometric pressure is a part of inclusion criteria in the study. For ease of operation, the vendor-provided machine calculation of percent predicted DLco adjusted for Hb and ambient barometric pressure is used for determination of eligibility, and remain as part of the source documents; these data are not included in the database. Every effort has been made to match the formula the vendor uses in the machine, however, the screening DLco for analyses purposes is calculated separately using the data available in the database. In the rest of this SAP, DLco (% predicted) always refers to the percent predicted DLco adjusted for hemoglobin and ambient barometric pressure.

Spleen volume – Abdominal magnetic resonance imaging

Spleen volume will be assessed by abdominal MRI (to quantify the degree of splenomegaly and hepatomegaly) at screening and approximately every 6 months since the first infusion in the study; for further details of timing, please refer to the protocol. The MRIs will be collected and read centrally by a third party blinded to patient number and study visit.

The vendor will provide spleen volume in the unit of cubic millimeters. This will be converted to the standard unit of cubic centimeters for reporting purposes, by dividing by 1000. The analysis will also use the unit of Multiples of Normal (MN), and the algorithm used is as follows (weight coming from the respective visit; if not available, the closest one available):

Spleen Volume (MN) = Spleen volume (cm^3) / [2*weight (kg)]

At screening one of the inclusion criteria requires spleen volume in multiples of normal. For operational purposes, the vendor provides these data (in MN) to the site at screening, using the same formula above, taking the spleen volume (cubic millimeters) and the weight (kg) from the screening visit. These data from vendor are used for determination of eligibility, and remain as part of the source documents; these data are not included in the database. For purposes of this analysis, the screening spleen volume is calculated separately using available spleen volume (cubic millimeters) and weight (kg) from the screening visit.

Splenomegaly-related score (primary for US)

The splenomegaly-related score is defined as the 7-day mean of the daily sum of scores of the following 5 symptom assessments: abdominal pain, abdominal discomfort, early satiety, abdominal body image, and ability to bend down. The 5 symptoms were selected from the Myelofibrosis-Symptom Assessment Form. Using a numerical rating scale of 0 (absent) to 10 (worst imaginable), these questions assess the effect of splenomegaly-related symptoms within the last 24 hours of assessment.

An eDiary will be used for collecting these responses. This is collected over 7 evenings before the clinic visit - at screening, at randomization visit, then at every quarterly visit during the study. At screening the patient must fill out at least six of the seven days of the eDiary; if not, the patient gets to repeat the eDiary for 7 days. If the patient did not fill out the 7^{th} day in the eDiary, the patient will fill in last day (Day 7) responses while at the clinic.

The sum of the individual responses will be used in the evaluation of total daily score. A missing response is not allowed in the eDiary, but in case of an incomplete questionnaire (eg, eDiary switched off before finishing the questionnaire), the score for that day will be missing. The mean of the 7 consecutive daily scores from the eDiary will be used as the splenomegaly-related score. If there are days with missing scores, then the mean of the available daily scores from the same patient and same time point will be used if at least 4 out of 7 days (>50%) of data are not missing, otherwise, the mean of the 7 consecutive daily score will be set to missing.

2.1.4.2 Secondary efficacy endpoint(s)

- Percentage change in liver volume (in MN) from baseline to Week 52
- Percentage change in platelet counts from baseline to Week 52
- Week 52 change from baseline in fatigue severity as measured by item 3 of the BFI scale
- Week 52 change from baseline in pain severity as measured by item 3 of the BPI scale
- Week 52 change from baseline in dyspnea severity as measured by the Functional Assessment of Chronic Illness Therapy (FACIT)-Dyspnea tool
- Change in splenomegaly-related score from baseline to Week 52 (except in US where this endpoint is part of the primary “combination spleen endpoint”)

Liver volume - abdominal magnetic resonance imaging

Liver volumes will be assessed by abdominal MRI (to quantify the degree of splenomegaly and hepatomegaly) at screening and approximately every 6 months since the first infusion in the study; for further details of timing, please refer to the protocol. The MRIs will be collected and read centrally by a third party blinded to patient number and study visit.

The vendor will provide liver volume in the unit of cubic millimeters. This will be converted to the standard unit of cubic centimeters for reporting purposes, by dividing by 1000. The analysis will also use the unit of MN, and the algorithm used is as follows (weight coming for respective visit; if not available, the closest one available):

Liver Volume (MN) = Liver volume (cm^3) / [25*weight (kg)]

Platelet count - hematology

Platelet count is part of hematology and hemogram panel with laboratory testing of blood samples. Hemogram panel evaluates hemoglobin, hematocrit, platelet and white blood cells (total and deferential) only. Platelet count is evaluated at screening, randomization visit, at every visit during dose escalation (both in PAP and ETP) period, and at every quarterly visit. At quarterly visits except for yearly visits, and during dose-escalation visits, one sample of hematology is collected pre-infusion. At yearly visits, two samples - one for hematology, one for hemogram - at least 24 hours and up to 2 weeks apart, before infusion, are evaluated. During screening period, two samples - one for hematology, one for hemogram - at least 24 hours and up to 2 weeks apart are evaluated. In addition, about 24 hours post- infusion one sample is collected at all visits except for screening. In all the visits where 2 samples are collected, platelet count and hemoglobin will be reported as average between the 2 samples. For details of sample collection schedule, please refer to the protocol.

BFI scale - Item 3

The BFI is a validated, self-administered questionnaire that was originally developed to assess fatigue severity in cancer patients (3). The BFI Item 3 is about the worst fatigue severity in past 24 hours, using numeric rating scales from 0 to 10, 0 being no fatigue, 10 being the worst fatigue. This particular item is collected in the eDiary prior to baseline and quarterly visits. The mean of the 7 consecutive daily measurements recorded in the eDiary will be used as the BFI Item 3 score. If the patient did not fill out the eDiary on the day before the clinic visit (Day 7), the patient will fill in Day 7 responses while at the clinic. If there are missing daily measurements, then the mean of the available measurements within the same patient and the same time point will be used if at least 4 out of 7 days (>50%) of data are not missing, otherwise, the mean of the 7 consecutive daily score will be set to missing.

BPI-SF scale - Item 3

The BPI-SF is a validated, self-administered questionnaire designed to measure a patient's perceived level of pain (4, 5, 6, 7). The BPI-SF Item 3 is about the worst pain severity in past 24 hours, using numeric rating scales from 0 to 10, 0 being no pain, 10 being the worst pain. This

particular item is collected in the eDiary prior to baseline and quarterly visits. The mean of the 7 consecutive daily measurements recorded in the eDiary will be used as the BPI-SF Item 3 score. If the patient did not fill out the eDiary on the day before the clinic visit (Day 7), the patient will fill in Day 7 responses while at the clinic. If there are missing daily measurements, then the mean of the available measurements within the same patient and the same time point will be used if at least 4 out of 7 days (>50%) of data are not missing, otherwise, the mean of the 7 consecutive daily score will be set to missing.

FACIT-Dyspnea score

The complete FACIT-Dyspnea short form consists of 10 symptom-specific questions, and each of these questions has a related question asking about functional limitations (8). For the secondary endpoint, only the symptom-specific questions of the FACIT-Dyspnea will be used, and the functional limitation questions will not be used. The answer to each question can take the following values:

A) Numeric values representing severity ranging from 0 (no shortness of breath) to 3 (severely short of breath). These responses will be entered into the scoring algorithm as described below:
   - A response that the activity was not attempted specifically because of the severity of shortness of breath associated with it will be entered into the scoring algorithm as a 3 (severe shortness of breath).
   - A response that the activity was not attempted for any other reason than shortness of breath will not be assigned a numeric value and will be excluded from the scoring algorithm.
   - The raw score for the FACIT-Dyspnea will be determined by the following rules:
     i. If all 10 questions have numeric values then the raw score is the sum of the 10 values.
     ii. If 50% or fewer of the questions have numeric results then the raw score will not be determined, ie, assigned a missing value.
     iii. Otherwise the available values will be added, the sum will be multiplied by 10, and then divided by the number of questions answered.

B) The raw score as calculated above is then converted to a scale score using a conversion table provided by the developers of the questionnaire (8). The scale scores for FACIT-Dyspnea have a range of 27.7 to 75.9; conversion table is provided in Section 7.1.

FACIT-Dyspnea questionnaire is part of Day 7 in eDiary and is administered prior to baseline and at every quarterly visit since the first infusion. If the patient misses Day 7 eDiary, he is allowed to complete Day 7, including FACIT, at the clinic visit.

Splenomegaly-related score

This is a primary endpoint in US; for details please refer to Section 2.1.4.1.

2.1.4.3 Additional efficacy endpoints

Liver function tests

Percentage change from baseline to Week 52 in Liver function tests (pre-infusion): alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, and direct bilirubin.

Pulmonary function tests

Pulmonary function tests (PFT) will be performed at screening and at approximately every 6 months since the first infusion to measure lung volumes and air flow; for further details of timing, please refer to the protocol.

Pulmonary function testing will include, but will not be limited to, assessment of forced vital capacity (FVC), forced expiratory volume in the first 1 second (FEV_1) of the FVC maneuver, and total lung capacity (TLC). Observed values of PFT parameters will be reported by the central vendor from machine reading. Percent predicted values will be computed based on following algorithm (height values are from the available one closest to the PFT evaluation date/time, and age is calculated at the PFT evaluation date):

Percent predicted values of FEV1 and FVC will use the Global Lung Initiative 2012 reference equations (9).

Predicted TLC (liters) will use the following formula for adults (1):

Adults (≥18 years of age): = 0.0795*height (cm) + 0.0032*Age (years) – 7.333 for males
= 0.0590*height (cm) – 4.537 for females

Percent predicted TLC = 100*Observed TLC / Predicted TLC

Pulmonary function imaging – high-resolution computed tomography

High-resolution computed tomography (HRCT) scans of the chest will be obtained at screening and approximately every 6 months since the first infusion; for further details of timing, please refer to the protocol. This evaluation is meant to quantify the degree of possible diffuse lung disease, classically referred to as an interstitial lung disease (ILD).

Images will be collected centrally by a medical imaging core laboratory where they will be digitized and coded and read by a third party blinded to patient number and study visit. The mean density in Hounsfield units will be reported at each lung level using standard HRCT software. The lung fields will be assessed by a central reader and scored subjectively for the degree of diffuse lung disease (ILD) (0=normal, 1=mild, 2=moderate, or 3=severe).

The bilateral lung boundaries are determined from the following 4 pre-defined levels on the HRCT images.

- Level 1 includes the aortic arch.
- Level 2 includes the carina.

- Level 3 is midway between the carina and 1 centimeter above the hemidiaphragm.
- Level 4 is 1 centimeter above the hemidiaphragm.

A qualitative assessment will be made of the components of diffuse lung disease, classically referred to as interstitial lung disease (ILD): ILD (read specifically as reticular appearance), ground glass appearance, reticulo-nodular density (read specifically as nodular appearance), and pleura thickening of the right and left lungs for each pre-defined levels above according to the following criteria:

- 0 = No disease.
- 1 = Mild (affecting 1% to 25% of the lung volume).
- 2 = Moderate (affecting 26% to 50% of the lung volume).
- 3 = Severe (affecting 51% to 100% of the lung volume).
- Not evaluable.

The mean scores at each scheduled visit for each of the 4 qualitative assessments are calculated as follows:

Mean score across levels for left (or right) lung = Sum of scores for X levels for left (or right) lung / X

Mean score across 4 levels and both lungs = (Mean score across X levels for left lung + Mean score across X levels for right lung) / 2.

No overall assessment of diffuse lung disease caused by ASMD is obtained. Rather, the key HRCT radiological findings are each scored individually. In the case of ILD (read specifically as reticular appearance), ground glass appearance, and reticulo-nodular (read specifically as nodular appearance), an improvement in any one of these findings may be indicative of a radiologic improvement in the diffuse lung disease.

Pulmonary function imaging – chest X-ray

A chest X-ray (posterior-anterior and lateral views) will be performed at screening and approximately every 6 months since the first infusion at sites, except for the patients in Germany (since Germany radiation board did not allow this); for further details of timing, please refer to the protocol.

Chest X-rays will be collected at the site and sent to a medical imaging core laboratory where they will be digitized (if not already done at the site), coded, and read by a third-party reader blinded to patient number and study visit. The lung fields will be scored subjectively for the degree of ILD - interstitial, reticular, nodular density, consolidation and pleura thickening of the right and left lungs separately according to the following criteria:

- 0 = No infiltrative lung disease.
- 1 = Mild (affecting 1% to 25% of the lung volume).

- 2 = Moderate (affecting 26% to 50% of the lung volume).
- 3 = Severe (affecting 51% to 100% of the lung volume).

Mean score across both lungs = Sum of 2 lung scores / 2

Treadmill ergometry

Treadmill ergometry is performed at screening and approximately every 6 months since the first infusion; for further details of timing, please refer to the protocol. Completion of the assessment may depend upon patient cooperation.

Treadmill ergometry assessments will include maximum workload (watts), percent predicted maximum workload (%), working time (min), maximum heart rate (breaths/min), percent predicted maximum heart rate (%), maximum O_2 saturation (%), maximum respiratory rate (breaths/min), maximum ventilation (L/min), maximum tidal volume (mL), percent predicted maximum O_2 uptake (mL/min), maximum O_2 uptake (%), maximum CO_2 output (mL/min), and maximum respiratory exchange ratio.

Fasting lipids (including lipoproteins of interest)

Fasting lipids evaluations are done at screening and approximately every 3 months since the first infusion. The parameters include: total cholesterol, LDL, HDL, VLDL, triglycerides, apolipoprotein B, apolipoprotein A1, and lipoprotein [a]. A derived parameter of non-HDL will be evaluated as follows: total cholesterol minus HDL (10).

Hematology

Platelet counts and hemoglobin are considered efficacy parameters. The hematology and hemogram sample collection schedule has been described under platelet count which is a secondary endpoint (Section 2.1.4.2). The rest of the hematology parameters, such as hematocrit, white blood cell counts (WBC), and red blood cell count are part of safety endpoints. Hematology samples evaluate all hematology parameters, while hemogram samples evaluate only platelet, hemoglobin and total and deferential WBC.

