NCT02782741

STATISTICAL ANALYSIS PLAN

A Phase 3 randomized, multicenter, multinational, double-blinded study comparing the efficacy and safety of repeated biweekly infusions of avalglucosidase alfa (neoGAA, GZ402666) to alglucosidase alfa in treatment-naïve patients with late onset Pompe disease

GZ402666-EFC14028

STATISTICIAN: [Name]

DATE OF ISSUE: 14-Apr-2020

Total number of pages: 85

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITION OF TERMS

6MWT:	six-minute walk test
ADA:	anti-drug antibody
AESI:	adverse event of special interest
ALT:	alanine aminotransferase
AST:	aspartate aminotransferase
ATC:	anatomic category
BMS:	biomedical system
BP:	bodily pain
EAIR:	exposure adjusted incidence rate
e-CRF:	electronic case report form
EF:	emotional functioning
EQ-5D-5L:	5-Level EuroQol in 5 dimensions
ETP:	extension treatment period
GH:	general health
GLI:	global lung initiative
GMFCS-E&R:	gross motor function classification system-expanded and revised
GMFM-88:	gross motor function measure-88
GSD:	glycogen storage disease
GSGC:	gait, stair, gower's maneuver, and chair
HHD:	hand held dynamometry
HLGT:	high-level group term
HLT:	high-level term
IAR:	infusion associated reaction
IMP:	investigational medicinal product
IRT:	item response theory
LLT:	lower-level term
MAR:	missing at random
MCS:	mental component summary
MedDRA:	medical dictionary for regulatory activity
MEP:	maximum expiratory pressure
MH:	mental health
MIP:	maximum inspiratory pressure
mITT:	modified intent-to-treat
MMRM:	mixed model repeated measure
MRD:	minimal required dilution
NYHA:	the New York Heart Association
PAP:	primary analysis period
PCS:	physical component summary
PCSA:	potentially clinically significant abnormality
PDIS:	Pompe disease impact scale
PDSS:	Pompe disease symptom scale

PF:	physical functioning
PFT:	pulmonary function test
PGIC:	patient global impression of change
PK:	pharmacokinetics
PP:	per protocol
PR:	interval from the beginning of the P wave until the beginning of the QRS complex
PT:	preferred term
QMFT:	quick motor function test
QRS:	interval from start of the Q wave to the end of the S wave
QT:	interval between the start of the Q wave to the end of the T wave
QTc:	QT interval corrected for heart rate
RE:	role-emotional
RP:	role-physical
R-PAct:	Rasch-built Pompe-specific activity scale
SAE:	serious adverse event
Sch F:	school functioning
SD:	standard deviation
SF:	social functioning
SF-12:	12-item short form health survey
SMQ:	standard MedDRA query
SOC:	system organ class
ULN:	upper limit of normal
VT:	vitality
WHO-DD:	World Health Organization-Drug Dictionary
WMW:	wilcoxon-mann-whitney

1 OVERVIEW AND INVESTIGATIONAL PLAN

This document includes the details of the statistical analyses planned for the data during both the blinded treatment phase and the open-label extension phase.

1.1 STUDY DESIGN AND RANDOMIZATION

EFC14028 is a Phase 3, multicenter, multinational, randomized, double-blind study comparing the efficacy and safety of avalglucosidase alfa and alglucosidase alfa (both at 20 mg/kg qow) in treatment-naïve patients with LOPD Ages 3 and above.

Approximately 96 patients will be randomized to receive either avalglucosidase alfa or alglucosidase alfa in a blinded manner. Randomization will be in a 1:1 ratio with stratification factors based on baseline FVC, gender, age (<18 versus ≥18), and region (Japan versus Ex-Japan) among patients of age ≥18. Randomization will be performed within each of the following 6 strata:

1. Age <18,
2. Age ≥18, all genders and FVC (% predicted), Japan,
3. Age ≥18, male and FVC (% predicted) <55%, ex-Japan,
4. Age ≥18, female and FVC (% predicted) <55%, ex-Japan,
5. Age ≥18, male and FVC (% predicted) ≥55%, ex-Japan,
6. Age ≥18, female and FVC (% predicted) ≥55%, ex-Japan.

The majority of enrolled patients are expected to be ≥18 years of age and ex-Japan. For regulatory reasons, patients ≥18 years of age from Japan are included as a separate stratum to ensure that at least one such patient is exposed to avalglucosidase alfa. The other strata (age, FVC % predicted, and gender) were chosen to balance the baseline demographics and characteristics of patients thought to influence treatment outcomes in the Phase 3 AGLU02704 (LOTS) study of Myozyme®/Lumizyme® (alglucosidase alfa). Few patients <18 years of age are expected to be enrolled and so additional strata were not deemed necessary.

If at the end of the recruitment, less than 4 pediatric patients aged 3 to < 18 years are enrolled, up to 2 additional pediatric patients will be enrolled directly in the open-label avalglucosidase alfa long term follow-up phase where they will receive avalglucosidase alfa.

Patients will be observed in a 49 week blinded treatment phase, in which patients will receive an IV infusion of 20 mg/kg avalglucosidase alfa or alglucosidase alfa qow, followed by an open-label treatment extension phase, in which all patients will receive an IV infusion of 20 mg/kg avalglucosidase alfa qow for up to 96 weeks week. The 49 week blinded treatment phase will be considered the primary analysis period (PAP).

An extended open-label avalglucosidase alfa long term follow-up period will last up to 144 additional weeks (or up to avalglucosidase alfa approval in the country of the patient,

whichever comes first) for all patients (extended open label avalglucosidase alfa long term follow up Phase).

The primary analysis will be performed when all patients have completed the PAP. The release of the randomization list will be performed after the approval of the final Statistical Analysis Plan (SAP) and lock of the database.

1.2 OBJECTIVES

1.2.1 Primary objectives

The primary objective of the study is to determine the effect of avalglucosidase alfa treatment on respiratory muscle strength as measured by FVC % predicted in the upright position, as compared to alglucosidase alfa.

Estimand for this study is defined as the difference between avalglucosidase alfa treatment and alglucosidase alfa in mean FVC % predicted change from baseline to Week 49 regardless of whether intercurrent events have occurred. This estimand corresponds to a ‘treatment policy strategy’.

Analysis method for estimand (estimator) is based on mixed model for repeated measures (MMRM) assuming missing at random (MAR). All observed data during the study that include the data collected after study treatment discontinuation will be used in the analyses.

1.2.2 Secondary objectives

The key secondary objective is to determine effect of avalglucosidase alfa treatment on functional endurance as measured by the 6MWT. Additional secondary objectives are to determine the safety and effect of avalglucosidase alfa treatment on muscle inspiratory strength (MIP), muscle expiratory strength (MEP), lower extremity muscle strength (by hand-held dynamometry [HHD]), quick motor function test (QMFT), and health-related quality of life (the 12-item short form health survey [SF-12]).

1.2.3 Other objectives

Other objectives are to determine the pharmacokinetics (PK), exploratory pharmacodynamics, pharmacogenetics and effect of avalglucosidase alfa treatment on motor function (gross motor function measures [GMFM-88] and gait, stair, gower’s maneuver, and chair (GSGC), upper extremity muscle strength (by HHD), health-related quality of life (5-Level EuroQol in 5 dimensions [EQ-5D-5L] and PedsQL Generic Core Scale), patient reported outcomes (Pompe Disease Symptom Scale [PDSS]/Pompe Disease Impact Scale [PDIS] and Rasch-built Pompe-specific activity scale [R-PAct]), and patient global impression of change (PGIC).

1.3 DETERMINATION OF SAMPLE SIZE

Sample size calculations are based on non-inferiority test of the primary efficacy endpoint of change from baseline to Week 49 in FVC (% predicted) upright position, with the following assumptions:

Primary endpoint is normally distributed with a common standard deviation (SD) of 5.1% predicted, which is estimated based on the data from previous Phase 3, placebo-controlled LOTS trial (AGLU02704).

Mean treatment difference (avalglucosidase alfa–alglucosidase alfa) of 2.0% predicted, assumed based on a conservative estimate when comparing studies AGLU02704 and TDR12857 (Section 1.3.2).