Bone disease assessments

Bone marrow burden

An MRI of lumbar spine and bilateral femur is performed at screening and at yearly visits. Three parameters are evaluated for each region: T1 weighted score, T2 weighted score, extent of involvement - femur only, infiltration pattern - spine only. The scale of these results is as follows:

Spine T1 score: 0=slightly hyperintense, 1=isotense, 2=slightly hypointense, 3=hypointense
Spine T2 score: 0=isointense, 1=slightly hyperintense or slightly hypointense, 2= hyperintense or hypointense
Infiltration pattern - infiltration: 0=None, 1=Patchy, 2=Diffuse
Infiltration pattern - fat in basivertebral vein region: 0=Preservation of fat, 1=Absence of fat

Femur T1 score: 0=Slightly hyperintense or isointense, 1=Slightly hypointense, 2=Hypointense
Femur T2 score: 0=Isointense, 1=Slightly hypointense or slightly hyperintense, 2=Hypointense or hyperintense, 3=Mixed type
Extent of Involvement: 0=No site involvement, 1=Diaphysis, 2=Proximal epiphysis / apophysis, 3=Distal epiphysis

Total bone marrow burden (BMB) score for each region is derived by adding up the individual scores. Total BMB score is derived by adding the total BMB score of two regions, and ranges from 0 to 16 [0 to 4: mild; 5 to 8: moderate; 9 to 16: marked to severe].

Dual-energy X-ray absorption parameters

Dual-energy X-ray absorption (DXA) of lumbar spine and bilateral femur is performed at screening and at yearly visits. Three parameters are evaluated for each region, and for left and right side separately: Bone mineral density (BMD), T-score and Z-score. The scale of these results is as follows:

BMD: a numerical value in the unit of g/cm^2
T-score: [score >= -1] = Normal, [score >-2.5 to <-1] = Osteopenia,
[score <= -2.5] = Osteoporosis
Z-score: [score >-2] = Normal, [score <= -2] = Below normal

A total score is obtained by averaging the two side scores within each region.

Bone biomarkers

Bone-specific Alkaline phosphatase and C-telopeptide levels are evaluated at screening and approximately every 3 months since the first infusion.

Efficacy biomarkers

Serum chitotriosidase, ligand 18 (CCL-18), and ACE are evaluated at screening and approximately every 3 months after the first infusion of the study drug.

Chitotriosidase values will be normalized prior to study analyses based on the patient's chitotriosidase genotype:

Chitotriosidase genotype	Normalization of chitotriosidase activity
Normal	Reported value to be used
Heterozygous	Value multiplied by 2
Homozygous mutation	Value set to “no value”

Physician’s global assessment

The physician will assess the disease severity of the patient at screening, as mild, moderate, severe. At approximately every six months exposure of the study treatment, the disease severity

will be assessed. At these clinic visits, the investigator will also evaluate the patient's current clinical status compared to screening (baseline) by marking 1 of 7 categories: "marked improvement of daily activities", "moderate improvement of daily activities", "mild improvement of daily activities", "no change", "mild worsening of daily activities", "moderate worsening of daily activities" or "marked worsening of daily activities".

NMR of HDL

Evaluation of HDL by NMR will be performed at Screening, then every six months since first infusion for 2 years, yearly thereafter.

Liver function via echo-Doppler ultrasound

Detected signs of portal vein hyperpressure is considered an efficacy measure and will be assessed through liver function via echo-Doppler ultrasound at Screening, Week 52 and Week 104.

2.1.5 Safety endpoints

The safety analysis will be based on the reported adverse events and other safety information, such as clinical laboratory data, vital signs, ECG, etc.

Observation period

The observation period will be divided into epochs:

- The **screening** epoch (or pretreatment epoch) is defined as the time from the signed informed consent date up to the start of the first infusion of investigational medicinal product (IMP) in the study (excluded).
- The **treatment** epoch is defined as the time from the first infusion of IMP (included) to the end of the study.
  - The treatment epoch for PAP is defined as the time from the first infusion of IMP (included) to:
    1. the time just prior to first infusion in ETP if the patient has an infusion in ETP;
    2. the last date the data are available if the patient has no infusion in ETP.
  - The treatment epoch for ETP is defined as the time from the first infusion in ETP to the end of the study.

Based on these epochs the safety data collected in this study will be classified as pre-treatment and treatment emergent.

The on-study observation period is defined as the time within treatment epoch.

2.1.5.1 Adverse events variables

Adverse event observation period

- Pretreatment adverse events are adverse events (AEs) that started before the first administration of IMP.
- Treatment-emergent adverse events are adverse events that developed or worsened or became serious during the treatment epoch. These will be defined separately for PAP and ETP.

If due to incomplete information on AE start date/time, the period designation cannot be made unambiguously (for imputation of missing or incomplete dates, please refer to Section 2.6.2), the AE is assumed to be treatment-emergent (and treatment-emergent in PAP, if so required for assumption).

All adverse events (including serious adverse events and adverse events of special interest) will be coded to a lower-level term (LLT), preferred term (PT), high-level term (HLT), high-level group term (HLGT), and associated primary system organ class (SOC) using Medical Dictionary for Regulatory Activities (MedDRA).

Adverse events are collected from the time of signed informed consent until the end of the study.

Adverse events of special interest include:

Infusion associated reactions

Protocol-defined infusion associated reactions (IAR): Any AE that occurs during the infusion or within up to 24 h after the start of the infusion and is considered related or possibly related to study treatment as judged by the investigator or the sponsor is defined as a protocol-defined IAR. An event occurring >=24 h after the start of an infusion may be judged an IAR at the discretion of the investigator or sponsor. Sponsor identification will reconcile with investigator identification of IAR before the database lock as much as possible. It is, however, possible that the sponsor identifies an AE as an IAR which the investigator did not agree to be an IAR; in that case, the sponsor decision is deemed final.

For this study, IARs may present as hypersensitivity reactions, acute phase reactions (APR), and cytokine release syndrome (CRS).

IARs, specifically APRs and some hypersensitive reactions may be associated with changes in specific clinical laboratory parameters (see Section 10.4.1.3 of the study protocol), and will be reconciled with investigator identification of those events as marked in the AE log as appropriate, facilitated by periodic clinical/safety review of AEs.

Algorithm-defined IAR: All AEs that start between the start of infusion and the end of infusion plus 24 hours are defined as algorithm-defined IARs, irrespective of the perceived relation with study treatment. Any TEAEs with missing time but which happened on the same day or the day after the infusion will be considered as algorithm-defined IARs.

Pregnancy: Pregnancy of a female patient enrolled in the study as well as pregnancy occurring in female partner of a male patient enrolled in this study is captured as an adverse event of special interest. The eCRF provides urine pregnancy test results as positive or negative. Labs provide serum pregnancy test results. Pregnancy is entered in the database as an AE, and will be counted from the AE page.

Symptomatic overdose: This is captured as an adverse event in this study, and will be reconciled with the clinical review of dosing data. For details, please see Section 10.4.1.3 in the protocol. This will be counted from the AE page.

Laboratory values: The following laboratory values and symptoms will be considered AESIs:

- Any increase in AST, ALT, total bilirubin, or AP >3x baseline (prior to olipudase alfa therapy) and > ULN
- Any increase in total bilirubin or AP >1.5x baseline in the presence of AST or ALT >2x ULN
- Any increase in ALT or AST >3x the ULN combined with an increase in ALT or AST >2x baseline (prior to olipudase alfa therapy) with symptoms of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash, or eosinophilia (> ULN)

Adverse events of other interests include:

Infections: TEAEs under the system organ class of Infections and Infestations

Bleeding: TEAEs under the standard MedDRA queries (SMQs) of Hemorrhages (identified using SMQ code 20000038 broad and narrow scope).

2.1.5.2 Deaths

The deaths observation period are per the observation periods defined above (Section 2.1.5).

- Death on-treatment: deaths occurring during the treatment epoch
- Death on-study: deaths occurring between informed consent date in this study and the end of the study

Death is reported as a fatal outcome in AE log page in eCRF.

2.1.5.3 Laboratory safety variables

Clinical laboratory data consists of blood analysis, including hematology, clinical chemistry, and urinalysis. Clinical laboratory values will be converted into standard international units. The US conventional units as well as international units will be used in all listings and tables for continuous variables.

Blood samples for clinical laboratories are scheduled to be taken at screening, all visits (every two weeks) during the dose escalation period (Week 0-16 – both in PAP and ETP), every

3 months thereafter, and in early termination visit if applicable, unless otherwise specified. The following clinical laboratory tests will be assessed by the site as specified in Section 9.2.2 of the study protocol. For more detailed schedule of assessment and window of time points, please refer to the protocol Section 1.2. The laboratory parameters will be classified as follows:

- Clinical Chemistry (before infusion): sodium, potassium, calcium, chloride, blood urea nitrogen, creatinine, lactate dehydrogenase, total protein, albumin, glucose, cholesterol, phosphorus, and creatine kinase
- Liver Function Tests (before infusion, 24 h and 48 h after the end of infusion): alanine aminotransferase (ALT), aspartate aminotransferase (AST), alkaline phosphatase, gamma glutamyl-transferase (GGT), total bilirubin, and direct bilirubin. The pre-infusion assessments on ALT, AST, total bilirubin and direct bilirubin are considered as additional efficacy endpoints in this study, while, the 24 h and 48 h post-infusion assessments on these tests are considered as safety endpoints.
- Hematology (before infusion, 24 hours after end of infusion; in yearly visits additional sample of hemogram before infusion):
  - Hematology sample: complete blood count with differential, including hematocrit, hemoglobin, red blood cell, platelet, white blood cell, neutrophil, lymphocyte, monocyte, eosinophil, and basophil counts (as well as neutrophil, lymphocyte, monocyte, eosinophil, and basophil as percentages of the WBC)
  - Hemogram sample: platelet, hemoglobin, WBC
- Coagulation Studies (before infusion, 24h and 48h after the end of infusion): prothrombin time (Protime; PT) including PT-INR, partial thromboplastin time (aPTT), and D-dimer
- Urinalysis (before infusion): dipstick for glucose, protein, hemoglobin, leukocytes, ketones, and bilirubin
- Additional Tests (screening only): hepatitis B surface antigen, hepatitis C antibody, human immunodeficiency virus antibody,
- Additional Tests (every 4 weeks): serum and urine β-human chorionic gonadotropin (β-HCG) pregnancy test (women of childbearing potential only)

Note: If a urine pregnancy test is positive, dosing of study drug will be withheld until results are confirmed with a serum β-HCG test.

- Safety biomarkers (before infusion, 24h and 48h after the end of infusion; no testing at screening): hsCRP, iron, ferritin, cardiac troponin I, IL-6, IL-8, ceramide in plasma and calcitonin.
- Immune response assessments (every 4 weeks during dose-escalation phase - both PAP and ETP, then at quarterly visits): anti-olipudase alfa IgG antibodies.

If an IAR is suggestive of a hypersensitivity reaction, additional laboratory parameters will be evaluated as soon as possible: tryptase activity, complementation activation, IgG and IgE. A skin testing may also be performed.

If an IAR is suspected to be a CRS, additional laboratory parameters will be evaluated as soon as possible: additional multiplex assay, calcitonin (see Section 2.1.5.6 inflammatory and vascular biomarkers).

2.1.5.4 Vital signs variables

Vital signs include: systolic and diastolic blood pressure (mmHg), heart rate (beats/minute), respiratory rate (breaths/minute), and temperature (°F or °C). At dose-escalation visits and at quarterly visits, vital signs will be measured before infusion, half-way through infusion, at the end of infusion and at specific intervals after the end of infusion until the patient is discharged from the hospital; for details of time points and associated window of time, please refer to the protocol Section 1.2.

2.1.5.5 Electrocardiogram variables

ECGs are reviewed by an expert at each site then centrally read by an independent expert for measurement of parameters and overall interpretation. Pre-infusion ECG on the day of first infusion will be done in triplicate; all other ECGs - screening, Day 1 (post-infusion only) and at quarterly visits - are done as singlet before the infusion start, 4 h, 12 h and 24 h after the end of infusion.

ECG parameters include, but not limited to, heart rate, intervals of PR, QRS, QT and QTc (according to Bazett/Fridericia), QRS axis.

2.1.5.6 Other safety endpoints

Doppler echocardiogram

A standard 2-dimensional and M-mode Doppler echocardiogram will be conducted during the Screening period, at Weeks 14, 52, 68 and 104, and at early termination if applicable; for details please refer to Sections 1.2 and 9.2.5 of the study protocol. Examination will include ventricular cavity size, valve characterization, ejection fraction, ventricular wall thickness, regional wall motion, diastolic function, pericardium characterization, pulmonary blood pressure and blood flow.

Physical examination

A complete physical examination will be performed during the screening period, before infusion at Day 1 and at every quarterly visit, and will include a neurological examination (mental status, cranial nerves, muscle strength, sensation, deep-tendon reflexes, and coordination) and an assessment of the patient's general appearance; skin; head, eyes, ears, nose, and throat; and examinations of lymph nodes, heart, lungs, abdomen, and extremities/joints.

An abbreviated physical examination of the patient's general appearance only will be performed before and after infusion at every infusion visit.

Patients will be assessed for weight at screening, every infusion visit during dose escalation (both PAP and ETP) and at quarterly visits. Height will be assessed at Screening and every yearly visit. Body mass index (BMI) will be calculated at the every visit that weight is measured, using the latest height and current weight available.

Liver biopsy

Liver biopsy is done at screening, Week 52 and Week 104. Liver biopsies are evaluated by a single blinded Sanofi Genzyme pathologist for liver fibrosis grading and for sphingomyelin accumulation. Fibrosis staging will be performed using the Laennec scoring system (11). Most of the pathology reports may not be included in the database, but will be received by the clinical personnel. Sphingomyelin accumulation from liver tissue will be provided with the liver biopsy results, and is discussed under pharmacodynamic endpoints (Section 2.1.7).

Liver function via echo-Doppler ultrasound

The liver ultrasound imaging with Doppler is performed at Screening, Week 52 and Week 104. The assessments include, but not limited to:

- Portal vein: diameter, direction of flow, velocity
- Presence of: liver surface irregularities, nodules, ascites, dysmorphia
- Presence and location of portosystemic venous collaterals
- Congestion index (CI)
- Hepatic artery: diameter, resistance, pulsatility index (PI)
- Hepatic vein waveform and profile
- Arterio-portal velocity ratio
- Detected signs of portal vein hyperpressure (this assessment is considered an efficacy assessment and will be analyzed under additional efficacy part)

Inflammatory and vascular biomarker

The inflammatory and vascular biomarkers are measured as part of Meso Scale Discovery panel within 24 hours before the start of infusion, then 24 h and 48 h (+/- 3 hours) after end of infusion at every visit during the dose escalation period (both in PAP and ETP) and at quarterly visits. Further, if a patient is suspected to have a cytokine release syndrome, a multiplex assay panel will be evaluated, which includes the following:

A) Inflammatory biomarkers:

- Chemokine panel: Thymus activation-regulated chemokine (TARC), Eotaxin-3, Eotaxin-1, Macrophage inflammatory protein 1 alpha (MIP-alpha), Macrophage inflammatory protein 1 beta (MIP-beta), Macrophage-Derived chemokine (MDC), Interferon gamma-induced protein 10 (IP-10), Monocyte chemotactic protein-1 (MCP-1), Monocyte chemotactic protein 4 (MCP-4), Interleukin 8 (IL-8)

- Cytokine panel: Granulocyte macrophage colony stm factor (GM-CSF), Tumor necrosis factor beta (TNF-beta), Vascular endothelial growth factor (VEGF), Interleukin 1 alpha (IL-1 alpha), Interleukin 12/23 subunit P40 ((IL12/IL23p40), Interleukin 15 (IL-15), Interleukin 16 (IL-16), Interleukin 17 alpha (IL-17A), Interleukin 5 (IL-5), Interleukin 7 (IL-7)
- Pro-inflammatory panel: Tumor necrosis factor alpha (TNF-alpha), Interleukin 1 beta (IL-1 beta), Interleukin 10 (IL-10), Interleukin 12 subunit P70 (IL12p70), Interleukin 13 (IL-13), Interleukin 2 (IL-2), Interleukin 4 (IL-4), Interleukin 6 (IL-6), Interleukin 8 (IL-8), Interferon gamma (IFN-gamma). [Please note IL-6 and IL-8 are considered safety biomarkers in this study.]

B) Vascular biomarkers include:

- Vascular injury panel: C Reactive Protein (CRP), Serum amyloid A protein (SAA), Intercellular adhesion molecule 1 (ICAM-1), Vascular cell adhesion molecule 1 (VCAM-1)
- Angiogenesis panel: Basic fibroblast growth factor (bFGF), Phosphatidylinositol-glycan biosynth. F (PIGF), Soluble fms-like tyrosine kinase-1 (FIT-1), TEK (TIE2) recombinant human protein, Vascular endothelial growth factor (VEGF), Vascular endothelial growth factor C (VEGF-C), Vascular endothelial growth factor D (VEGF-D)

2.1.6 Pharmacokinetic variables

Pharmacokinetic endpoints will include evaluation in plasma of the following parameters: C_{max}, AUC0-τ, $AUC_{0\text{-last}}$, $t_{1/2}$, CL, and V_{β}. Definition of each parameter is provided in Section 9.3.2.4 of the study protocol. The process of sample collection has been outlined in Section 9.3.2.1.

Concentrations of olipudase alfa in plasma will be determined using a validated enzyme-linked immunosorbent assay method.

2.1.7 Pharmacodynamic/genomics endpoints

The pharmacodynamic endpoints include:

- Change from baseline to Week 52 in sphingomyelin in liver tissue
- Percent change from baseline to Week 52 in:
  - Sphingomyelin in plasma
  - Lyso-sphingomyelin in plasma
- Percent change from baseline to Week 52 in Sphingosine-1 phosphate in dried blood spot (DBS)

The plasma and/or DBS samples are collected at every visit during dose-escalation (both PAP and ETP) and at quarterly visits thereafter.

Sphingomyelin accumulation is evaluated from liver tissue collected during liver biopsy at screening, Week 52 and Week 104. Sphingomyelin accumulation will be quantified by computer morphometry of high-resolution light microscopy images (11). Up to 10 blocks will be produced from each patient's liver biopsy at each time point with values representing % of tissue staining for sphingomyelin for each block (stored as source data), and a mean +/- SD for each biopsy will be provided by the blinded pathologist.

2.1.8 Quality-of-life endpoints (patient-reported outcome)

Quality of life questionnaires include BFI, BPI-SF, Niemann-Pick B Health Assessment Questionnaire (NPB-HAQ) v2.0, Short Form-36 Health Survey (SF-36); FACIT-Dyspnea SF; for details please refer to Sections 9.1.4.13 and 9.1.4.14 in the protocol.

- The BFI is a validated, self-administered questionnaire that was originally developed to assess fatigue severity in cancer patients. The BFI has 9 items, using numeric rating scale from 0 to 10. This questionnaire data is collected at screening and every 6 months since the first infusion.
  - Three items assess the severity of the fatigue now and at its worst and usual during past 24 hours. If 2 out of 3 items are available, the item scores are averaged to provide fatigue severity scale score; if less than 2 item scores are available, the scale score is left missing.
  - Six items assess the fatigue interference with activities of life during the preceding 24 hours. If 4 out of 6 items are available, the item scores are averaged to provide fatigue interference scale; if less than 4 item scores are available, the scale score is left miss
  - Question 3 in BFI is part of eDiary and is collected over 7 days at baseline and at quarterly visits; this is part of secondary endpoints and has been discussed in Section 2.1.4.2.

- The BPI-SF is a validated, self-administered questionnaire designed to measure a patient's perceived level of pain. The BPI-SF consists of 15 items that use a numeric rating scale from 0-10, 0 being no pain and 10 being the worst, to assess pain severity and pain interference in the past 24 hours and the past week. This questionnaire data is collected at screening and every 6 months since the first infusion.
  - There are 4 questions that are used for the scale of pain severity - Q3, 4, 5, 6 in Ecrf - which ask for "worst" and "least" pain within past 24 hours, an average pain score and the pain severity "right now" respectively. If 2 or more responses are available (12, 13), the individual question scores are averaged to provide the scale score for pain severity; if only one response to the 4 questions is available (or no response is available at all), the scale score is left missing.
  - Item #9 in eCRF has seven questions asking about how the pain interfered with general activity, mood, sleep, etc. If 4 or more responses are available, the individual question scores are averaged to provide scale score for pain interference; if less than 4 responses are available, the scale score is left missing.

  - Question 3 in BPI-SF is part of eDiary and is collected over 7 days at baseline and at quarterly visits; this is part of secondary endpoints and has been discussed in Section 2.1.4.2.

- The NPB-HAQ v2.0 is a disease-specific questionnaire that covers various aspects of fatigue, pain, respiratory, and quality of life as well as questions specific to ASMD symptoms and physical activity. The 3 domains of this questionnaire include: resource utilization, symptoms and physical activity, with several questions in each, with qualitative or quantitative response. This questionnaire data is collected at screening and every 3 months following the first infusion.
- The SF-36 v2.0 is a 36-item, validated, multidimensional, generic health-related quality of life measure that has been validated for adults in numerous healthy and ill populations internationally (14, 15). The SF-36 consists of 8 scales, including physical functioning, role physical, bodily pain, mental health, role emotional, social functioning, vitality, and general health and includes two summary measures of physical health and mental health derived from scale aggregates. These data including 8 individual scales and 2 summary measures are scored externally by QualityMetrics who had the proprietary rights to do so. This questionnaire data is collected at screening and every 6 months since the first infusion.
- FACIT-Dyspnea Short form consists of 10 symptom-specific questions, and each of these questions has a related question asking about functional limitations (8). In Part 1, the patient rates how short of breath the patients felt in 7 days in doing different activities such as dressing, walking, household chores; this is already part of secondary endpoints and the scoring has been described in Section 2.1.4.2. Part 2 asks to rate how limiting the activities have been due to dyspnea, and is scored similar to part 1 (conversion scale provided in Section 7.1); this functional limitation section is part of additional efficacy endpoints. This questionnaire data is collected at screening and every 6 months since the first infusion.
- Patient Global Impression of Symptom Severity (PGIS) is a patient-rated 5-question instrument on a 5-point scale, higher score indicating worse. PGIS asks patients to rate abdominal problems, bodily pain, fatigue, and shortness of breath “over the past week”. This questionnaire data is collected at screening and every 3 months since the first infusion.
- Patient Global Impression of Change (PGIC) is a patient-rated 4-question instrument on a 7-point scale, higher score indicating worse, to rate the change in the symptoms since the start of study infusion. This questionnaire data is collected at screening and every 3 months since the first infusion.

2.1.9 Health economic endpoints

- Health-related productivity questionnaire is a validated, self-administered questionnaire that was developed to measure how treatment of disease impacts an individual’s ability to participate in the workforce and complete daily household duties. This questionnaire data is collected at screening and every 3 months since the first infusion. A part of this questionnaire is about the workforce participation before the start of this study; this is collected only at screening.

- EQ-5D is a standardized measure of health status developed by the EuroQoL Group to provide a simple, generic measure of health for clinical and economic appraisal (16). The EQ-5D-5L consists of 5 dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has 5 levels: no problems, slight problems, moderate problems, severe problems, and extreme problems. The respondent is asked to indicate his/her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the 5 dimensions. This decision results in a 1-digit number expressing the level selected for that dimension. The digits for 5 dimensions can be combined in a 5-digit number describing the respondent's health state. In addition, there is visual analogue scale 0-100, 0 indicating the worst and 100 indicating the best that is scored for the patient's health perception on the day of the visit; eCRF collects a numeric value between 0 and 100. This questionnaire data is collected at screening and every 6 months since the first infusion.

For further details, please refer to Section 9.1.4.13 in the protocol.

2.2 DISPOSITION OF PATIENTS

This section describes patient disposition for both patient study status and the patient analysis populations.

Screened patients are defined as patients who have signed the informed consent form (ICF). Eligible patients are defined as screened patients who have met all eligibility criteria.

Randomized patients consist of all patients with a signed ICF who have had a treatment kit number allocated and recorded in the IXRS database, regardless of whether the treatment kit was used.

For patient study status, the total number of patients in each of the following categories will be presented in the clinical study report by treatment group, using a summary table, and percentages will be provided using the number of randomized patients as the denominator (except for screened, eligible and non-randomized patients where a percentage will not be provided). For PAP CSR, the summary of the following categories will be provided up to the data cutoff date for the database lock of PAP. For final CSR, the summary of the following categories will also be provided:

- Screened patients
- Eligible patients
- Nonrandomized but treated patients
- Randomized patients
- Randomized but not treated patients
- Randomized and treated patients
- Patients who received rescue therapy during PAP

- Patients who did not complete the study treatment period as per protocol
- Patients who discontinued study treatment by main reason for permanent treatment discontinuation
- Status at last study contact
  - Completed PAP and enrolled in ETP
    - Enrolled in ETP but did not complete ETP
    - Completed ETP
  - Completed PAP but did not enroll in ETP
  - Did not complete PAP

A patient is considered lost to follow-up if he/she is not assessed at the last protocol planned visit and if the time from the last successful contact to the last protocol planned visit is greater than 30 days.

2.2.1 Randomization and drug dispensing irregularities

Randomization and drug-dispensing irregularities occur whenever:

1. A randomization is not in accordance with the protocol-defined randomization method, such as a) an ineligible patient is randomized, b) a patient is randomized twice.

   OR

2. A patient is given a study drug not allocated by the protocol-defined randomization, such as a) a patient at any time in the study is given a different treatment than as randomized (eg, patients randomized to placebo treatment group received olipudase alfa), or b) a nonrandomized patient is treated with study drug.

   OR

3. A patient receives an incorrect dose of olipudase alfa (not matching what the investigator requested, or not matching what IXRS instructed pharmacist to prepare, or infusion prepared for 1 patient given to another patient).

All randomization and drug-dispensing irregularities will be documented in the clinical study report. Whether any of these constitute a major or critical protocol deviation, and whether any of these adversely impact efficacy, is deferred to the decision by the sponsor's clinical team before the database lock in PAP. Nonrandomized, treated patients, if any, will be described separately.

Randomization and drug-dispensing irregularities to be prospectively identified include but are not limited to:

Randomization and drug allocation irregularities
Study drug given without IRT transaction
Erroneous study drug administration
Randomization by error
Patient randomized twice

2.2.2 Protocol deviations

Protocol deviations are qualitative and quantitative in nature. Qualitative deviations are evaluated and reported in a list by study monitoring team and reviewed by the study team. Quantitative deviations are programmed in SAS from the data, and reviewed by the study team. In both cases, the study team determines whether a deviation is minor, major or critical before the unblinding in PAP (and any that occurs in ETP at the end of study before the database lock). A summary table and a data listing of major or critical deviations will be provided, including, but not limited to, prohibited medication use as well as non-compliance.

2.3 ANALYSIS POPULATIONS

Patients treated without being randomized will not be considered randomized and will not be included in any efficacy analysis. Patients who are dispensed study drug without calling the IRT or before calling the IRT are considered nonrandomized patients. They are excluded from any population for analysis, including safety. However, if these patients experienced any significant safety event, they will be documented separately in the clinical study report.

For any patient screened more than once, only the data associated with the second screening will be used in any analysis population. The safety experience associated with any later randomization will be assessed separately.

The **randomized population** includes any patient who had been allocated to a randomized treatment arm by IRT regardless of whether study drug was administered.

2.3.1 Efficacy populations

2.3.1.1 Modified intent-to-treat population

The **modified intent-to-treat (mITT) population** is a subset of the randomized population and will include randomized patients who received at least 1 infusion (partial or total). [Note that the design of the study requires patients to be hospitalized for infusion, and the randomization happens when the patient is hospitalized; therefore, it is unlikely that mITT population will differ from the traditional ITT population which constitutes of all randomized patients irrespective of whether one receives a dose of study drug.] Patients will be analyzed in the treatment group to which they are randomized.

2.3.1.2 Per-protocol population

The **per-protocol population** is a subset of the mITT population that has no critical or major protocol deviations that are expected to interfere (the list to be finalized before the database lock and unblinding in PAP) with assessments of the primary efficacy endpoints.

2.3.1.3 Other efficacy population

The **mITT-C population** is the subset of mITT population which excludes patients who had exposure to [redacted] of olipudase alfa in the active treatment arm.

2.3.2 Safety population

The **safety population** is defined as a subset of the randomized population who received at least 1 infusion (partial or total), analyzed according to the treatment actually received during the PAP, irrespective of the treatment to which the patient has been randomized. For example, if a placebo patient was rescued to receive olipudase alfa in PAP, this patient will be counted as olipudase alfa treated patient. The safety population will not be changed after patients enter the ETP. Olipudase alfa infusions require special vials to be mixed with the vehicle while placebo infusions do not require these vials, and these vial numbers are called off by IXRS, it is unlikely that there will be any error in treatment received.

The **rescue population** is defined as patients who have used rescue strategy in PAP (ie, unblinded for treatment and continued on olipudase alfa irrespective of the randomized treatment assignment). Data after rescue in PAP will be provided separately for the rescue population.

In addition, nonrandomized but treated patients will not be part of the safety population; however, their safety data will be presented separately.

2.3.3 Other analysis populations

The **pharmacokinetic population** consists of mITT patients who have evaluable drug concentration data.

The **pharmacodynamic population** consists of mITT patients who have at least one evaluable pharmacodynamic marker data available post-baseline.