A 2-sided 5% significance level,

Expected percent of missing data of up to 10% (estimated based on studies AGLU02704 and EMBASSY),

A non-inferiority margin of 1.1% predicted, which is based on the estimated alglucosidase alfa effect from the placebo-controlled study AGLU02704 (Section 1.3.1).

A total sample size of 96 (1:1 randomization ratio) will provide approximately 80% power to demonstrate non-inferiority of avalglucosidase alfa versus alglucosidase alfa, when the true treatment difference (avalglucosidase alfa–alglucosidase alfa) is 2.0% predicted. If the non-inferiority criterion is met, a test for superiority will be performed. If the true difference between avalglucosidase alfa and alglucosidase alfa is 3.6% predicted (see Table 3), the study will have more than 85% power to demonstrate superiority of avalglucosidase alfa to alglucosidase alfa.

Calculations were made using East 6.3 software.

1.3.1 Determination of the non-inferiority margin for the primary efficacy endpoint

Three studies of alglucosidase alfa/avalglucosidase alfa were identified with treatment-naïve late onset Pompe patients (LOPD) as the intended study indication (Table 1).

Table 1 - LOPD treatment naïve studies

Study	Study Design	Main results (FVC% predicted upright)	Key inclusion criteria
LOTS (AGLU02704) (Study 1)	Randomized, double-blind, Myozyme (20 mg/kg qow) vs. placebo, N = 90	Mean chg from BL at Week 52 was 1.651 (Myozyme) and -1.865 for placebo, with a LS mean difference of 3.5.	-Age ≥8 years -6MWT ≥40 meters -FVC (%) upright between 30%-80% -GAA enzyme deficiency and 2 GAA gene mutations
EMBASSY (Study 2)	Single arm study at Myozyme 20 mg/kg qow, N = 16	Mean chg from BL at Week 26 was 1.8	-Age ≥18 -6MWT ≥50 meters -FVC (%) upright ≥50% -GAA deficiency and.or confirmed GAA gene mutation and no known cardiac hypertrophy
TDR12857 (Study 3)	Open label, multi-dose cohorts with avalglucosidase alfa at 5, 10, 20 mg/kg qow	Group 1 (naïve LOPD) 10 + 20mg/kg cohorts combined: Mean chg from BL at Week 25 was 5.21 Subgroup with baseline FVC <85% (N = 5): Mean chg from BL at Week 25 was 4.39 (10 + 20mg/kg), or 6.15 (20 mg/kg only)	-Age ≥18 -6MWT ≥50 meters -FVC (%) upright ≥50% -GAA deficiency and.or confirmed GAA gene mutation and no known cardiac hypertrophy Group 1: LOPD treatment naive Group 2: LOPD switch from Myozyme

The LOTS study was the only randomized, placebo-controlled study with similar inclusion/exclusion criteria as the intended Phase 3 study and was used to determine the non-inferiority margin.

The estimated alglucosidase alfa treatment effect as compared to placebo from the LOTS study is presented in Table 2 below. The change from baseline in FVC (% predicted) was modeled using a mixed model repeated measure (MMRM) approach including the randomization strata (by FVC and 6MWT cutoffs), age, gender, treatment, visit, and treatment-by-visit interaction as fixed effects. An unstructured covariance matrix shared across treatment groups was used to model the within-subject errors. The Kenward-Roger approximation was used to estimate the degrees of freedom. The treatment effects of alglucosidase alfa vs placebo at each specified time point were estimated using the least-square means.

Table 2 - Estimated alglucosidase alfa treatment effect (FVC % predicted) from LOTS data

Population	Timepoint	LS mean difference	Estimated SD	80% CI for the difference
All patients				
	9 months (38 weeks)	3.77	4.74	(2.36, 5.18)
	12 months (52 weeks)	3.64	4.98	(2.14, 5.15)
	18 months (78 weeks)	3.73	5.35	(2.09, 5.37)

Based on all randomized patients from the LOTS study, the difference between alglucosidase alfa and placebo in change of FVC (% predicted) from baseline at Month 12 (52 weeks) was 3.64 with an 80% confidence interval of (2.14-5.15).

A non-inferiority margin of 1.1 was selected for this study with the following rationale

- A NI margin of 1.1 retained approximately 50% of the treatment effect based on lower bound of the 80% confidence interval.
- A recent literature review on the clinical relevance of outcome measures used in LOPD revealed that in 2/3 of the studies in which LOPD patients were treated with alglucosidase alfa, the changes from baseline in % FVC (predicted) were above or within the minimal clinically important difference (MCID) established for another restrictive respiratory disease, IPF, of 2% to 6% (1). The current proposed NI margin of 1.1 is smaller than this reference range.

Although the NI margin was determined based on the LOTS study with FVC % predicted inclusion criterion of 30-80%, a subgroup analysis conducted in LOTS (CSR in-text Table 11-9) suggested that the estimated treatment difference between alglucosidase alfa and placebo (based on last observed data within 18 months) was larger among those patients with baseline FVC (% predicted) ≥55%, comparing to those patients with baseline FVC (% predicted) <55% (4.74% versus 2.25%). Therefore, including a small percent of patients (capped at 15%) with slightly higher baseline FVC (80-85%) in the current study is not expected to compromise the estimated alglucosidase alfa treatment effect, and the proposed NI margin.

1.3.2 Estimated treatment differences between avalglucosidase alfa and alglucosidase alfa on change from baseline in FVC (% predicted upright) at Month 12 (Week 49)

The estimated treatment difference between avalglucosidase alfa and alglucosidase alfa is limited by the following factors:

- Small number of patients treated with avalglucosidase alfa in TDR12857 study, with 6 months post-baseline assessments,
- No randomized comparison data available between avalglucosidase alfa and alglucosidase alfa,
- Difference in inclusion/exclusion criteria between studies TDR12857 and EFC14028.

Table 3 provides the estimated treatment difference between avalglucosidase alfa and alglucosidase alfa in various subpopulations. Assuming that the difference between avalglucosidase alfa and alglucosidase alfa does not change between Month 6 (Week 26) and Week 49 for the primary endpoint of FVC (% predicted) upright, our estimated potential treatment difference (avalglucosidase alfa-alglucosidase alfa) is predicted to range from 2% to 3.6% based on the lower limit of 80% confidence interval and the point estimate of difference when comparing combined Study 1 and 2 versus Study 3 patients. Therefore, we have used 2.0 as the assumed true treatment effect for power calculation of non-inferiority test.

Table 3 - Point estimates and confidence intervals of change in FVC % predicted (upright) at 6 months comparing avalglucosidase alfa vs. alglucosidase alfa (Mz) from previous studies

Treatment	N	Mean change from BL in FVC % predicted (upright)	Mean difference (80% CI)*
avalglucosidase alfa from Study 3 (10 mg/kg or 20 mg/kg dose level) vs. all Mz patients from Study 1 and 2			
avalglucosidase alfa	6	5.2	3.6, CI: (2.03, 6.98)
alglucosidase alfa	72	1.6	
avalglucosidase alfa from Study 3 20mg/kg patients vs. all Mz patients from Study 1 and 2			
avalglucosidase alfa	3	6.2	4.6, CI: (2.03, 8.33)
alglucosidase alfa	72	1.6	

*Confidence interval for the difference was based on Hodges-Lehmann exact confidence interval due to small sample size in the avalglucosidase alfa arm.

1.3.3 Determination of the dropout rate

Based on the LOTS study, 57 out of 60 (95%) alglucosidase alfa treated patients remained in the study at Month 12 with evaluable FVC data and 54 out of 60 (90%) patients remained on the study at Month 18. A separate uncontrolled study (EMBASSY) with a similar patient population enrolled a total of 16 patients for planned 6 months follow-up. Among those, 1 patient (6.3%) had missing data at 6 months. To allow potential variability between trials and to retain a reasonable power for the per-protocol population analysis, a 10% dropout rate was assumed which corresponds to a total sample size of approximately 96 patients.

1.3.4 Power calculations with different effect size assumptions

The power calculations for additional assumptions of effect size are presented in Table 4.