2.4 STATISTICAL METHODS

Continuous data will be summarized using the number of available data, mean, standard deviation (SD), median, minimum, and maximum for each treatment group. Categorical and ordinal data will be summarized using the number and percentage of patients. Summaries and summary plots are presented by randomized treatment group; by-patient plots and data listings will identify the randomized treatment group for each patient.

2.4.1 Demographics and baseline characteristics

Parameters described in Section 2.1.1 will be summarized by randomized treatment group and pooled treatment groups using descriptive statistics in mITT population.

Medical history will be summarized using body system and/or medical/surgical condition of specific interest, by randomized treatment group and pooled treatment groups in mITT population.

No specific description of the safety parameters will be provided within the baseline characteristics displays. The baseline values will be described along with each safety analysis.

No specific description of the efficacy parameters will be provided within the baseline characteristics displays. The baseline values will be described along with each efficacy analysis.

P-values on demographic and baseline characteristic data will not be provided.

2.4.2 Prior or concomitant medications

The prior and concomitant medications have been defined in Section 2.1.2.

Medications will be summarized by randomized treatment group using mITT population according to the WHO-DD, considering the first digit of the anatomic and therapeutic category (ATC) class (anatomic category) and the first 3 digits of the ATC class (therapeutic category). All ATC codes corresponding to a medication will be summarized, and patients will be counted once in each ATC category (anatomic or therapeutic) linked to the medication. Therefore patients may be counted more than once for the same medication, but for different ATC category. Prior and concomitant mediations will be summarized separately.

The table for prior medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the overall incidence across treatment groups. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

The tables for concomitant medications will be sorted by decreasing frequency of ATC followed by all other therapeutic classes based on the incidence in olipudase alfa treatment group. In case of equal frequency regarding ATCs (anatomic or therapeutic categories), alphabetical order will be used.

Prohibited medications during the study, if used, will be provided as part of a data listing for major and critical protocol deviations.

2.4.3 Smoking and alcohol drinking

Alcohol and tobacco usage data will be provided in listings.

2.4.4 Extent of investigational medicinal product exposure and compliance

The extent of IMP exposure and compliance will be assessed and summarized by randomized treatment group in mITT population (Section 2.3.1.1).

2.4.4.1 Extent of investigational medicinal product exposure

The extent of study drug exposure will be assessed by the duration of study drug exposure, and amount of dose received.

Study drug exposure for PAP is defined below:

For patients who did not initiate rescue therapy:

- Duration of study drug exposure for PAP = the first dose date in ETP (or 13 days after the last dose date in PAP if the patient is not continuing in ETP) – first dose date in PAP, regardless of unplanned intermittent discontinuations

For patients who had initiated rescue therapy during PAP:

- Duration of rescue therapy exposure in PAP = the first dose date in ETP (or the last dose date in PAP + 13 days if the patient is not continuing in ETP) – first rescue therapy dose date in PAP, regardless of unplanned intermittent discontinuations.
- Duration of study drug exposure for PAP before initiation of rescue therapy = the first rescue therapy dose date in PAP – first dose date in PAP, regardless of unplanned intermittent discontinuations.

Study drug exposure for PAP (before initiation of rescue therapy if any) will be summarized descriptively. In addition, duration of treatment exposure in PAP may be summarized categorically, eg, <14 weeks, <26 weeks, etc. Duration of olipudase alfa exposure after rescue initiation will also be summarized.

Cumulative study drug exposure of olipudase alfa is defined as follows:

- Duration of cumulative study drug exposure = disposition date of the study (or the cutoff date for database lock if the patients are still ongoing) – first dose date of olipudase alfa infusion in the study + 1, regardless of unplanned intermittent discontinuations.

Duration of cumulative study drug exposure will be summarized descriptively as well as categorically.

Dose information for PAP will be assessed by the following variables for patients who received any dose of olipudase alfa:

The cumulative dose for PAP is the sum of total amount of olipudase alfa in mg received in all infusions in PAP. Total amount of olipudase alfa in each infusion is obtained from the patient dose level in mg/kg multiplied by weight, adjusted for partial infusion. When the question on the eCRF “Was the total volume infused?” indicates “No” (ie, infusion was partial), eCRF does not collect the volume actually infused (to maintain blinding, the color of the liquid is covered from

view); this needs to be imputed from the infusion rate and infusion time as follows: if the infusion rate is xx ml/hour and the duration of that infusion was zz hour, then the volume infused is imputed as (xx times zz) ml. The sum of such individual infusion volume over different infusions within a visit will constitute the partial volume infused. If the resulting volume is greater than total volume prepared, it will be imputed as equal to total volume. In case of partial infusion, the total amount of olipudase alfa in the infusion is considered as proportional to the volume infused.

Total amount of olipudase alfa in each infusion is obtained from the patient dose level in mg/kg multiplied by weight, adjusted for partial infusion:

= Dose level (mg/kg) * weight (kg) if no partial infusion

= Dose level (mg/kg) * weight (kg) * partial volume infused (mL) / Total volume prepared (ml) if partial infusion

Cumulative dose in PAP for each patient will be summarized. In addition, the number and percentage of patients for whom the target dose of 3 mg/kg was not achieved, and separately for those who reach target dose but have to reduce dose later for safety reasons will be provided.

Any incorrect dosing will be part of protocol deviations.

The cumulative dose of olipudase alfa over the entire study will be summarized.

2.4.4.2 Compliance

Compliance to the treatment regimen will be monitored in terms of the percentage of scheduled infusions the patient receives during the treatment periods, irrespective of whether the infusion was total or partial. This is calculated as:

100 x (number of infusions received) / (number of infusions intended to be received)

For PAP, the number of infusions intended to be received is 27 over 52 weeks for a patient who complete PAP.

Noncompliance in PAP is defined as missing 2 or more consecutive infusions in PAP (excluding for cause misses, such as when the missed doses are required by the protocol due to a DLT) or 4 or more infusions in total in the 52-week PAP treatment period.

Noncompliance in PAP+ETP is defined as missing 4 or more infusions in every 52-week period following PAP treatment period until the end of study.

Treatment compliance for overall PAP will be summarized descriptively as quantitative variable using mITT population and randomized treatment group.

The percentage of patients whose compliance is ≤85% (missing 4 or more infusions out of scheduled 27 infusions in PAP) will be provided. In addition, number and percentage of patients with noncompliance, and by reasons (missed 2 or more consecutive infusions not for cause (“not for cause” means there is no justifiable reasons for a missed dose, eg., DLT), or missed 4 or more infusions in total in the 52-week PAP) will be provided.

Similar summaries will be provided for PAP+ETP.

A case of overdose will be reported as an AE and will be summarized with AE. An overdose is defined in Section 10.4.1.3 of the protocol.

2.4.5 Analyses of efficacy endpoints

Efficacy data will be analyzed by randomized treatment group in mITT population at scheduled time point. The statistical analyses will compare olipudase alfa versus placebo within PAP; the difference between two treatment groups is calculated as olipudase alfa minus placebo.

Primary analyses period is 52 weeks.

Efficacy analyses will use mITT population unless specified otherwise. This will include patients who used rescue therapy, but their data after rescue will NOT be included in the analysis. Sensitivity analyses including data beyond rescue therapy may be undertaken if appropriate.

Baseline values are generally the latest value before the first infusion of the study drug, except for platelets and hemoglobin which are part of hematology parameters; for hematology parameter baseline definition, please refer to Section 2.6.1.

2.4.5.1 Analysis of primary efficacy endpoint(s)

The percentage change in DLco (% predicted) from baseline to 52 weeks will be analyzed in the mITT population using a mixed model for repeated measures (MMRM) (17, 18). The MMRM will include baseline DLco, baseline age, treatment arm, study visit, and study visit by treatment arm interaction as covariates; have an unstructured variance-covariance matrix; and be fit using restricted maximum likelihood estimation. If this model fails to converge, the following variance-covariance structures will be tested in this order: Toeplitz (equal variances and a separate correlation for each level of separation between the time points), AR(1) (first-order autoregressive, equal variances and exponentially decreasing correlations), CS (compound symmetry, equal variances and equal pairwise correlations across fixed time points). The first (co)variance structure yielding convergence will be used as the primary analysis. Comparisons between treatment arms will be made using least-square mean contrasts at the 52 week visit with denominator degrees of freedom estimated using the Kenward-Roger approximation (19). When a variance-covariance structure other than unstructured is used, then the denominator degrees of freedom will be estimated using the between within method (DDFM=BW in SAS PROC Mixed).

Example SAS code for the MMRM is:

```
ods output Diffs = out1
        LSMeans = out2;
proc mixed data = DATAIN method = reml;
  class TRTPN WEEK USUBJID;
  model PCHG_DLCO = BL_DLCO BL_AGE TRTPN|WEEK / s ddfm=kr;
  repeated WEEK / type = un subject = USUBJID  r;
  lsmeans TRTPN TRTPN*WEEK / cl diff;
run;
```

where TRTPN = treatment group, WEEK = (Week 26, Week 52), USUBJID = unique patient ID, PCHG_DLCO is the percent change from baseline in DLco (%predicted), BL_DLCO is the baseline DLco (% predicted) value, BL_AGE is the baseline age,

type =

UN = unstructured covariance

TOEP = Toeplitz covariance

AR(1) = first-order autoregressive covariance

CS = compound symmetry covariance

The MMRM assumes data missing-at-random (MAR) and will include all DLco observations except for measurements made after the initiation of rescue therapy. In addition, for patients who do not have Week 52 value available but have a PFT value measured after Week 38 infusion (eg, patients discontinuing the study or initiating rescue therapy after Week 38 but before Week 52 may have the PFT measurements available from early discontinuation visit or right before rescue therapy starts), these values will be used as Week 52 value for analysis purposes; Week 52 DLco will not be considered missing in such situations. This idea is based on the clinical concept that PFT measurements are similar within a 12-week period.

The percentage change in spleen volume (MN) and the change in the splenomegaly-related score (for US, SRS is part of primary endpoints) will be analyzed using an analogous MMRM model.

The summary plots of the LS means from the MMRM of the percentage change in DLco (% predicted), the percentage change in spleen volume (MN) and the change in the splenomegaly-related score from baseline up to Week 52 with 95% CIs will be provided.

Sensitivity analyses for the primary efficacy endpoints to support robustness of results will include the following:

1. The MMRM described above will be run using the per-protocol population. This analysis will demonstrate whether the results vary depending on the population analyzed.
2. The MMRM described above will exclude observations made after the initiation of rescue therapy. As a sensitivity analysis, the observations collected after the initiation of rescue therapy will be included in the MMRM; the treatment arm for rescued patients will remain

as randomized treatment.
The MMRM planned for the primary efficacy analysis assumes MAR. To assess the robustness of the primary results to that assumption a pattern mixture model will be used (see Section 2.6.6 for details).

3. The MMRM planned for the primary efficacy analysis assumes multivariate normality. To assess the robustness of conclusions under this assumption, nonparametric testing method will be used: A Wilcoxon-Mann-Whitney (WMW) will be used to compare the primary efficacy endpoints; missing Week 52 data will be imputed using LOCF, excluding patients who initiate rescue therapy.

Additional supportive analysis for splenomegaly-related scores domains

The symptom domain score will be calculated based on the first three questions in the splenomegaly-related score questionnaire (abdominal discomfort, satiety and pain under the rib). The 7-day mean of the daily sum of scores of these three questions will be defined as the symptom domain score. The impact domain score will be calculated based on the last two questions (bending down, body image). The 7-day mean of the daily sum of scores of these two questions will be defined as the impact domain score. The 7-day mean of the daily scores of each question in the splenomegaly-related score questionnaire will be defined as the score for each question. The change from baseline to Week 52 in these two domain scores and each question in the splenomegaly-related score questionnaire will be analyzed using the MMRM, similarly to the primary analysis of SRS.

Additional supportive analysis for DLco and spleen volume (MN)

To evaluate the treatment effect on the absolute change on DLco (% predicted) and spleen volume (MN), the change from baseline on these two endpoints up to Week 52 will be analyzed using an analogous MMRM model.

2.4.5.2 Analyses of secondary efficacy endpoints

Liver volume

The MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for percentage change in liver volume from baseline to Week 52, using mITT population. Supportive summary in mITT-C population will be provided. Sensitivity analyses are not planned, but may be undertaken if necessary.

Platelet counts

The MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for percentage change in platelet count from baseline to Week 52, using mITT population. Data collected at the quarterly visits will be included in the MMRM model. Supportive summary in mITT-C population will be provided.

BFI scale - Item 3

The MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for BFI - Item 3 – for change from baseline to Week 52, using mITT population. Supportive summary in mITT-C population will be provided. Sensitivity analyses are not planned, but may be undertaken if necessary.

BPI scale - Item 3

The MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for BPI - Item 3 – for change from baseline to Week 52, using mITT population. Supportive summary in mITT-C population will be provided. Sensitivity analyses are not planned, but may be undertaken if necessary.

FACIT-Dyspnea symptom score

The MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for FACIT-Dyspnea symptom score, change from baseline to Week 52, using mITT population. Supportive summary in mITT-C population will be provided. Sensitivity analyses are not planned, but may be undertaken if necessary.

Splenomegaly-related score

This is part of primary endpoints in US; all analyses for this endpoint are described in Section 2.4.5.1.

The summary plots of the LS means from the MMRM for all the secondary efficacy endpoints in change or percentage change from baseline up to Week 52 with 95% CIs will be provided.

2.4.5.3 Multiplicity issues

Primary endpoints

The hypothesis testing of the primary endpoints will proceed as follows:

- Compare olipudase alfa to placebo on DLco and spleen volume. The overall 5% significance level for the two hypothesis tests will be maintained using the Hochberg method [MTP2 is maintained for 2-sided, 2-endpoint testing(20)]:
  - Higher of the two p-values will be compared to 0.05; if this p <0.05, both DLco and spleen volume are significant.
  - If higher p-value >=0.05, and lower p-value <0.025, then the endpoint associated with lower p value is significant but one associated with higher p value is not significant.

For Europe and ROW: from the sponsor perspective, the study will be declared positive if at least one of the primary endpoints is statistically significant by the above method.

For US only: If spleen volume is statistically significant based on the above testing, comparison of olipudase alfa versus placebo on SRS will be considered; if the two-sided p-value is <0.15, then the combination spleen endpoint is considered significant. Otherwise, the combination spleen endpoint is not significant. From the sponsor perspective, the study will be declared positive if either DLco or combination spleen endpoint is significant.

Secondary endpoints (sequential testing)

If statistical significance is reached on **both** the DLco and spleen volume using the Hochberg method (and the trend in splenomegaly-related score is established in case of USA), then hypothesis testing of the secondary efficacy endpoints will proceed using sequential testing at 5% level with the order as specified below. At any step when the endpoint is not significant at 5% level, the formal testing in subsequent steps will stop; the p-values for the subsequent endpoints in the sequence will be considered exploratory, and hence interpreted at the nominal level. This controls the overall type I error level at 5%, using the closed testing principle. If either of the primary endpoints is not significant, then hypothesis testing for the secondary endpoints will be considered exploratory, and p-values will be interpreted at the nominal level.

For Europe and ROW, the order of testing for secondary endpoints is as follows:

1. Percentage change in liver volume from baseline to Week 52
2. Percentage change in platelet count from baseline to Week 52
3. Change in BFI scale, Item 3 - from baseline to Week 52
4. Change in BPI scale, Item 3 - from baseline to Week 52
5. Change in FACIT-Dyspnea symptom score from baseline to Week 52
6. Change in SRS from baseline to Week 52

In the US, the order of testing for secondary endpoints is as follows:

1. Percentage change in liver volume from baseline to Week 52
2. Percentage change in platelet count from baseline to Week 52
3. Change in BFI scale, Item 3 - from baseline to Week 52
4. Change in BPI scale, Item 3 - from baseline to Week 52
5. Change in FACIT-Dyspnea symptom score from baseline to Week 52

2.4.5.4 Subgroup analyses for the primary endpoints in PAP

The consistency of the treatment effect in PAP will be evaluated as follows for percentage change from baseline to Week 52 of spleen volume (MN) and percentage change from baseline to Week 52 of % predicted DLco separately:

- Spleen volume by baseline spleen volume severity (severe vs not severe). Severity is defined as baseline spleen volume >15 multiple of normal (MN).
- % predicted DLco by baseline % predicted DLco severity (severe vs not severe). Severity is defined as baseline % predicted DLco <40%.

In addition, for the endpoints of percentage change from baseline to Week 52 of spleen volume (MN), SRS, and % predicted DLco, the consistency of treatment effect in PAP will be assessed by the following subgroups:

- Baseline ALT or AST abnormality (ALT or AST ≥1 ULN vs ALT and AST <1 ULN)
- Baseline total bilirubin abnormality (total bilirubin ≥1.5 ULN vs total bilirubin <1.5 ULN)

For the percentage change from baseline to Week 52 of spleen volume (MN), whether or not the presence of portal hypertension has an effect on olipudase alfa treatment will be assessed by the following subgroup:

- Presence vs absence of portal hypertension at baseline (detected signs of portal vein hyperpressure is “likely severe portal hypertension” or “likely portal hypertension” for presence of portal hypertension vs detected signs of portal vein hyperpressure is “unlikely portal hypertension” for absence of portal hypertension)

Descriptive summaries on the observed values, change from baseline and percent change from baseline to Week 52 of spleen volume (MN) for patients with severe portal hypertension at baseline will be provided as well if there are at least 3 patients identified with severe portal hypertension at baseline. If the number of patients with severe portal hypertension at baseline is low (<3), only individual patient data will be provided in listings.

Due to the small sample size in these subpopulations, treatment difference within the subgroup will be difficult to interpret. Therefore, the focus of these analyses will be on assessment of interaction between subgroups and the treatment.

For each subgroup, same MMRM models for the primary efficacy endpoints using data from the specific subgroup only will be conducted to provide estimate of treatment effect within each

subgroup. The summary plots of the LS means from the MMRM of the percentage change in DLco (% predicted) and spleen volume (MN) from baseline up to Week 52 with 95% CIs by subgroups will be provided.

To assess the interactions between subgroup and treatment, the same MMRM models for the primary efficacy endpoints described in Section 2.4.5.1 will be applied with the addition of subgroup, subgroup-by-visit interaction, subgroup-by-treatment interaction, subgroup-by-treatment-by-visit interaction. The p-value for overall subgroup-by-treatment interaction, subgroup-by-treatment-by-visit interaction, subgroup-by-visit interaction as well as the p-value for subgroup-by-treatment interaction at week 52 will be provided.

2.4.5.5 Responder analyses for the primary endpoints in PAP

For the responder analyses specified below, binary measure of responder (yes vs no) at Week 52 will be analyzed with logistic regression including treatment and baseline as covariates. Patients who had missing data at Week 52 will be counted as non-responder. Number and % of responder, as well as odds ratio, 95% CI and p-value from the logistic regression model will be provided.

Responder analyses for spleen volume

In addition to evaluating the percent change from baseline, a responder analysis will be performed to assist in the interpretation of a clinically meaningful result. In Gaucher disease, therapeutic goals for splenomegaly include reduction in spleen volume of 30-50% within year 1 of enzyme replacement therapy (21). Therefore a responder analysis will be performed, setting the threshold for response at <= -30% change from baseline in spleen volume at 52 weeks.

Responder analyses for % predicted DLco

International guidelines for clinically meaningful changes in DLco have been published for idiopathic pulmonary fibrosis (IPF) including 2017 Spanish IPF guidelines and the 2017 French IPF guidelines indicate a decrease of >15% DLCO in absolute values is associated with increased risk of mortality (22, 23). At the same time, the Connective Tissue Disease- associated interstitial lung disease (CTD-ILD) -OMERACT CTD-ILD working group has consensus guideline indicates that a relative 15% change constitutes a clinically meaningful change (24).

In addition, Pelligrino et al (25), set a clinically meaningful threshold at 10% and similar values in the range of approximately 10-14% has been reported reviewing datasets from larger studies (26, 27).

Therefore, a responder analysis will be performed using a threshold of 15% improvement in DLco from baseline. This means if a patient has a change from baseline on DLco (% predicted) >=15% at Week 52, then this patient will be a responder.

Responder analyses for SRS, BFI scale - Item 3, BPI scale - Item 3 and FACIT-Dyspnea symptom score

Thresholds for meaningful within-patient improvement on the SRS, BFI scale - Item 3, BPI scale - Item 3, and FACIT-Dyspnea symptom score were calculated using blinded data from

the DFI12712 study (methods described in PRO Psychometric Statistical Analysis Plan). The primary and secondary improvement thresholds for each measure are as follows:

- SRS: Primary threshold is change from baseline at Week 52 <= -12.5 points; Secondary threshold is change from baseline at Week 52 <= -18 points.
- BFI scale - Item 3: Primary threshold is change from baseline at Week 52 <= -1.7 points; Secondary threshold is change from baseline at Week 52 <= -2.5 points.
- BPI scale - Item 3: Primary threshold is change from baseline at Week 52 <= -2.5 points; Secondary threshold is change from baseline at Week 52 <= -4.0 points.
- FACIT-Dyspnea symptom score: Primary threshold is change from baseline at Week 52 <= -7.0 points; Secondary threshold is change from baseline at Week 52 <= -5.6 points.

Responder analyses will be conducted using these thresholds. For each measure, patients who meet or exceed the meaningful change threshold at Week 52 will be considered a responder and those who do not meet or exceed the threshold will be considered non-responders. For patients who have missing data at Week 52, they will be considered as non-responder. Separate responder analyses will be conducted using the primary and secondary thresholds for each measure.

Cumulative distribution functions (CDFs) will be generated by treatment group for SRS, BFI scale - Item 3, BPI scale - Item 3 and FACIT-Dyspnea symptom score. The CDFs will show the cumulative percent of patients on the y-axis and change from baseline to Week 52 scores on the x-axis. The CDFs will allow a visual inspection of the cumulative proportion of patients in each treatment group who achieved various levels of change on the PRO measure at Week 52.

2.4.5.6 Compliance of eDiary data during PAP

In addition, to assess the compliance for eDiary data including SRS, BFI scale - Item 3, and BPI scale - Item 3, the following summaries will be provided for each treatment group.

- The number and percentage of patient with >=4 days of non-missing diary entries at each baseline and post-baseline collection visits up to Week 52 separately.
- The number and percentage of patient with >=4 days of non-missing diary entries at both baseline and Week 52.
- The number and percentage of patient with >=4 days of non-missing diary entries at baseline and all scheduled post-baseline visits up to Week 52.

Compliance on the FACIT-Dyspnea symptom score will be calculated as (a) the number and percentage of patients who have non-missing values at baseline and each post-baseline visits up to Week 52 separately, (b) the number and percentage of patients who have non-missing values at both baseline and Week 52, and (c) the number and percentage of patients who have non-missing values at baseline and all scheduled post-baseline visits up to Week 52.

2.4.5.7 Additional efficacy analyses in PAP

No multiplicity adjustment will be conducted for additional efficacy endpoints. Nominal p-values will be reported in the tables. No sensitivity analyses will be performed for the additional efficacy

endpoints. All of the analyses specified below will be done in mITT population using data obtained before initiation of rescue therapy.

Liver function tests

The percentage change in ALT, AST, direct bilirubin, and total bilirubin from baseline up to Week 52 at pre-infusion time point will be analyzed using MMRM model described in Section 2.4.5.1. Data collected at quarterly visits will be included in the MMRM model.

Pulmonary function tests

The percentage change in FEV1 (% predicted), FVC (% predicted) and TLC (% predicted) from baseline up to Week 52 will be analyzed using the MMRM model described in Section 2.4.5.1.

Pulmonary imaging - HRCT

A qualitative assessment will be made for the following features: interstitial lung disease (reticular pattern), ground glass appearance, reticulo-nodular density (nodular densities) and pleural thickening of the right and left lungs based on a scale 0-3, as explained in Section 2.1.4.3.

The change in the two types of mean scores for each feature from baseline up to Week 52 will be analyzed using the MMRM model described in Section 2.4.5.1.

Pulmonary function imaging – chest X-ray

The lung features (interstitial (overall ILD), reticular, nodular, consolidation and pleura) will be scored subjectively for the degree of diffuse lung disease (infiltrative lung disease) in a scale of 0-3 for each side of lungs, as explained in Section 2.1.4.3.

For chest x-ray imaging, interstitial provides an overall assessment of interstitial lung disease with reticular and nodular may provide more supportive information on type of interstitial involvement. Consolidation provides a measure of air space disease and is assessed in addition to interstitial involvement. As HRCT provides a more robust characterization of the diffuse lung disease, only interstitial (overall ILD) of the chest X-ray will be analyzed; however data for all features will be provided in data listings.

The change in mean scores across two lungs for interstitial, which provides a measure of overall interstitial involvement, from baseline up to Week 52 will be analyzed using the MMRM model described in Section 2.4.5.1.

Treadmill ergometry

The parameters of cycle ergometry assessments are noted in Section 2.1.4.3. The change in percent predicted values (or observed values if the percent predicted value is not available) from baseline up to Week 52 will be analyzed using the MMRM model described in Section 2.4.5.1.

Fasting lipids (including lipoproteins of interest)

The percentage change from baseline up to Week 52 values will be analyzed using MMRM model described in Section 2.4.5.1. Summary plots of the LS means from the MMRM of the change from baseline up to Week 52 with 95% CIs will be provided.

Hematology

Platelet and hemoglobin are efficacy parameters. Baseline definitions of hematology parameters are defined in Section 2.6.1. In the visits where two samples are to be collected, the average of two results will be used for summary and plots. Analysis of platelets has already been described under secondary efficacy parameters. For hemoglobin, the MMRM model as used in the primary analysis of spleen volume under primary efficacy endpoints will be used to compare treatment groups for percentage change from baseline up to Week 52, using mITT population. Data collected at quarterly visits will be included in the MMRM model.

Bone disease assessments

Bone marrow burden

The change from baseline to the last observation during PAP will be analyzed using ANCOVA including covariates of baseline and treatment for total BMB score, femur score and spine score separately.

DXA parameters

The change from baseline to the last observation during PAP will be analyzed using ANCOVA model including covariates of baseline and treatment for BMD, T-score and Z-score by the region of femur and spine separately.

Bone biomarkers

Bone-specific alkaline phosphatase and C-telopeptide are bone biomarkers. The change from baseline to Week 52 for these bone biomarkers will be analyzed using MMRM model described in Section 2.4.5.1.

Efficacy biomarkers

Normalized values of serum chitotriosidase, and results from ligand 18 (CCL-18), and angiotensin-converting enzyme (ACE) are efficacy biomarkers. The percentage change from baseline up to Week 52 for efficacy biomarkers will be analyzed using MMRM model described in Section 2.4.5.1.

Physician's global assessment

At the scheduled post-baseline visits, the investigator will also evaluate the patient's current clinical status compared to screening (baseline) by marking 1 of 7 categories and those 7 categories will be converted to numeric values as follows:

- "Marked improvement of daily activities" = 3.
- "Moderate improvement of daily activities" = 2.
- "Mild improvement of daily activities" = 1.
- "No change" = 0.
- "Mild worsening of daily activities" = -1.
- "Moderate worsening of daily activities" = -2.
- "Marked worsening of daily activities" = -3.

Observed numeric values at each post-baseline visits will be analyzed using MMRM model described in Section 2.4.5.1. And in the MMRM model, the baseline value will be Physician's global assessment of severity at screening with value of 1=mild, 2=moderate, 3=severe.

In addition, as supportive analysis, at Week 52, patients will be categorized as responder if having numeric value of 3, 2, or 1, otherwise, patient will be categorized as non-responder if having numeric value of 0, -1, -2, -3 or missing. This binary measure of responder (yes vs no) at Week 52 will be analyzed with logistic regression including treatment group and baseline as covariates.

NMR of HDL

The NMR of HDL data will be provided in data listings.

Liver function via echo-Doppler ultrasound

A frequency shift table for detected signs of portal vein hyperpressure will be provided.

2.4.5.8 PAP+ETP efficacy analyses

All of the efficacy endpoints throughout the study treatment epoch including all longitudinal time points from both PAP and ETP will be summarized by randomized treatment groups in PAP, using mITT population and the same baseline definition for both randomized treatment groups (latest observation prior to the first infusion of the IMP in the study, see Section 2.6.1 for the details of hematology parameters and possibly ECG parameters). ANCOVA model with baseline and treatment group as covariates will be used to assess the means of each treatment group and the 95% CIs of both treatment groups as well as the difference between two groups at each time point. It is understood that all patients in ETP are on olipudase alfa. Summary plots and by-patient plots will be provided as appropriate.

For the primary efficacy endpoints, in addition to the above summaries, the descriptive summaries over time while patients are treated with olipudase alfa using mITT population will be provided.

Since patients who were randomized to placebo group start receiving olipudase alfa when they enter the ETP, their baseline will be defined as the latest observation prior to the first infusion of olipudase alfa.

2.4.6 Analyses of safety data

The summary of safety results will be presented by actual treatment received using safety population. The epochs for safety analysis have been defined in Section 2.1.5. The assessment of safety measures [redacted] will be conducted through pooled analyses and presented in the integrated summary of safety (ISS).