Table 4 - Power for non-inferiority test and superiority test under different effect size and standard deviations

Standard Deviation assumption (based on LOTS patients at Month 12)	Treatment difference	Power for NI test (NI margin = 1.1)	Power for superiority test
5.0 (estimated from the MMRM model with all patients data)			
	1.5	66.4%	27.9%
	2.0	81.1%	45.0%
	2.5	91.0%	63.0%
	3.0	96.4%	78.5%
	3.5	98.8%	89.4%
5.1 (estimated from within arm SD from evaluable alglucosidase alfa patients only)			
	1.5	64.7%	27.0%
	2.0	79.6%	43.5%
	2.5	89.9%	61.3%
	3.0	95.8%	76.9%
	3.5	98.5%	88.2%

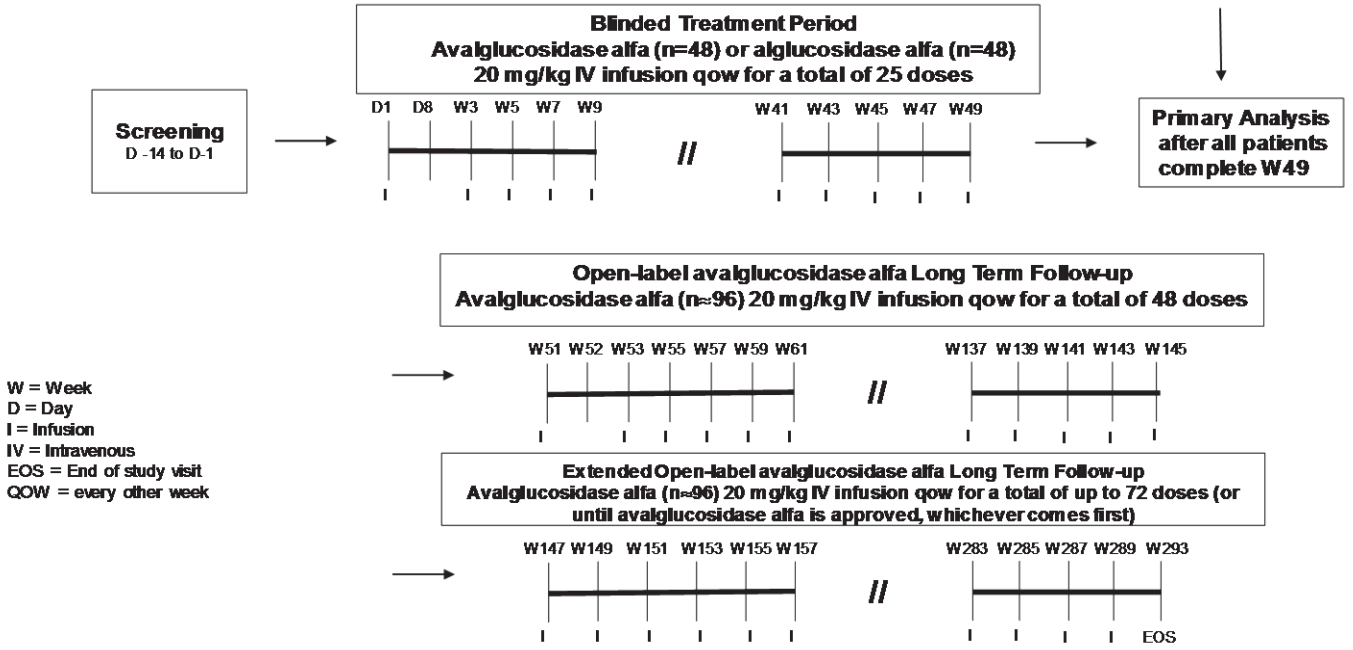
1.4 STUDY PLAN

For each patient who signs the informed consent, the study consists of:

- Screening/baseline visit: occurs within 14 days prior to randomization,
- Blinded treatment period (PAP), except for the subgroup of patients aged 3 to < 18 years enrolling directly in the open-label long term follow-up phase: a 49-week blinded treatment phase, in which patients will receive an IV infusion of 20 mg/kg avalglucosidase alfa or alglucosidase alfa qow. This phase of the study will be referred as PAP in this document.
- Open-label avalglucosidase alfa long term treatment period: an up to 96-week open label avalglucosidase alfa long term follow-up. All patients will receive an IV infusion of 20 mg/kg avalglucosidase alfa qow for long-term follow-up.
- An extended open-label avalglucosidase alfa long term follow-up period will last up to 144 additional weeks (or up to avalglucosidase alfa approval in the country of the patient, whichever comes first) for all patients (additional open label avalglucosidase alfa long term follow up Phase).
- Both open-label period and extended open-label period will be referred to as the Extension Treatment Period (ETP) in this document.
- An up to 4 week post treatment observation period.

Study flowchart is described in Figure 1 below. The schedule of efficacy assessments is described in Section 7.

Figure 1 – General 2Study Design



1.5 MODIFICATIONS TO THE STATISTICAL SECTION OF THE PROTOCOL

The current SAP is based on protocol amendment 3 (dated 10 Apr 2019). This section summarizes major changes to the protocol statistical section with emphasis on changes after study start (after the first patient was enrolled). The protocol history table below gives the timing, rationale and key details of major changes to the protocol statistical section.

Protocol amendment statistical changes

Amendment Number	Date Approved	Description of Statistical Changes	Brief Rationale
3	10-Apr-2019	Updated definition of Safety Population to include up to 2 pediatric patients who will be enrolled directly in the open-label avalglucosidase alfa long term follow-up phase	In order to comply with Health Authority requirements to enroll a certain number of pediatric patients, up to 2 additional pediatric patients will be enrolled directly in the open-label avalglucosidase alfa long-term follow-up phase where they will receive avalglucosidase alfa
3	10-Apr-2019	Added following superiority test to the hierarchical testing '% predicted change from baseline to Week 49 for MEP'.	We will have chance to make claim in the label for the secondary endpoint MEP
3	10-Apr-2019	Removed non-inferiority test of 6MWT in the fixed sequence testing to control multiplicity.	Updated in accordance with the feedback received from Regulatory Agency.

1.6 STATISTICAL MODIFICATIONS MADE IN THE STATISTICAL ANALYSIS PLAN

The following main modifications have been made in this version of SAP

- Placebo-based imputation is removed since they may bias towards equivalence and thus is not conservative for NI trial.
- Added potential sensitivity analyses with respect to the constancy assumption based on covariate-adjustment fixed margin approach and/or covariate-adjustment synthesis method as proposed by Nie and Soon (2) based on feedback from health authority.
- Added the concept of estimand based on feedback from health authority.
- Additional results on Type I error rates and power were added in Section 2.5.1 in response to FDA question/request for information item no.1 dated 02 September 2016.
- The SAP is revised to cover statistical analyses in both PAP and ETP.

- In response to FDA question/request for information item no.2 dated 02 September 2016:
  - Background data on determination of dropout rate were added in Section 1.3.3
  - Section 2.5 was revised to add the sentence that "If the pure ITT (all randomized patients) population is different from the mITT population, we plan to perform a sensitivity analysis in this population as well to assess the robustness of the results."
- The Appendix on Potentially Clinically Significant Abnormalities Criteria was updated to correct a typo on children diastolic blood pressure criteria, i.e., increase/decrease from baseline should be ≥10mmHg instead of ≥20mmHg.
- In response to FDA question/request for information dated 19 July 2019:
  - Added Global Lung Initiative (GLI) 2012 reference equations to calculate FVC % predicted values
  - Added new threshold of 27.5 meters to define the 6MWT responder in addition to the current 3 thresholds.
- In response to FDA question/request for information dated 14 January 2020:
  - Proposed the treatment policy estimand to assess the treatment effect regardless of occurrence of intercurrent events. All collected efficacy data will be used in analysis to be aligned to the treatment policy estimand.
- In response to meeting discussion with FDA on 20 February 2020,
  - Added a within patient comparison in FVC (% predicted) between Week 49 and Week 97 for patients who were initially randomized to Alglucosidase alfa in PAP and then crossed over to Avalglucosidase alfa in ETP
  - Updated immunogenicity analysis. Some revisions of immunogenicity analysis are made to be consistent with integrated immunogenicity analysis plan.
- In response to FDA advice/request for information dated 13 April 2020:
  - Added an analysis of covariance (including baseline FVC, age, gender, treatment group as covariates) for the endpoint of change from baseline in FVC (% predicted) at Week 49 or the last visit prior to initiation of an alternative treatment prior to Week 49. For those patients who prematurely discontinued prior to Week 49 but with unknown alternative treatment information, their last value prior to dropout will be used in the analysis.
  - Added analysis to assess the within-subject changes from baseline in PD marker with respect to the changes in the status and/or titer value of ADA/NAb (neutralizing antibodies).