General common rules

All safety analyses will be performed on the safety population as defined in Section 2.3.2, unless otherwise specified, using the following common rules:

- The baseline value is defined as the last available value before the first infusion of the study drug, except for hematology parameters and possibly ECG parameters. For complete definition of baseline, refer to Section 2.6.1.
- The analysis of the safety variables will be essentially descriptive and no systematic testing is planned.

The potentially clinically significant abnormality (PCSA) values are defined by the Sponsor according to predefined criteria/thresholds based on literature review for clinical laboratory tests, vital signs, and ECG; this list is provided in Section 7.2.

- PCSA criteria will determine which patients had at least 1 PCSA during the treatment-emergent adverse event period, taking into account all evaluations performed during the treatment-emergent adverse event period, including unscheduled or repeated evaluations. The number of all patients with any post-baseline evaluation will be the numerator for the on-treatment PCSA percentage.
- The treatment-emergent PCSA denominator by group for a given parameter will be based on the number of patients assessed for that given parameter in the treatment-emergent adverse event period by treatment group on the safety population.
- Before determining a value as PCSA, the value will be rounded to the precision that the PCSA criterion is defined.

2.4.6.1 Adverse events

Pre-treatment adverse events are defined as those that have an onset date/time before the first infusion date/time of the patient. Treatment emergent adverse events are those that have an onset date/time in the treatment epoch. Any AE reported in the database with onset date after the end of the treatment epoch will be classified as post-treatment.

The epochs for safety analysis have been defined in Section 2.1.5. Treatment-emergent adverse events will be presented for different periods under different treatment groups as follows (and as appropriate):

1. Adverse events started in PAP treatment epoch, by randomized treatment group.
2. Adverse events started in either PAP or in ETP while taking olipudase alfa. This will be presented by randomized treatment group as well as combined (ie, all patients treated with olipudase alfa).

Generalities

The primary focus of adverse event reporting will be on treatment-emergent adverse events. Pretreatment adverse events will be described separately.

If an adverse event date/time of onset (occurrence, worsening, or becoming serious) is incomplete, an imputation algorithm will be used to classify the adverse event as pretreatment, or treatment-emergent. The algorithm for imputing date/time of onset will be conservative and will classify an adverse event as treatment emergent unless there is definitive information to determine it is pretreatment. Details on classification of adverse events with missing or partial onset dates are provided in Section 2.6.2. If an AE worsens in severity, two AEs of same verbatim but with different start/end dates and different severity are captured separately on the AE log page in the eCRF.

The AE log page in the eCRF captures the date when the event become serious. If and when this date is non-missing, it indicates that the AE started as non-serious and became serious at a later date. To account for non-serious and serious parts of the AE collected in one AE log page, this AE will be split into two AEs – one non-serious, and one serious AE – with the same AE number, with the following modifications in data for these two AEs:

- Non-serious AE: AE end date, AE outcome, date of death or hospitalization, autopsy information, criterion for serious adverse event (SAE) will be blank.
- Serious AE: AE start date will use the date the event became serious.

Action taken towards study drug and action taken toward AE will be considered in relation to the date the event became serious, the last infusion date and dosing information, and a determination will be made, in consultation with safety monitor of the trial, as to whether the recorded fields belong to non-serious AE or serious AE; then the decision will be implemented in analysis. Any such decision will be rationalized in the CSR.

The adverse events summary will be presented by SOC and PT sorted by the internationally agreed SOC order and decreasing frequency of PTs within SOCs within the olipudase alfa treatment arm, unless otherwise specified. Multiple occurrences of the same event in the same patient will be counted only once in the tables within a particular treatment phase. The denominator for computation of percentages is the safety population within each treatment group.

Start day of treatment emergent AE relative to first infusion = Start date of AE – first infusion date + 1

Start day of pre-treatment AE relative to first infusion = Start date of AE – first infusion date

2.4.6.1.1 Treatment-emergent adverse events

An overview of treatment-emergent adverse events (TEAEs) for PAP will be provided to include (potentially related is defined as either 'related' or 'possibly related' as assessed by the investigator):

- TEAEs
- TEAEs potentially related to study treatment
- TEAEs by severity
- TEAEs leading to treatment discontinuation
- TEAEs leading to study withdrawal
- TEAEs leading to dose reduction
- TEAEs leading to study treatment interruption
- Protocol-defined IARs
- Algorithm-defined IARs
- Treatment-emergent serious adverse events
- TEAEs with fatal outcome (deaths)
- Treatment-emergent pregnancies
- TEAEs which are considered symptomatic overdose
- Dose-limiting toxicities (collected in the adverse event log page in the eCRF, see protocol Section 8.1.6 for details)

The following summaries will be produced by primary SOC and PT:

- TEAEs
- TEAEs by severity (ie, mild, moderate, or severe)
- TEAEs potentially related to study treatment
- TEAEs potentially related to study treatment, by the worst severity (ie, mild, moderate, or severe)

The most common TEAEs are defined as those with percentages of events >=2% and number of patients >=2 in the olipudase alfa treatment group and in addition, the percentage of patients with the specific TEAEs in olipudase alfa treatment group is greater than placebo. The most common TEAEs by SOC and PT will be provided in descending order of olipudase alfa incidence, followed by placebo incidence.

Summary of TEAEs by the latest infusion dose immediately preceding the start of AE will be provided by SOC and PT.

Post-rescue PAP TEAEs by SOC and preferred term will be provided separately in a data listing using the rescue population.

2.4.6.1.2 Adverse events of special interest

Protocol-defined IAR: The following summaries will be produced by primary SOC and PT, and by treatment group for PAP:

- Protocol-defined IARs
- Protocol-defined IARs by the worst severity
- Protocol-defined IARs started during infusion
- Protocol-defined IARs started between 0 hours (exclusive) and 3 hours (exclusive) post end of infusion
- Protocol-defined IARs started between 3 hours (inclusive) and 24 hours (inclusive) post end of infusion
- Protocol-defined IARs started between 24 hours (exclusive) and 72 hours (inclusive) post end of infusion
- Protocol-defined IARs started after 72 hours post end of infusion

In case there are IARs identified by the sponsor (before database lock) that the investigator did not agree to be an IAR, sponsor-identified IARs may be added to the list of investigator-identified IARs to define the protocol-defined IARs, and additional summaries may be produced.

A subset of protocol-defined IARs will also be defined programmatically by searching for preferred terms within Hypersensitivity reaction SMQs (identified using SMQ code 20000214 Broad and Narrow scope). The summaries of Hypersensitivity IARs will be provided by primary SOC and PT.

Algorithm-defined IAR: Summary of algorithm-defined IARs by primary SOC and PT will be provided for PAP.

2.4.6.1.3 Adverse events of other interest

Infections: Infection AEs are already part of system organ class of Infections and Infestations in all TEAE displays. No separate display is planned.

Bleeding: Bleeding related TEAEs will be summarized using the standard MedDRA queries (SMQs) Hemorrhages (identified using SMQ code 20000038 broad and narrow scope).

2.4.6.1.4 Serious adverse events

The following summaries will be provided by primary SOC and PT for PAP:

- Treatment-emergent serious adverse events

2.4.6.1.5 Adverse events leading to permanent treatment discontinuation

In this study, if the treatment is discontinued permanently, the patient is withdrawn from the study. A temporary dose interruption is not considered permanent discontinuation of study treatment. Any AEs leading to permanent treatment discontinuation will be summarized by primary SOC and PT - for PAP, ETP and cumulative over PAP and ETP. The listing of such AEs will include the olipudase alfa dose of the infusion prior to or during the occurrence of the AE leading to treatment discontinuation. Any AE leading to discontinuation of study treatment post-rescue in PAP will be provided in data listing.

2.4.6.1.6 Adverse events leading to dose reduction

Any AEs leading to dose reduction will be listed including the olipudase alfa dose of the infusion prior to or during the occurrence of the AE leading to dose reduction, where applicable.

2.4.6.1.7 Adverse events leading to drug interruption

Any AEs leading to drug interruption will be listed including the olipudase alfa dose of the infusion prior to or during the occurrence of the AE leading to drug interruption, where applicable.

2.4.6.2 Laboratory data

For parameters that are reported as below the quantification limit (BQL), the results often convey a clinically meaningful change. Therefore, a numerical value is necessary to be imputed for including these in descriptive statistics calculations. For all these parameters either the lower limit of quantification (LLOQ) is available in the database, or the result is reported as <xx. The numeric value for results of BQL will be imputed as LLOQ/2; the numeric value for results reported as <xx will be imputed as xx/2.

In hematology, platelets and hemoglobin are part of efficacy parameters, and have been described under Section 2.1.4. For the rest of hematology parameters, the baseline will be defined the same way (Section 2.6.1), and any visit which is scheduled to collect two samples pre-infusion for WBC, the average of the two values will be used as pre-infusion value.

Observed values, change from baseline and percent change from baseline to scheduled study time points will be summarized in separate groups of hematology (excluding hemoglobin and platelets which are considered efficacy parameter, hence the analyses have been described in Section 2.4.5), liver function tests, coagulation, etc. as described in Section 2.1.5 (excluding lipids which have been described as part of efficacy, Section 2.4.5.7). Summary plots over time will also be provided as needed.

Laboratory values will be classified as normal, above normal, or below normal based on normal ranges when provided by the laboratory. Shifts from baseline to scheduled time points, any time during the on-treatment period and the last available value, in the abnormality of any parameters will be provided.

PCSA values for clinical chemistry, hematology and liver function tests (provided in Section 7.2) during the study (from first infusion to last available observation) will be summarized. PCSA values for hematology, liver function test and clinical chemistry parameters will be provided in separate listings.

2.4.6.3 Vital signs

Vital signs will include heart rate (HR, beats/minute), systolic and diastolic blood pressure (SBP/DBP, mmHg), respiratory rate and body temperature (°F or °C). Each vital sign parameter will be summarized as observed parameter value and change from baseline at each scheduled time point.

PCSA values (provided in Section 7.2) during the study (from first infusion start to last available observation) will be summarized.

2.4.6.4 Electrocardiogram variables

For ECG parameters, the baseline is defined in Section 2.6.1. The observed data and change from baseline will be summarized at each scheduled time point.

PCSA values (provided in Section 7.2) will be summarized during the study (from first infusion start to last available observation).

2.4.6.5 Echocardiogram with Doppler

Shift tables will be created for valve characterization (including aortic regurgitation, mitral regurgitation, pulmonic regurgitation, tricuspid regurgitation). Summary of change from baseline in cardiac index, ejection fraction (%), pulmonary artery systolic pressure will be provided.

2.4.6.6 Physical examinations

Complete physical examinations will be classified as normal or abnormal. Shifts from baseline to scheduled time points, any time during the on-treatment period and the last available value, in the abnormality of any complete physical examinations (except overall neurological exam - please see Section 2.4.6.7) will be provided. A listing of the abbreviated physical examination data will be provided.

Observed values and change from baseline for weight, height and BMI will be summarized for all scheduled time points.

2.4.6.7 Extended neurological examinations

Neurological examinations (Section 2.1.5.6) will be classified as normal/abnormal findings. Frequency shift table will be provided. The overall evaluation of neurological exam collected in the eCRF under physical exam will be included in this section.

2.4.6.8 Liver biopsy

Liver biopsy is planned at screening, Week 52 and Week 104. Liver biopsies are evaluated by a Sanofi Genzyme pathologist for liver fibrosis grading and for sphingomyelin accumulation. Fibrosis staging will be performed using the Lathe pathology reports. Liver biopsy report will not be stored in clinical database, but will be received by clinical personnel at the site and kept in the patient local study file. Data of fibrosis staging will be recorded and transferred as external data.

The fibrosis stage will be summarized with number of patients and percentage for each stage at each visit. Cirrhosis and liver failure could be interpreted from this data. Change from baseline in Sphingomyelin accumulation in liver tissue that is assessed from liver biopsy will be summarized under pharmacodynamic endpoints (Section 2.1.7).

2.4.6.9 Liver function via echo-Doppler ultrasound

Liver ultrasound Doppler data will be provided in listings.

2.4.6.10 Inflammatory and vascular biomarkers

Results of inflammatory and vascular biomarkers will be provided in data listings. IL-6 and IL-8 are part of safety biomarkers and will not be included here; for details of IL-6/IL-8 analysis, please refer to Section 2.4.6.11.

2.4.6.11 Safety biomarkers

Observed values, change from baseline and percent change from baseline will be summarized at each scheduled time point. Summary plot for the mean ceramide levels in plasma over time also be presented. By-patient plots for the observed values over time will be provided for the highly sensitive C-Reactive protein, iron, ferritin, troponin I, calcitonin, Interleukin 6 and Interleukin 8 will be provided.

2.4.6.12 Immune response assessments

For ADA summaries during PAP period, the baseline is the last non-missing ADA assessment prior to the first IMP. For ADA summaries during PAP+ETP for duration on olipuase alfa exposure, the baseline is the last non-missing ADA assessment prior to the first oliupdase alfa infusion.

Data listings will be provided to display results of anti-olipudase IgG antibody, neutralizing antibody, anti-olipudase alfa IgE antibody, serum tryptase activity, complement activation, and skin testing performed. If IgG antibody titer value is reported as <xx, the numeric value will be imputed as xx, the minimum required dilution (MRD), and used for summary purposes. Listing will provide results as reported as well as results as imputed.

A patient with at least one anti-drug antibody (ADA) response available in the database is considered evaluable.

A patient whose ADA status is positive at baseline is considered to have pre-existing ADA.

A patient whose ADA status is positive anytime post-baseline and is negative or missing at baseline is considered to have treatment-induced ADA.

A patient whose ADA status is positive at baseline (pre-existing ADA) and the ADA titer level anytime post-baseline is significantly higher than that at baseline is considered to have treatment-boosted ADA. A difference in titer values between 2 samples representing greater than or equal to twice the dilution level is considered significant. In other words, the post-baseline titer value divided by the baseline titer value will be greater than or equal to 4 to be considered significant. For example, if baseline titer value is 50, the post-baseline titer value has to be >=200.

For patients with treatment-induced ADA, the following 3 ADA responses are defined.

Transient ADA response is defined as:

- Treatment-induced ADA detected only at one sampling time point post-baseline (excluding the last sampling time point).
- Treatment-induced ADA detected at two or more sampling time points post-baseline, where the first and the last ADA-positive samples (irrespective of any negative samples in between) are separated by a period of less than 16 weeks, and the subject's last sampling time point is ADA-negative.

Persistent ADA response is defined as:

- Treatment-induced ADA detected at two or more sampling time points post-baseline, where the first and the last ADA-positive on-treatment sample (irrespective of any negative samples in between) are separated by at least 16 weeks.
- Treatment-induced ADA detected in the last two sampling time points, irrespective of the time period in between.

The following subclassifications for persistent ADA response will be considered as well. Since this study is ongoing, the final assessment will be the final assessment at the time of data cutoff date for initial BLA submission.

- Low response – if a patient's peak titer ≤400 and positive at final assessment. This represents the first titer that is greater than a 4-fold increase from the assay minimum required dilution (MRD). Titers within this range would be considered as Low response.
- Intermediate response - if a patient had persistent ADA response but peak titer is >400 and positive at final assessment.
- Tolerized - if a patient had persistent ADA response, but negative at the final assessment.

Indeterminate ADA response is defined as:

- Only the last sampling time point is positive and all previous samples are negative.

Pre-existing ADA = 100*(Number of patients with pre-existing ADA at baseline) / number of evaluable patients

ADA Incidence = 100*(treatment-boosted + treatment-induced ADA positive patients) / number of evaluable patients

The following information will be provided:

- Number of ADA evaluable patients
- Number (%) of patients with ADA positive/negative at baseline
- Number (%) of patients that never develop ADA at any time (ie, ADA is always negative during baseline and post-baseline visits)
- Number (%) of patients with treatment-boosted ADA
- Number (%) of patients with treatment-induced ADA
- Treatment emergent ADA: Number (%) of patients with treatment-boosted/treatment-induced ADA
- Duration of ADA response: Number (%) of patients with transient/persistent/indeterminate ADA response (each category separately)
- Number (%) of patients within persistent ADA response subclassifications: low response, intermediate response, tolerized (each category separately)
- Patients with treatment-induced ADA will be further characterized
  - Peak ADA titer: Median and $25^{th}/75^{th}$ quantiles (min, max) of the ADA titer for peak ADA titer of treatment-induced ADA positive patients
  - Last ADA titer (final assessment): Median and $25^{th}/75^{th}$ quantiles (min, max) of the last ADA titer of treatment induced ADA positive patients
- Number (%) of patients with positive neutralizing antibody (NAb)
  - NAb Inhibition of catalytic activity
  - NAb inhibition of cellular uptake

The assessment of ADA incidence and characterization [redacted] will be conducted through pooled analyses and presented in the integrated summary of immunogenicity. The assessment of immunogenicity and safety [redacted] will be presented in the integrated summary of safety (ISS).

2.4.6.13 Subgroup analyses for key safety measures in PAP

The following subgroup analyses for key safety measures will be performed:

- Baseline ALT or AST abnormality (ALT or AST ≥1 ULN vs ALT and AST <1 ULN)
- Baseline total bilirubin abnormality (total bilirubin ≥1.5 ULN vs total bilirubin <1.5 ULN)

The key safety measures that will be analyzed by the above subgroup include the followings:

- Overview of treatment-emergent adverse events
- Summary of treatment-emergent adverse events by primary SOC and PT
- Summary of protocol-defined infusion-associated reactions by primary SOC and PT

2.4.6.14 PAP+ETP safety analyses

For categorical measures, such as AE, data throughout both PAP and ETP will be summarized as three groups:

- **Previous placebo group** includes patients who are randomized to placebo. For patients in this group, only the data during olipudase alfa exposure at ETP will be summarized. The baseline will be defined as the last non-missing value prior to first olipudase alfa infusion at ETP.
- **Previous olipudase alfa group** includes patients who are randomized to olipudase alfa. For patients in this group, all the data during PAP and ETP with olipudase alfa exposure will be summarized. The baseline will be defined as the last non-missing value prior to first olipudase alfa infusion at PAP.
- **Overall Olipudase alfa group** includes patients from previous placebo group and Previous Olipudase alfa group.

For continuous measures collected at longitudinal time point, such as labs and vital signs, data throughout both PAP and ETP will be summarized as two groups:

- **Placebo/olipudase alfa group** includes patients who are randomized to placebo. For patients in this group, all data during PAP and ETP will be summarized at each longitudinal time point. The baseline will be defined as the last non-missing value prior to first infusion of IMP at PAP.
- **Olipudase alfa/Olipudase alfa group** includes patients who are randomized to olipudase alfa. For patients in this group, all the data during PAP and ETP with olipudase alfa exposure will be summarized at each longitudinal time point. The baseline will be defined as the last non-missing value prior to first infusion of IMP at PAP.

2.4.7 Analyses of pharmacokinetic and pharmacodynamic variables

2.4.7.1 Pharmacodynamic analysis

Pharmacodynamic endpoints include sphingomyelin and its metabolites, including sphingomyelin accumulation in liver tissue. These endpoints will be analyzed using pharmacodynamics population. Observed, change from baseline values and percent change from baseline values will be summarized at scheduled visits. Summary plots of the mean values over time for all pharmacodynamics endpoints will be provided.

2.4.7.2 Pharmacokinetic analysis

Plasma concentration-time data will be analyzed by non-compartmental methods, or nonlinear mixed effects modeling, based upon patient age and data suitability. Values will be reported for individual patient and summarized by dose level, process materials and study week as appropriate.

For ease of presentation, mean values will be arithmetic mean unless specified. Concentration values below the plasma assay limit will be treated as zero in calculating mean values. Mean values below lower limit of quantification (LLOQ) will be reported as LLOQ in the tables and not plotted in the figures if after C_{max}. Mean calculations and their associated statistics will be generated from unrounded numbers and may differ slightly from those values that would be determined using the rounded numbers displayed in the tables. Values expressed in all tables will be for ease of presentation and will not be meant to imply accuracy to more than 3 significant figures.

Pharmacokinetic parameters of olipudase alfa may be summarized using descriptive statistics such as arithmetic mean, geometric mean, SD, standard error of the mean (SEM), coefficient of variation (CV [%]), minimum, median, maximum, and number of observations for each study visit, and dose level under the responsibility of pharmacokinetics, dynamics and metabolism (PKDM) using WinNonlin Professional, or Phoenix, or NONMEM (VII or above) running on a Linux cluster.

Associations between ADA variables (eg, ADA status) and PK of olipudase alfa (eg, AUC) will be explored for the olipudase alfa treatment group at selected PK visits. If applicable, plot of olipudase alfa PK (eg, AUC) versus the selected PK visit may be provided by ADA classifications for the olipudase alfa treatment group. Nab data may be explored as appropriate.

2.4.8 Analyses of quality of life (patient reported outcome)/health economics variables

All of the analyses specified below will be done in mITT population. Observations collected after the initiation of the rescue therapy will be excluded from the analysis. Exploratory p-values will be provided, if applicable, from MMRM model described in Section 2.4.5.1 for analyses of data from PAP.

Brief fatigue inventory

The change from baseline up to Week 52 on severity and interference of fatigue as measured by BFI (individual questions as well as scale score) will be analyzed using the MMRM model described in Section 2.4.5.1.

Brief Pain Inventory – Short Form (BPI-SF)

The change from baseline up to Week 52 on severity (average of Questions 3, 4, 5, 6) and interference (Q9) of pain as measured by BPI-SF (individual questions as well as scale score) will be analyzed using the MMRM model described in Section 2.4.5.1.

SF-36

The change from baseline up to Week 52 on 8 scales measured by SF-36 - physical functioning, role physical, bodily pain, mental health, role emotional, social functioning, vitality, and general health - as well as the 2 summary measures of physical health (PCS) and mental health (MCS) will be analyzed using the MMRM model described in Section 2.4.5.1.

Niemann-Pick B Health Assessment Questionnaire (NPB-HAQ)

Three domains of NPB-HAQ – resource utilization, symptoms and physical activity - will be summarized separately with each question at each scheduled time point. For resource utilization, in the last three months of each scheduled time point, the following measures will be summarized: the number of physician visits (total number of physician visits by adding up all the visits from different specialists), the number of emergency room visits, the number of days of paid caregiver assistance and the number of DAYS of care/help from a family member(s) and/or friend(s). For symptom in the last three months of each scheduled time point, frequency and level of bothersome of each symptom will be listed. For physical activity in the last three months of each scheduled time point, the number and percentage of patients at each frequency of physical activities will be summarized.

FACIT-Dyspnea

The symptom scores have been discussed under secondary efficacy endpoints. Functional limitation scores will be provided in data listing.

Health-related productivity

Each question (about employment, household chores, homework/classes) will be summarized separately at each scheduled time point. At each scheduled time point, the following number of hours missed due to ASMD or its treatment will be summarized: the number of hours missed of work, the number of hours missed of house hold chores, the number of hours missed of homework or class. If the answer for ASMD or its treatment(s) keep you from work, house hold chores, homework or class is “no”, then number of hours missed will be imputed as 0 for the summary.

EQ-5D-5L

Each question on a 1-5 scale (for mobility, self-care, usual activities, pain/discomfort, anxiety/depression), higher score indicating worse, will be summarized separately at each scheduled time point, providing number and % of patients at each level. The change from baseline up to Week 52 on the VAS scale score on “your health today” (0 indicating the worst, and 100 indicating the best) will be analyzed using the MMRM model described in Section 2.4.5.1.

PGIS/PGIC

For PGIC, 7 categories of response will be converted to the following numeric values:

- Very much Better = 3
- Moderately Better = 2

- A Little Better = 1
- No Change = 0
- A Little Worse = -1
- Moderately Worse = -2
- Very much Worse =-3

Observed numeric values at each post-baseline visits will be analyzed using MMRM model described in Section 2.4.5.1. And in the MMRM model, the baseline value will be PGIS at screening with value of 0=none, 1=mild, 2=moderate, 3=severe, 4=very severe.

In addition, as supportive analysis, at week 52, patients will be categorized as responder if having numeric value of 3, 2, or 1, otherwise, patient will be categorized as non-responder if having numeric value of 0, -1, -2, -3 or missing. This binary measure of responder (yes vs no) at Week 52 will be analyzed with logistic regression including treatment and baseline as covariates.

The PGIS at baseline and post-baseline visits will be converted to a numeric score on 0-4 scale (0=none, 1=mild, 2=moderate, 3=severe, 4=very severe). Therefore, change from baseline at each post-baseline visits using MMRM model described in Section 2.4.5.1.

2.4.8.1 PAP+ETP quality of life analyses

All of the quality of life endpoints throughout the study treatment epoch including all longitudinal time points from both PAP and ETP will be summarized by randomized treatment group in PAP, using mITT population. ANCOVA model with baseline and treatment group as covariates will be used to assess the means of each treatment group and the 95% CIs of both treatment groups as well as the difference between two groups at each time point. It is understood that all patients in ETP are on olipudase alfa.

2.5 PATIENT NARRATIVES

Patient narratives criteria for this study are as follows:

- Deaths
- Serious adverse event
- Adverse Events leading to permanent treatment discontinuation
- Adverse events of special interest including the followings:
  - Protocol defined IARs (as specified in Section 2.1.5.1)
  - Pregnancy of female patient or partner of a male patient
  - Symptomatic overdose of study drug
  - Any increase in AST, ALT, total bilirubin, or AP >3x baseline (prior to olipudase alfa therapy) and > ULN

- Any increase in total bilirubin or AP >1.5x baseline in the presence of AST or ALT >2x ULN
- Any increase in ALT or AST >3x the ULN combined with an increase in ALT or AST >2x baseline (prior to olipudase alfa therapy) with symptoms of fatigue, nausea, vomiting, right upper quadrant pain or tenderness (for adults only), fever, rash, or eosinophilia (> ULN)

Patients who meet these criteria will be listed and provided to the medical writer.

2.6 DATA HANDLING CONVENTIONS

2.6.1 General conventions

In general, the baseline value is defined as latest value prior to the start of first infusion, except for hematology parameters and possibly ECG parameters. For hematology parameters, the baseline value is defined as the average of all available values before the start of first infusion. Other exception may be the ECG parameters - Day 1/Week 0 (the first infusion visit) collects ECG in triplicate before the infusion; if 2 or more than 2 of these triplicate collections of ECG are before the first infusion, the average of these ECG parameter results will be considered as the baseline value; otherwise, the baseline value of ECG parameters will be taken from the latest observation before the first infusion.

Age at ASMD diagnosis = number of years between the date of birth and the date of ASMD diagnosis.

2.6.2 Missing data

In general, no imputation is planned for missing data. The following approaches are default methods for missing data handling. Some exploratory analyses can be planned with different strategies for treating missing outcomes.

- Categorical data at baseline will be summarized using counts (n) and percentages (%). The number of patients with missing data may be mentioned, but will not be included in the denominator for the calculation of percentages unless otherwise specified. When relevant, the number of patients with missing data is presented.
- Continuous data: The analyses and summaries for variables with continuous scales will be based on observed data only.

Handling of computation of treatment duration if investigational medicinal product end of treatment date is missing

For the calculation of the treatment duration, the date of the last dose of study treatment is reported on the last infusion page in the eCRF.

Handling of medication missing/partial dates

No imputation of medication start/end dates or times will be performed. If a medication date or time is missing or partially missing and it cannot be determined whether it was taken prior to the start of first infusion or concomitantly with study, it will be considered a prior as well as a concomitant medication.

Handling of adverse events with missing data

- If the AE is missing start date and/or time, the treatment-emergent adverse event will be determined by the following conservative principle: An adverse event will be considered a treatment-emergent adverse event if it cannot be confirmed that the event is not a treatment-emergent adverse event due to missing data. Further, if the PAP or ETP period could not be determined, this AE will be considered as TEAE in PAP.
- If the timing of AE with respect to infusion start and/or end time could not be determined unambiguously, the AE will be assumed to be an algorithm-defined IAR.
- If the assessment of the relationship is missing, then the relationship of possibly related is assumed for the summary table, but no imputation will be done at the data level.
- If the worst severity grade is missing, a missing category will be added in the summary table.

2.6.3 Windows for time points

Statistical analysis will be based on the nominal visit which should follow schedule defined in protocol Section 1. Visits will not be windowed for analysis purposes, except for unscheduled visits in some cases as described in Section 2.6.4.

2.6.4 Unscheduled visits

In general, unscheduled visit measurements will not be included in the by-visit summaries, but will be used for computation of baseline and/or summaries of “any time” or “end of study” time points. If a scheduled post-baseline visit measurement of an efficacy parameter is missing, but an unscheduled visit value for the same parameter is available within 3 months (91 days) of scheduled 6-month or 12-month visit, the unscheduled visit value will be used to substitute the missing scheduled visit value before any missing data imputation algorithm is implemented. If the parameter is scheduled for evaluation every 3 months, a 45 day window will be used instead of 91 days. If two unscheduled visits qualify for missing scheduled visit value, the one closest in date to target visit date will be used. If these two unscheduled visits happened in the same distance to the target visit date with one before and one after, the one before the target visit date will be used. Laboratory, vital signs, ECG and pharmacodynamic parameters are collected at several times during a visit with respect to the infusion time, therefore a window for an unscheduled visit to substitute missing schedule values will not be considered. For ADA analyses, unscheduled visits will be included in the by-visit summaries.