2 STATISTICAL AND ANALYTICAL PROCEDURES

The PAP analysis will be performed after all patients have completed the 12-month (49 week) blinded treatment phase and will include a formal database lock and unblinding of the study treatment arms, and analyses of efficacy, safety, pharmacodynamic, and pharmacokinetic data. A CSR will be prepared after the PAP analysis.

After the primary analysis for PAP is completed, interim analyses may be performed during the open-label extension period to provide additional information for regulatory purpose. those additional analyses after PAP will be considered supportive purpose.

The final ETP analyses will be performed after all patients completed ETP. A final database lock will occur, and a final CSR will be prepared after the ETP analyses.

2.1 ANALYSIS ENDPOINTS

2.1.1 Demographic and baseline characteristics

The baseline value is defined as the last non-missing value prior to first treatment, unless otherwise specified. In ETP analyses, when needed, the ETP baseline value is defined as the last non-missing value prior to first infusion in ETP.

Demographic characteristics

Demographic variables are gender (Male, Female), race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Island, White, Not reported), regions, weight, height, BMI, head circumference, age at study entry (informed consent), and randomization strata.

Medical/Surgical history

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

Disease characteristics at baseline

Pompe disease history includes age at Pompe disease diagnosis, age at first symptoms of Pompe disease, and time from these events to 1st infusion of study drug.

Pompe medical history includes,

- Cardiovascular history (congestive heart failure, current NYHA (the New York Heart Association) heart failure classification, evidence of cardiac involvement),
- Ear, nose, throat medical history (enlarged tongue, hearing loss, hepatomegaly, gastroesophageal reflux),
- Respiratory history (tracheostomy history, pneumonia, sleep disturbances, sleep apnea, use of non-invasive respiratory support),

- Musculoskeletal history (muscle weakness, scoliosis, joint contractures, and current ambulatory status, assisted walking device and wheelchair use).
- Family history of Pompe disease history includes confirmed Pompe disease in other family members.

Vital signs

Vital signs at baseline are heart rate, systolic and diastolic blood pressure, respiratory rate, temperature, and oxygen saturation. The baseline values for vital signs are the last non-missing assessment prior to the first study treatment by comparing the infusion time and measurement time.

Electrocardiogram variables

The baseline ECG is defined as the average of triplicate values at the baseline visit and prior to the first infusion of the study drug.

Immunogenicity status

The immunogenicity evaluation at baseline includes antibody status (positive or negative) and magnitude (titer) of the anti-drug antibodies (ADA).

2.1.2 Prior or concomitant medications

Medications and therapies taken by the patient during the 30-day period prior to the Screening/Baseline evaluation visit and during the course of the study will be recorded in the electronic Case Report Forms (e-CRF).

All medications will be coded using the World Health Organization-Drug Dictionary (WHO-DD) using the version in effect at Sanofi at the time of database lock.

- Prior medications are those the patient used prior to first investigational medicinal product (IMP) intake. 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 IMP, from 1st study treatment to the end of treatment + 30 days. A given medication can be classified both as a prior medication and as a concomitant medication. Concomitant medications do not include medications started during the post-treatment period (as defined in the observation period in Section 2.1.4).
- Post-treatment medications are those the patient took in the period running from the 30 days after last IMP intake up to the end of the study.

The use (if any) of mechanical ventilation (including both invasive and noninvasive), as well as assistive device use will be recorded on the e-CRF.

Any technical details related to computation, dates, imputation for missing dates are described in Section 2.10.

2.1.3 Efficacy endpoints

The baseline value for efficacy endpoints is the last non-missing assessment prior to study treatment and no later than the randomization date. The ETP baseline value for efficacy endpoints is the last non-missing assessment prior to the first study treatment in ETP.

2.1.3.1 Primary efficacy endpoint

The primary efficacy endpoint is the change from baseline to Week 49 in FVC (% predicted) in the upright position. FVC will be reported in absolute value in liters, as well as the percent of predicted normal values based on Global Lung Initiative (GLI) 2012 reference equations (3). The FVC percent of predicted values will be calculated based on FVC in liters, gender, race (classified as Caucasian, Asian and African-American, and Other/Mixed), age (at least one decimal place in years), and height (in cm) at the time of each assessment and will be reported centrally from Biomedical Systems (BMS), following pulmonary software specification and user requirement (4).

The FVC percent predicted value is calculated as:

(actual FVC measurement/predicted value of FVC) * 100.

The GLI-2012 regression equations and lookup tables are used to calculate predicted values of FVC (1). FVC is predicted according to the following equation in the EFC14028 trial:

$M = \exp(a0 + a1 \cdot \ln(\text{Height}) + a2 \cdot \ln(\text{Age}) + a3 \cdot \text{black} + a4 \cdot \text{NEAsia} + a5 \cdot \text{SEAsia} + a6 \cdot \text{Other} + \text{Mspline})$

where

Black = 1 if a subject is African American, otherwise = 0

NorthEastAsia = 1 if a subject is from North East Asia, otherwise = 0

SouthEastAsia = 1 if a subject is from South East Asia, otherwise = 0

Other = 1 if subject is ‘other ethnic group’ or mixed ethnicity, otherwise = 0

Coefficients a(n) depend on sex and are given by lookup table

Mspline is age-varying coefficients, given by lookup table for each type of sex

For the analysis purpose, the age will be calculated based on (assessment date - birth date + 1)/365.25. Height will be based on the most recent valid value at or prior to the assessment date.

Since baseline FVC (% predicted) is used for randomization stratification purposes, no patient should have a missing baseline value. The schedule of assessments is provided in Section 7.

2.1.3.2 Secondary efficacy endpoint(s)

The key secondary efficacy endpoint for the study is the total distance (meters) walked during six-minute walk test (6MWT). The additional secondary efficacy endpoints are:

- Pulmonary function testing: maximum expiratory pressure (MEP) and percent predicted, maximum inspiratory pressure (MIP) and percent predicted,
- Hand-held dynamometry (lower extremity strength): lower extremity summary score and individual muscle group and percent predicted from the muscle groups of Hip Flexion, Hip Extension, Hip Abduction, Hip Adduction, Knee Flexion, Knee Extension, Ankle Dorsiflexion, Ankle Plantar Flexion,
- Quick motor function test: total score of the QMFT,
- 12-item short-form health survey: the physical component summary (PCS) and mental component summary (MCS) scales from the survey.

The planned schedule of assessments for secondary efficacy endpoints is provided in Section 7.

Six-minute walk test

The 6MWT (5) will be performed to assess functional capacity in the late-onset Pompe disease population. The primary outcome measure of the 6MWT is the distance walked in 6 minutes recorded in meters. Test equipment and administration techniques will be standardized among the investigational sites. The percent predicted value will be calculated based on the normal reference equations from (6) and (7), which cover the age range in this study. A similar approach was used by Montes et al (8). For analysis purposes, the age at each assessment will be calculated based on (assessment date - birth date + 1)/365.25. Height (which is to be assessed annually for patients with age ≥18 years and every 3 months for patients with age <18 years) will be based on the most recent valid value at or prior to the assessment date.

Table 5 - Equations for calculating reference value for % predicted total distance walked in 6MWT

Reference	Age at baseline	Gender	Equation
(6)	≥18 years	Male and Female	868.8 - 2.99 * age - 74.7*sex
(7)	<18 years	Male	196.72 + 39.81 * age – 1.36 * age^2 + 132.28 * height
	<18 years	Female	188.61 + 51.50 * age – 1.86 * age^2 + 86.10 * height

Age in years; sex = 0 if male and sex = 1 if female; height in meters.

Additional information, including the amount of time walked (to quantify endurance, as all patients may not complete the full 6-minute walk), as well as the use and change of assistive device will also be recorded.

Pulmonary function testing

Secondary efficacy endpoints measured by pulmonary function tests (PFT) will include the assessment of MEP and MIP in the upright position. Additional parameters of respiratory function, including assessment of FVC, MIP, and MEP in the supine position. FEV1, and PEF in the upright and supine positions will be considered exploratory or tertiary endpoints.

All PFTs will be reported centrally from BMS following pulmonary software specification and user requirement (4).

The percent predicted values for MIP and MEP will be calculated (9, 10). The normal reference values used for % predicted calculation will be dependent on each patient's age, gender, and weight (Table 6). For each patient, the same equations used for the baseline will be used to calculate the predicted values (eg, if a patient started at ≤17 years of age and turns 18 during the study, he/she will still use the pediatric formula after turning to 18). For analysis purposes, the age will be calculated based on (assessment date - birth date + 1)/365.25. Weight for age <18 will be based on the most recent valid value at or prior to the assessment date. The percent (%) predicted value is defined as (absolute measurement/normal reference value) * 100.

Since the reference values will be available only for patients of age 7 years or older, the % predicted MIP and MEP will not be derived for patients aged <7.

Table 6 - Reference equations for predicted value of MEP and MIP

Test/Age at baseline	Male	Female
MEP		
≥18 years old (9)	174 – 0.83 * age	131 – 0.86 * age
7–17 years old (10)	35 + 5.5 * age	24 + 4.8 * age
MIP		
≥18 years old (9)	120 – 0.41 * age	108 – 0.61 * age
7–17 years old (10)	44.5 + 0.75 * weight	40 + 0.57 * weight

Age in years; weight in kg.

Hand-held dynamometry (lower extremity muscle strength)

Equipment for hand-held dynamometry (11) using the make technique will be standardized across sites. During the treatment period, the assessment will be completed before the IMP infusion. Stabilization procedures for all muscles groups will be followed to avoid use of compensatory muscles. To complete a make test, the examiner holds the dynamometer stationary while the patient exerts a maximal force against the dynamometer. The patient makes a gradual increase in force and then completes an isometric hold for 4-5 seconds.

Lower extremity strength in the muscle groups detailed below will be tested by the same physical therapist or trained assessor.

- Hip Flexion,
- Hip Extension,
- Hip Abduction,
- Hip Adduction,
- Knee Flexion,
- Knee Extension,

- Ankle Dorsiflexion,
- Ankle Plantar Flexion.

The limb tests will be completed bilaterally to account for differences in the generated force for the dominant and non-dominant limb. Every muscle group will be measured two times and the highest value will be reported in the e-CRF. Patients may use noninvasive ventilation during the dynamometry assessment.

Measurements from selected muscle groups will be summed to generate the summary score of the lower extremity muscle strength and lower extremity strength percent predicted. Specifically, no healthy predicted values are available for ankle plantar flexion and hip adduction thus the following 6 muscle groups (both left and right side) are used in calculating the summary score.

- Hip Flexion,
- Hip Extension,
- Hip Abduction,
- Knee Flexion,
- Knee Extension,
- Ankle Dorsiflexion

The reference equations to determine healthy predicted results for each muscle group are provided in Table 7, Table 8 and Table 9. For the patients with age >16 at baseline, the predicted muscle strength will be calculated using the adult reference equations in Table 7. For the patients with baseline age between 3 and 16 years old, the predicted muscle strength will be calculated using the pediatric reference questions in Table 8 and Table 9. For pediatric patients, the same reference equations will be used to calculate the predicted results from both left and right sides within each muscle group. The predicted muscle strength normal reference is only available for patients of 6 years or older for the majority of muscle groups from existing publications.

Table 7 - Adult (age >16) muscle strength reference equations to determine healthy predicted results in Newton (N)

Muscle Group	Test side	Equations*
Hip Flexion (12)	R	9.8067 [-(age×.33) + (gender×19.19)+((weight/height^2) × .66) + 34.44]
Hip Flexion (12)	L	9.8067 [-(age×.29) + (gender×18.75)+((weight/height^2) × .47) + 36.05]
Hip Extension (12)	R	9.8067 [-(age×.21)+(gender×15.19)+((weight/height^2) × .14) + 33.52]
Hip Extension (12)	L	9.8067 [-(age×.23)+(gender×15.02)+((weight/height^2) × .17) + 33.88]
Hip Abduction (13)	R & L	195.24 – 62.4(1-gender) – 1.184(age) + .198 (weight x 9.8067)
Knee Flexion (12)	R	9.8067 [-(age×.16)+(gender×8.78)+((weight/height^2) × .08) + 22.47]
Knee Flexion (12)	L	9.8067 [-(age×.17) + (gender×7.67) + ((weight/height^2) × .14) + 21.10]
Knee Extension (12)	R	9.8067 [-(age×.38) + (gender×18.44) + ((weight/height^2) × .62) + 34.41]
Knee Extension (12)	L	9.8067 [-(age×.38) + (gender×17.68) + ((weight/height^2) × .62) + 33.61]
Ankle Dorsiflexion (12)	R	9.8067 [-(age×.12) + (gender×7.93) + ((weight/height^2) × .11) + 22.87]
Ankle Dorsiflexion (12)	L	9.8067 [-(age×.13) + (gender×8.12) + ((weight/height^2) × .15) + 21.91]

* Age in years; gender: male=1, female=0; weight in kgs; height in meters.

Table 8 - Pediatric muscle strength reference values in Newton (N) (14)

		Age												
Muscle Group	Gender	4	5	6	7	8	9	10	11	12	13	14	15	16
Hip flexors	Boy			182	182	225	232	261	245	198	289	337	301	395
	Girl			162	184	175	195	177	264	232	308	281	288	301
Hip abductors	Boy			128	124	131	153	174	151	158	225	306	356	312
	Girl			109	122	117	124	104	140	171	227	244	257	244
Knee extensors	Boy			156	157	185	194	267	239	225	296	370	362	396
	Girl			148	177	166	173	198	265	250	346	280	325	373
Knee flexors	Boy	111	105	158	180	185	195	268	218	201	273	307	327	382
	Girl	92	99	154	171	160	180	175	246	221	301	271	282	336
Foot dorsal flexors	Boy	71	76	104	130	137	141	154	149	170	218	257	267	291
	Girl	75	76	95	114	121	137	130	178	177	214	207	220	232

Table 9 - Pediatric muscle strength reference values in Newton (N) (15)

		Age Groups					
Muscle Group	Gender	3.5-5.5	5.5-7.5	7.5-9.5	9.5-11.5	11.5-13.5	13.5-15
Hip extensors	Boy	80	131	170	210	226	277
	Girl	76	124	157	174	226	267

If a patient is not covered by one of the reference equations, specifically 15-16 years old for hip extension, the reference equation for the oldest age group available (13.5-15 years) will be used.

The summary score of the lower extremity strength will be the sum of the 12 measurements from the 6 muscle groups (left and right measurements from each muscle group). The percent predicted muscle strength will be calculated separately for the 12 measurements from the 6 muscle groups. The percent predicted value of the total score for the lower extremity strength will be the average of the 12 percent predicted values from the 6 muscle groups. If any of the 12 measurements from the 6 muscle groups in lower extremity strength is missing, the summary score of the lower extremity strength is missing. In that case, the percent predicted of the summary score is also missing.

- ***Quick motor function test***
- The QMFT is an observer administered test comprising 16 items specifically difficult for patients with Pompe disease (16). The items are scored separately on a 5-point ordinal scale (ranging from 0 to 4) with a total score of all items ranging between 0 and 64 points with higher score representing better outcome. If any of the 16 items is missing, the total score of the QMFT will be considered as missing.
- ***12-Item short-form health survey***

The SF-12 is comprised of a subset of 12 items from the SF-36 to reproduce the PCS and MCS scales (17). The SF-12 will be administered to assess health-related quality of life in patients 18 years or older at screening/baseline.

More specifically, SF-12 consists of 4 2-item health domain scales, including, physical functioning (PF), role-physical (RP), role-emotional (RE) and mental health (MH), as well as 4 1-item health domain scales, including, bodily pain (BP), general health (GH), vitality (VT) and social functioning (SF). The PCS and MCS scores for a patient are weighted sums of the 12 item scores (18).