2.6.5 Pooling of centers for statistical analyses

Not applicable to the primary analysis. Investigation of the effects of geographic regions may be performed on an exploratory basis.

2.6.6 Statistical technical issues

A sensitivity analysis of the primary efficacy endpoints using a pattern mixture model and different imputations for values missing during the on-treatment period versus values missing after treatment discontinuation is planned. The description of the analysis refers to the DLco endpoint; differences in the implementation for the spleen volume and splenomegaly-related scores are noted. On-treatment period is defined as the time period between the start of first infusion of the study drug up to the day of last infusion in PAP +13 days for patients who did not continue in ETP, or up to the day before first infusion in ETP if the patient continued into ETP.

A pattern-mixture model sensitivity analysis of the percentage change in DLco (% predicted) from baseline to 52 weeks will be conducted, with a different imputation strategy applied for missing DLco values during the on-treatment period and missing DLco values after treatment discontinuation (ie, after the day of last infusion +13 days) based on the following assumptions:

- Patients within 13 days of their last IMP infusion would continue to show benefit from treatment similar to that observed at the scheduled time point. Therefore, DLco values missing during the on-treatment period will be considered “Missing At Random” and imputed using a model estimated using all samples collected on-treatment.
- Patients who stop taking their study treatment will no longer benefit from it in the future, and thus will tend to have DLco values returning to baseline. Thus DLco values missing after treatment discontinuation will be imputed based on a patient’s own baseline value.

Missing DLco values will be imputed 100 times to generate 100 complete data sets. The percent change from baseline to 52 weeks will be derived from observed and imputed DLco at this time point. The complete data sets will be analyzed using an ANCOVA model with treatment group, baseline DLco and baseline age as covariates. The results from the 100 analyses will be combined using Rubin’s formulae.

Imputation of missing data during the on-treatment period

Missing DLco values during the on-treatment period will be imputed from other on-treatment measurements assuming MAR, using SAS® PROC MI.

Only DLco values collected during the on-treatment period will be included in the imputation model. This way, missing DLco values during the on-treatment period will be imputed based solely on observed on-treatment calculated DLco values.

The imputation model will include baseline DLco, baseline age and all observed DLco values at weeks 26 and 52 and will be stratified by treatment group of placebo and olipudase alfa. Since the

pattern of missing data could be non-monotone, a MCMC method will be used. A minimum value of 0 will be specified in order to avoid negative imputed DLco values.

Sample SAS code is provided below:

```
proc mi data=DATAIN out=DATAOUT nimpute=100 minimum=0;
  var DLCO_BASE AGE_BASE DLCO_W26 DLCO_W52;
  by TRTPN;
run;
```

For the splenomegaly-related score endpoint which is collected at Weeks 14, 26, 38 and 52, the above code will be modified accordingly.

As stated above, the input dataset DATAIN will include only DLco values collected during the on-treatment period. Any DLco values collected after the treatment discontinuation (after last infusion +13 days) will be excluded from the input dataset. In practice, the MI procedure will generate imputed values for all missing values (whether on-treatment or post-treatment), but only imputed values during the-on-treatment period will be kept in the final datasets that will be analyzed using ANCOVA. Imputed values during the post-treatment period will be discarded and replaced by imputed values described in the next section.

Imputation of missing data after treatment discontinuation

Missing DLco values during the post-treatment period will be imputed assuming DLco values will, on average, return to baseline values.

For each patient, missing post-treatment DLco values will be imputed 100 times, using a random draw from a normal distribution, with mean equal to a patient's own baseline value and variance equal to the conditional variance at the specific time-point, given the baseline value.

Let Y_0 and Y_1 denote the DLco at baseline and at the specific time-point respectively. Since Y_0 and Y_1 are assumed to have a bivariate normal distribution, the conditional variance of Y1 given Y_0 is:

$$Var(Y_1 | Y_0 = y_0) = \sigma_1^2 (1 - \rho^2)$$

where denotes the variance of Y_1 and ρ the coefficient of correlation between Y_0 and Y_1.

The conditional variance will be estimated from observed data within the same treatment arm at the specific time-point.

During the random generation process, a minimum value of 0 will also be applied in order to avoid negative imputed DLco values.

3 INTERIM ANALYSIS

There is no formal interim analysis before the primary timepoint (Week 52). Periodic unblinded safety data review is carried out by an external Data Monitoring Committee, aided by external vendor providing statistics and programming support. Efficacy will be assessed when all patients complete 52 weeks in PAP and data are unblinded to the study team in preparation for the CSR for PAP. For ETP data, efficacy analysis is exploratory.

4 DATABASE LOCK

The database is planned to be locked within 6 weeks after the last patient last visit in PAP.

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS Version 9.0 or higher. The exact version used will be documented in the clinical study report.

6 REFERENCES

1. Crapo RO, Morris AH. Standardized single breath normal values for carbon Monoxide Diffusing Capacity. Am Rev Respir Dis. 1981;123:185-9.

2. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J. 2005;26:720-35.

3. Mendoza TR, Wang XS, Cleeland CS, Morrissey M, Johnson BA, Wendt JK, et al. The rapid assessment of fatigue severity in cancer patients: Use of the Brief Fatigue Inventory. Cancer. 1999;85(5):1186-96.

4. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9(2):105-21.

5. Wang XS, Cleeland CS, Mendoza TR, Engstrom MC, Liu S, Xu G, et al. The effects of pain severity on health-related quality of life: A study of Chinese cancer patients. Cancer. 1999;86:1848-55.

6. Cleeland CS, Ryan KM. Pain assessment: Global use of the Brief Pain Inventory. Ann Acad Med Singapore. 1994;23(2):129-38.

7. Cleeland, CS. The brief pain inventory. MD Anderson, TX. Copyright: 2009.

8. Choi SW, Victorson DE, Yount S, Anton S, Cella D. Development of a conceptual framework and calibrated item banks to measure patient-reported dyspnea severity and related functional limitations. Value Health. 2011;14(2):291-306.

9. Global Lung Initiative (GLI) reference. Available from: URL: http://www.ers-education.org/guidelines/global-lung-function-initiative.aspx

10. Jacobson TA, Ito MK, Maki KC, Orringer CE, Bays HE, Jones PH, et al. National Lipids Assocation recommendations for patient-centered management for dyslipidemia: Part 1- full report. J Clin Lipidol. 2015;9:129-69.

11. Thurberg BL, Wasserstein MP, Schiano T, O'Brien F, Richards S, Cox GF, et al. Liver and skin histopathology in adults with acid sphingomyelinase deficiency (Niemann-Pick Disease Type B). Am J Surg Pathol. 2012;36(8):1234-46.

12. Kammerman LA, Cappelleri J, Bartlett J. Patient-Reported Outcomes. Wiley StatsRef: Statistics Reference. [Online]. 2017. DOI: 10.1002/9781118445112.stat07929

13. Sloan JA, Dueck AC, Erikson PA et al. Analysis and interpretation of results based on patient-reported outcomes. Value Health. 2007;10(2):S106-15.

14. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey: Conceptual framework and item selection. Med Care. 1992;30(6):473-83.

15. Maruish ME. User's Manual for the SF-36v2 Health Survey. 3rd ed. Lincoln, RI: QualityMetric, Inc; 2011.

16. EQ-5D User guide. Available from: URL:https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf

17. Mallinckrodt CH, Lane PW, Schnell, D, Peng Y, Mancuso JP. Recommendations for the Primary Analysis of Continuous Endpoints in Longitudinal Clinical Trials. Drug Information Journal. 2008;42:303-19.

18. Siddiqui O, Hung HMJ, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat. 2009;19:227-46.

19. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics.1997;53:983-97.

20. Samuel-Cahn E. Is the Simes improved Bonferroni procedure conservative? Biometrika. 1996;83(4):928-33.

21. Pastores GM, Weinreb NJ, Aerts H, Andria G, Cox TM, Giralt M, et al. Therapeutic goals in the treatment of Gaucher Disease. Semin Hematol. 2004;41(4 Suppl 5):4-14.

22. Xaubet A, Molina-Molina M, Acosta O, Bollo E, Castillo D, Fernández-Fabrellas E, et al. Guidelines for the medical treatment of idiopathic pulmonary fibrosis. Arch Bronconeumol. 2017;53(5):263-9.

23. Cottin V, Crestani B, Cadranel J, Cordier JF, Marchand-Adam S, Prévot G, et al. French practical guidelines for the diagnosis and management of idiotpathic pulmonary fibrosis - 2017 update. Full-length version. Rev Mal REspir. 2017;34(8):900-68.

24. Khanna D, Mittoo S, Aggarwal R, Proudman SM, Dalbeth N, Matteson EL, et al. Connective Tissue diseases associated interstitial lung diseases (CTD-ILD) -- report from OMERACT CTD-ILD working group. J Rheumatol. 2015;42(11):2168-71.

25. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J. 2005;26(5):948-68.

26. Horita N, Miyazawa N, Kojima R, Inoue M, Ishigatsubo Y, Kaneko T. Minimum clinically important difference in diffusing capacity of the lungs for carbon monoxide among patients with severe and very severe chronic obstructive pulmonary disease. COPD. 2015;12(1):31-7.

27. Punjabi NM, Shade D, Patel AM, Wise RA. Measurement variability in single-breath difusing capacity of the lung. Chest. 2003;123(4):1082-9.

7 LIST OF APPENDICES

7.1 FACIT-DYSPNEA SHORT FORM - CONVERSION FROM RAW TO SCALE SCORES

Raw dyspnea score	Scale score	SE	Raw functional limitations score	Scale score	SE
0	27.7	4.7	0	29.7	4.7
1	32.8	3.7	1	34.9	3.4
2	36.1	3.2	2	38.0	3.0
3	38.6	2.8	3	40.3	2.7
4	40.6	2.6	4	42.1	2.5
5	42.3	2.4	5	43.8	2.4
6	43.8	2.2	6	45.2	2.3
7	45.2	2.2	7	46.5	2.2
8	46.4	2.1	8	47.8	2.2
9	47.6	2.1	9	49.0	2.2
10	48.8	2.0	10	50.1	2.1
11	50.0	2.0	11	51.2	2.1
12	51.1	2.0	12	52.3	2.1
13	52.1	1.9	13	53.4	2.1
14	53.2	1.9	14	54.4	2.0
15	54.2	1.9	15	55.4	2.0
16	55.2	1.9	16	56.4	2.0
17	56.2	1.9	17	57.4	2.0
18	57.2	1.9	18	58.4	2.0
19	58.1	1.9	19	59.4	2.0
20	59.2	1.9	20	60.4	2.1
21	60.2	2.0	21	61.4	2.1
22	61.2	2.0	22	62.4	2.1
23	62.3	2.0	23	63.5	2.2
24	63.5	2.1	24	64.7	2.2
25	64.8	2.1	25	66.0	2.3
26	66.1	2.3	26	67.3	2.4
27	67.7	2.4	27	68.9	2.6
28	69.5	2.6	28	70.7	2.8
29	71.9	3.0	29	73.0	3.2
30	75.9	4.0	30	76.7	4.1

7.2 POTENTIALLY CLINICALLY SIGNIFICANT ABNORMALITIES CRITERIA

Liver function tests	
ALT	>3 x ULN >5 x ULN >10 x ULN >20 x ULN
AST	>3 x ULN >5 x ULN >10 x ULN >20 x ULN
Alkaline Phosphatase	>1.5 x ULN
Total Bilirubin	>1.5 x ULN >2 x ULN
ALT and Total Bilirubin	ALT >3 x ULN and Total Bilirubin >2 x ULN
Hematology	
WBC	<3.0 GIGA/L (non-Black), <2.0 GIGA/L (Black), >=16.0 GIGA/L
Lymphocytes	>4.0 GIGA/L
Neutrophils	<1.5 GIGA/L (non-Black) <1.0 GIGA/L (Black)
Monocytes	>0.7 GIGA/L
Basophils	>0.1 GIGA/L
Eosinophils	>0.5 GIGA/L or > ULN if ULN >=0.5 GIGA /L
Hemoglobin	Males: <=115 g/L (<=7.14 mmol/L) and 20% decrease from baseline; >=185 g/L (>=11.48 mmol/L) Females: <=95 g/L (<=5.9 mmol/L) and 20% decrease from baseline; >=165 g/L (>=10.24 mmol/L) Decrease from Baseline >=20 g/L (1.24 mmol/L)
Hematocrit	Males: <=0.37 v/v and 20% decrease from baseline; >=0.55 v/v Females: <=0.32 v/v and 20% decrease from baseline; >=0.5 v/v
RBC	>=6 TERA/L
Platelets	<100 GIGA/L and 20% decrease from baseline >=700 GIGA/L

ECG – PCSA criteria	
HR	<=50 bpm and decrease from baseline >=20 bpm
	>=120 bpm and increase from baseline >=20 bpm
PR	>=200 ms and increase from baseline >=20 ms
QRS	>=120 ms
QTc	Absolute values (ms) Males: 431-450ms; >450 ms; ≥500 ms Females: 451-470 ms; >470 ms; ≥500 ms Increase from baseline (Males and Females) 30-60 ms >60 ms
Clinical chemistry	
Creatinine	>=150 µmol/L (Adults)
	>=30% increase from baseline
	>=100% increase from baseline
Blood Urea Nitrogen	>=17 mmol/L
Chloride	<80 mmol/L
	>115 mmol/L
Sodium	<=129 mmol/L
	>=160 mmol/L
Potassium	<3 mmol/L
	>=5.5 mmol/L
Glucose	
Hypoglycemia	<=3.9 mmol/L and < LLN
Hyperglycemia	>=11.1 mmol/L (unfasted), >7 mmol/L (fasted)
Albumin	<=25 g/L
Vital signs	
Heart rate	<=50 bpm and decrease from baseline >=20 bpm
	>=120 bpm and increase from baseline >=20 bpm
Systolic BP	<=95 mmHg and decrease from baseline >=20 mmHg
	>=160 mmHg and increase from baseline >=20 mmHg
Diastolic BP	<=45 mmHg and decrease from baseline >=10 mmHg
	>=110 mmHg and increase from baseline >=10 mmHg
Urinalysis	
pH	<=4.6
	>=8

Signature Page

dfi12712-16-1-9-sap

Approve & eSign	[redacted]
Approve & eSign	[redacted]