When a patient fails to complete all 12 items, a Maximum Data Recovery method is used to first estimate missing item scores and then to predict the PCS and MCS summary scores (18). Specifically, for the 4 2-item health domain scales (PF, RP, RE, MH), a model using the item response theory (IRT) is developed to estimate a missing value on an item based upon a responder's response to an answered item. If neither item is completed, the corresponding scale is missing. For the 4 1-item scales, if the single item is not completed, the corresponding scale is missing. If no data are missing after this estimation step, the PCS and MCS scores can be calculated via the standard weighted sums; otherwise, an established and validated prediction model may be used to estimate the PCS and MCS scores for those responders who have observed or estimated scores on at least 7 out of the 8 health domains. If more than one scale is missing then neither PCS nor MCS can be estimated. Moreover, the PF scale cannot be missing to calculate the PCS score, while the MH scale cannot be missing to calculate the MCS score. For analysis purpose, the scoring of SF-12 will be provided by QualityMetric Inc, based on Scoring Software Version 4.5 of the SF-12v2 Health Surveys.

2.1.3.3 Additional efficacy endpoints

Additional efficacy endpoints include endpoints from GSGC composite functional assessment, GMFM-88, hand-held dynamometry of upper extremity muscle groups (upper extremity muscle strength), 5-Level EuroQoL survey, and pediatric quality of life inventory.

Gait, Stair, Gower's Maneuver, and Chair composite functional assessment

Functional performance will be measured using the GSGC score. The GSGC total score can be obtained by adding the item scores of the four functional tests; item scores range from 1 to 7 for three items and 1-6 for one item (Arising from a chair). Briefly, total scores can vary from a minimum of 4 representing normal performance to a maximum of 27 representing the poorest functional score. The GSGC performance test has been validated for use with glycogen storage disease (GSD) type II patients and patients with Duchenne muscular dystrophy (19). Each of the functional items is also evaluated according to the time required to complete the task.

- Time to walk 10 meters (Gait),
- Time to climb 4 stairs (Stair),
- Time to stand from sitting on the floor (Gower's Maneuver),
- Time to stand from sitting position in a chair (Chair).

If a qualitative item is missing, total score will be missing. If timed item is missing, then this may be due to the item not being able to be completed (quality score of 6 or 7).

Gross Motor Function Measure-88 (GMFM-88)

The GMFM-88 (20) will be administered for all patients at the time points specified in the protocol to evaluate changes in motor function. Of the five dimensions to the GMFM-88, 2 dimensions (D and E) will be evaluated in this study:

- Standing (13 items),
- Walking, Running & Jumping (24 items).

Items were selected to represent motor functions typically performed by children without motor impairments by 5 years of age. Each item is scored on a 4-point Likert scale (ie, 0 = cannot do; 1 = initiates [<10% of the task]; 2 = partially completes [10% to <100% of the task]; 3 = task completion). The score for each dimension is expressed as a percentage of the maximum score for that dimension. Total score is obtained by adding the percentage scores for each dimension and dividing the sum by the total number of dimensions. Therefore, each dimension contributes equally to the total score. If an item is missing, the dimension score will be missing.

This assessment will include the expanded and revised Gross Motor Function Classification System for the GMFM-88 (gross motor function classification system-expanded and revised [GMFCS-E&R]) adapted for adults (21). The GMFCS emphasizes concepts in the World Health Organization's International Classification of Functioning, Disability and Health. Emphasis is on performance in home, work and community settings without judgments about quality of movement or prognosis for improvement. The GMFCS is a 5 level classification system consisting of Levels I to V based on self-initiated movement, with emphasis on sitting, transfers, and mobility. The distinctions between levels are based on functional limitations, the need for

assistive mobility devices, and to a much lesser extent, quality of movement, and are designed to be meaningful in daily life (21).

The general headings for the 5 levels (analyzed as 1, 2, 3, 4, 5) are:

- Level I Walks without limitations,
- Level II Walks with limitations,
- Level III Walks using a hand-held mobility device,
- Level IV Self-mobility with limitations; may use powered mobility,
- Level V Transported in a manual wheelchair.

Hand-held dynamometry (upper extremity muscle strength)

Upper extremity strength in the following muscle groups will be included:

- Shoulder Flexion,
- Shoulder Extension,
- Shoulder Abduction,
- Shoulder Adduction,
- Elbow Flexion,
- Elbow Extension,
- Grip Strength.

The limb tests will be completed bilaterally to account for differences in the generated force for the dominant and nondominant limb. Every muscle group will be measured two times and the highest value will be reported in the e-CRF. Patients may use noninvasive ventilation during the dynamometry assessment.

The summary score of the upper extremity strength will be the sum of the 14 measurements from the 7 muscle groups (left and right measurements from each muscle group). If any of the 14 measurements from the 7 muscle groups in extremity strength is missing, the summary score of the upper extremity strength is missing.

The percent predicted muscle strength will be calculated separately for muscle groups for which percent predicted equations are available; the percent predicted value of the total score for the upper extremity strength will be the average of the 12 percent predicted values from the 6 muscle groups. If any of the 12 measurements from the 6 muscle groups in percent predicted total score is missing, the percent predicted total score is not available.

No healthy predicted values are available for shoulder adduction for adults thus the following 6 muscle groups (both left and right side) are used in calculating the summary score.

- Shoulder Flexion,
- Shoulder Extension,
- Shoulder Abduction,
- Elbow Flexion,

- Elbow Extension,
- Grip Strength.

The reference equations to determine healthy predicted results for each muscle group are provided in Table 10, Table 11, and Table 12. For the patients with age >16 at baseline, the predicted muscle strength will be calculated using the adult reference equations in Table 7. For the patients with baseline age between 3 and 16 years old, the predicted muscle strength will be calculated using the pediatric reference questions in Table 11 and Table 12. For pediatric patients, the same reference equations will be used to calculate the predicted results from both left and right sides within each muscle group. For patients with baseline age between 3 and 16 years old, the predicted muscle strength normal reference is not available for shoulder flexion, shoulder extension, and shoulder adduction from existing publications; therefore, the summary score of the upper extremity muscle strength and supper extremity strength percent predicted will not be calculated.

Table 10 - ADULT HHD Reference Equations to Determine Healthy Predicted Result (N) (12) (13)

Muscle / Action (Reference)	Age=years; Gender: Male=1, Female=0; weight=kg, height=meters
Shoulder Flexion, Right, Left (1)	R: 9.8067 [-(age×.14)+(gender×11.22)+((weight/height^2) ×.22)+16.85] L: 9.8067 [-(age×.12)+(gender×10.68)+((weight/height^2) ×.24)+14.68]
Shoulder Extension, Right, Left (1)	R: 9.8067 [-(age×.17)+(gender×16.26)+((weight/height^2) ×.17)+23.35] L: 9.8067 [-(age×.18)+(gender×14.64)+((weight/height^2) ×.29)+19.59]
Elbow Flexion, Right, Left (1)	R: 9.8067 [-(age×.13)+(gender×11.24)+((weight/height^2) ×.07)+22.78] L: 9.8067 [-(age×.11)+(gender×10.63)+((weight/height^2) ×.05)+19.66]
Elbow Extension, Right, Left (1)	R: 9.8067 [-(age×.08)+(gender×8.33)+((weight/height^2) ×.16)+12.37] L: 9.8067 [-(age×.07)+(gender×8.18)+((weight/height^2) ×.17)+11.32]
Grip Strength, Right, Left (1)	R: [-(age×.18)+(gender×16.90)+((weight/height^2) ×.23)+31.33]/2.20 L: [-(age×.16)+(gender×16.68)+((weight/height^2) ×.29)+26.60]/2.20
Shoulder Abduction, Right, Left (2)	178.90 – 77.1(gender) – 1.128(age) + .134(9.8067xweight)

Table 11 - Pediatric muscle strength reference values in Newton (N) (14)

		Age												
Muscle Group	**Gender**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Shoulder Abductors	Boy	62	55	97	92	98	110	136	110	118	159	205	219	253
	Girl	68	47	75	91	94	91	81	129	123	154	178	173	173
Elbow Extensors	Boy			73	85	90	89	120	103	104	128	158	175	182
	Girl			73	85	82	91	84	108	117	118	129	141	107
Elbow Flexors	Boy	78	70	103	121	124	134	173	153	160	195	253	287	276
	Girl	69	66	105	103	115	125	134	172	168	201	193	198	215

Table 12 - Pediatric grip strength reference values (Mathiowetz, et al, 1986) (22)

Average Performance of Normal Subjects on Grip Strength (lb)

		Males			Females		
Age	Hand	Mean	*SD*	Range	Mean	*SD*	Range
6–7	R	32.5	4.8	21–42	28.6	4.4	20–39
	L	30.7	5.4	18–38	27.1	4.4	16–36
8–9	R	41.9	7.4	27–61	35.3	8.3	18–55
	L	39.0	9.3	19–63	33.0	6.9	16–49
10–11	R	53.9	9.7	35–79	49.7	8.1	37–82
	L	48.4	10.8	26–73	45.2	6.8	32–59
12–13	R	58.7	15.5	33–98	56.8	10.6	39–79
	L	55.4	16.9	22–107	50.9	11.9	25–76
14–15	R	77.3	15.4	49–108	58.1	12.3	30–93
	L	64.4	14.9	41–94	49.3	11.9	26–73
16–17	R	94–0	19.4	64–149	67.3	16.5	23–126
	L	78.5	19.1	41–123	56.9	14.0	23–87
18–19	R	108.0	24.6	64–172	71.6	12.3	46–90
	L	93.0	27.8	53–149	61.7	12.5	41–86

Note: The mean scores for individuals, aged 14 to 19 years, may be slightly low (0–10 lb lower than they should be) due to instrument error detected after the study.

5-Level EuroQol in 5 dimensions

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome (23). Applicable to a wide range of health conditions and treatments, it provides a simple descriptive profile and a single index value for health status. The EQ-5D-5L descriptive system comprises 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, anxiety/depression). Each dimension has 5 levels of health/ability (type of response could be no problem, slight problem, moderate problem, severe problem; or unable to perform activity or no, slight, moderate, severe, or extreme). The EQ VAS records the respondent's self-rated health status on a vertical graduated (0-100) visual analogue scale. A score of 100 represents the best imaginable health state and 0 refers to the worst imaginable health state.

Patients who are ≥18 years of age at screening/baseline will complete this assessment.

Pediatric quality of life inventory

The PedsQL Generic Core Scale (24, 25) is an instrument that was designed to measure HRQoL in healthy children, as well as those with acute and chronic health conditions, covering an age spectrum in the range of 2–18 years. During the treatment period, the patients will continue on the age-specific assessment they first completed at screening/baseline, even if they exceed the age range for that specific tool over the trial duration. The PedsQL comprises parallel child self-report (age 5–7 years [young child], 8–12 years [child] and 13–18 years [adolescent]) and parent proxy-report (age 2–4 years [toddler], 5–7 years [young child], 8–12 years [child] and 13 to 18 years [adolescent]) formats. The PedsQL generic core scales were specifically designed to measure the core dimensions of health as delineated by the World Health Organization, as well as role (school/day care) functioning. It contains 23 items that are grouped to create four sub-scales of PF: 8 items), Emotional Functioning (EF: 5 items), (SF: 5 items), and School Functioning (Sch F: 5 items).

A 5-point response scale is applied across each item (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). Items are reverse scored and linearly transformed to a 0–100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0) with higher scores indicating better HRQoL. For scale and total scores, the mean is computed as the sum across all items divided by the number of items answered, thereby accounting for missing data if present. The Physical Health Summary Score (8 items) is the same as the PF Subscale. To create the Psychosocial Health Summary Score (15 items), the mean is computed as the sum of the items divided by the number of items answered in the Emotional, Social, and Sch F Subscales. If more than 50% of the items in the scale are missing, the Scale Score should not be computed. Imputing the mean of the completed items in a scale when 50% or more are completed is generally the most unbiased and precise method. To do this, count the number of missing values in the scale (call it nmiss). Next, sum the item scores and divide by the number of items in the scale minus nmiss.

The questionnaire was developed through focus groups and cognitive interviews. The results demonstrate the reliability and validity of the PedsQL 4.0 Generic Core Scales. Internal consistency reliabilities generally exceeded the standard of 0.70 for group comparisons. Across the ages, the Total Scale Score for self-report and proxy-report approached an alpha of 0.90, recommended for individual patient analysis, making the Total Scale Score suitable as a summary score for the primary analysis of HRQOL outcomes in clinical trials and other group comparisons. The Physical Health and Psychosocial Health Summary Scores are recommended for secondary analyses. The Sch F Subscales for proxy-report for ages 2 to 4 and self-report for ages 5 to 7 were the only two Subscales that did not approach or exceed 0.70. The Emotional, Social, and Sch F Subscales may be utilized to examine specific domains of functioning, with the caveat that until further testing is conducted, these Subscales should be used for descriptive or exploratory analyses.

2.1.4 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, physical examination, weight, height, and immunogenicity.

Observation period

The observation period will be divided into the following epochs:

- The pre-treatment epoch is defined as the time from the signed informed consent date up to first administration of IMP.
- The treatment epoch for PAP is defined as the time from the first administration of the study drug to,
  - The time just prior to the first administration of the study drug in ETP, or,
  - Up to four weeks (28 days) after the last infusion date if the patient does not go to ETP. If the patient enrolls in another study or receives commercially available ERT, the follow-up period may be reduced from 4 to 2 weeks.
- The treatment epoch for ETP is defined as the time from the first administration of the study drug in ETP to the last administration of the study drug + up to 4 weeks (28 days). If the patient enrolls in another study or receives commercially available ERT, the follow-up period may be reduced from 4 to 2 weeks.
- The post treatment epoch for PAP is defined as the time beyond the treatment epoch in PAP through the last study assessment; this epoch is defined only for patients who do not receive any infusion in ETP.
- The post treatment epoch for ETP is defined as the time beyond the treatment epoch in ETP through the last study assessment, defined only for patients who receive any infusion in ETP.

The on-study observation period is pre-treatment, treatment and post treatment epoch.

2.1.4.1 Adverse events variables

Adverse event observation period

- Pretreatment adverse events are adverse events that developed or worsened or became serious from the signed informed consent date up to first administration of IMP.
- Treatment-emergent adverse events are adverse events that developed or worsened or became serious during the treatment-emergent period. More specifically,
  - Treatment-emergent adverse events for PAP are adverse events that developed or worsened or became serious during the treatment epoch for PAP.

  - Treatment-emergent adverse events for ETP are adverse events that developed or worsened or became serious during the treatment epoch for ETP, defined for those patients who receive any infusion in ETP.

- Post-treatment adverse events in PAP are adverse events that developed or worsened or became serious during post treatment epoch in PAP and could only be observed in patients who do not get an infusion in ETP.
- Post-treatment adverse events in ETP are adverse events that developed or worsened or became serious during the post treatment epoch for ETP and could only be observed in for patients who receive at least one infusion in ETP.

All adverse events (including serious adverse events [SAE] 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 the version of Medical Dictionary for Regulatory Activities (MedDRA) currently in effect at Sanofi at the time of database lock.

Serious adverse events

An SAE is any untoward medical occurrence that at any dose:

- Results in death, or,
- Is life-threatening, or,
  Note: The term "life-threatening" in the definition of "serious" refers to an event in which the patient was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe.
- Requires inpatient hospitalization or prolongation of existing hospitalization, or,
- Results in persistent or significant disability/incapacity, or,
- Is a congenital anomaly/birth defect,
- Is a medically important event.
  Medical and scientific judgment should be exercised in deciding whether expedited reporting is appropriate in other situations, such as important medical events that may not be immediately life-threatening or result in death or hospitalization but may jeopardize the patient or may require medical or surgical intervention (ie, specific measures or corrective treatment) to prevent one of the other outcomes listed in the definition above.

Note: The following list of medically important events is intended to serve as a guideline for determining which condition has to be considered a medically important event. The list is not intended to be exhaustive:

- Intensive treatment in an emergency room or at home for:
  - Allergic bronchospasm,
  - Blood dyscrasias (ie, agranulocytosis, aplastic anemia, bone marrow aplasia, myelodysplasia, pancytopenia, etc),

  - Convulsions (seizures, epilepsy, epileptic fit, absence, etc).
- Development of drug dependence or drug abuse,
- ALT >3 x upper limit of normal (ULN) + total bilirubin >2 x ULN or asymptomatic ALT increase >10 x ULN,
- Suicide attempt or any event suggestive of suicidality,
- Syncope, loss of consciousness (except if documented as a consequence of blood sampling),
- Bullous cutaneous eruptions.

Adverse event of special interest (AESI)

An adverse event of special interest (AESI) is an AE (serious or non-serious) of scientific and medical concern specific to the Sponsor's product or program, for which ongoing monitoring and immediate notification by the Investigator to the Sponsor is required. Such events may require further investigation in order to characterize and understand them.

AESIs will include:

- Infusion-associated reactions:

IARs are defined as AESIs that occur during either the infusion or the observation period following the infusion which are deemed to be related or possibly related to the IMP. At the discretion of the Investigator, AEs occurring after completion of the post-infusion observation period that are assessed as related may also be considered IARs. See more detailed definition in Section 2.1.5.

- Pregnancy:
  - Pregnancy occurring in a female patient included in the clinical trial. Pregnancy will be recorded as an AESI with immediate notification in all cases. It will be qualified as an SAE only if it fulfills the SAE criteria.
  - Male patients will be instructed to notify the Investigator immediately if they discover that their sexual partner is pregnant.
  - In the event of pregnancy in a female participant, IMP should be discontinued.
  - Follow-up of the pregnancy is mandatory in a female participant or in a female partner of a male participant, until the outcome has been determined.
- Symptomatic overdose (serious or nonserious) with IMP:
  - An overdose (accidental or intentional) with the IMP is an event suspected by the Investigator or spontaneously notified by the patient and defined as at least twice the intended dose within the protocol defined therapeutic interval, adjusted according to the tested drug.
    Of note, asymptomatic overdose has to be reported as a standard AE.

- Clinical laboratory (change from baseline):
  - ALT or AST increase of ≥3 x the ULN if baseline is < ULN, or ALT or AST increase ≥2 x the baseline value if baseline is ≥ ULN,
  - A maximum ALT value of ≥400 IU/L or AST value of ≥500 IU/L or an increase in direct, indirect, or total bilirubin of ≥2 x ULN,
  - Serum creatinine increase of >1.5 x the baseline value (and final serum creatinine value is > ULN).

2.1.4.2 Deaths

The deaths' observation periods are per the observation periods defined above.

- On-study: deaths occurring during the on-study observation period (PAP and ETP respectively),
- On-treatment: deaths occurring during the on-treatment epoch (PAP and ETP respectively),
- Post-treatment: deaths occurring during the post treatment epoch (PAP and ETP respectively).

2.1.4.3 Laboratory safety variables

Clinical laboratory data consist of blood analysis, including hematology, clinical chemistry, and urinalysis. Clinical laboratory values after conversion will be analyzed into standard international units and international units will be used in all listings and tables.

Blood samples for clinical laboratories will be taken as specified in the study protocol. The laboratory parameters will be classified as follows:

- Hematology: red blood cell count, hematocrit, hemoglobin, white blood cell count with differential count (neutrophils, eosinophils, basophils, monocytes, and lymphocytes), platelets,
- Biochemistry:
  - Plasma/serum electrolytes: sodium, potassium, chloride, calcium,
  - Liver function: aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase, gamma-glutamyl transferase, total and conjugated bilirubin,
  - Renal function: creatinine, blood urea nitrogen, uric acid,
  - Metabolic panel: glucose, albumin, total proteins, total cholesterol, triglycerides,
  - Potential muscle toxicity: creatine kinase, creatine kinase with MB fraction, lactate dehydrogenase.

Urinalysis will include urine color, appearance, specific gravity, proteins, glucose, erythrocytes, leukocytes, ketone bodies, and pH to be assessed:

- Qualitatively: A dipstick is to be performed on a freshly voided specimen for qualitative detection using a reagent strip.
- Quantitatively: A quantitative measurement for protein, erythrocytes, and leukocytes count will be required in the event that the urine sample test is positive for any of the above parameters by urine dipstick (eg, to confirm any positive dipstick parameter by a quantitative measurement).

2.1.4.4 Vital signs variables

Vital signs include heart rate, systolic and diastolic blood pressure, respiratory rate, temperature, and oxygen saturation.

2.1.4.5 Electrocardiogram variables

Standard 12-lead ECGs will be recorded after at least 15 minutes in the supine position using an electrocardiographic device. The following will be assessed: heart rate, rhythm, interval from start of the Q wave to the end of the S wave (QRS), interval between the peaks of successive QRS complexes (RR), interval from the beginning of the P wave until the beginning of the QRS complex (PR), interval between the start of the Q wave and the end of the T wave (QT), QT interval corrected for heart rate (QTc) automatic correction evaluation (by the ECG device), QRS axis, left ventricular hypertrophy criteria, right ventricular hypertrophy criteria, repolarization charges, and overall cardiac impression for each patient. For Day 1 only, and prior to receiving IMP, 3 ECGs within 5 minutes will be performed with at least 1 minute between 2 replicates.

2.1.4.6 Physical examination

Physical examination parameters include assessments of the patient's general appearance; skin; head, eyes, ears, nose, and throat; examinations of lymph nodes, abdomen, extremities/joints, neurological and mental status; heart and respiratory auscultation; peripheral arterial pulse; and pupil, knee, Achilles, and plantar reflexes. Head circumference and Tanner stage of sexual maturation will be assessed as part of the physical examination in pediatric patients at baseline and yearly thereafter.

2.1.4.7 Body weight and height

Body weight will be measured in kilograms and collected in the e-CRFs monthly for pediatric patients and every 3 months throughout the duration of the study for adults. More frequent weight may be obtained at the discretion of the investigator.

Standing height will be measured in all patients at baseline and annually thereafter, and in pediatric patients additionally every 3 months up to Week 73 and then every 6 months. If possible, height will be measured in the morning of the study visit day, prior to the pulmonary function testing, and by using the same stadiometer for all measurements.

2.1.4.8 Immunogenicity

Patients in the avalglucosidase alfa treatment arm will be tested for anti-avalglucosidase alfa antibodies and patients in the alglucosidase alfa treatment arm will be tested for anti-alglucosidase alfa antibodies in PAP (up to Week 49). Cross reactivity will be assessed at Week 25 and Week 49 during the PAP. Patients who are positive for anti-avalglucosidase alfa antibodies will be tested to determine if the antibodies cross-react with alglucosidase alfa and patients who are positive for anti- alglucosidase alfa antibodies will be tested to determine if the antibodies cross-react with avalglucosidase alfa at these timepoints. In the open label follow-up phase and beyond, patients from each treatment arm will be tested for anti-avalglucosidase alfa antibodies.

Samples will be collected from patients for evaluation of ADAs every month during the blinded treatment period and for the first 6 months of the open-label avalglucosidase alfa long term follow-up phase, and then every 3 months throughout the duration of the study. In addition, samples will be collected from all randomized patients at Week 2 (Day 8) and Week 52 (1 week after the first ETP treatment) to monitor for an early antibody response. ADA seropositive patient serum will be assessed for neutralizing antibodies to avalglucosidase alfa and/or alglucosidase alfa, as appropriate, including inhibition of enzyme activity and uptake.

The qualitative sample status of the ADA will be assessed and be categorized into the following classes:

- ADA-negative sample: a sample is considered negative if ADAs are not detected (ie, negative in screening assay or reactive in screening but negative in confirmatory assay).
- ADA-positive sample: sample in which ADA is detected, ie, sample generates an assay signal equal to or greater than the cut-point in the screening assay and is tested positive in the confirmatory assay.
- The ADA titer of the positive samples will also be assessed. A titer represents a quasi-quantitative information on the level of ADA present in a sample. Confirmed positive samples are serially diluted until a negative result is achieved. The titer is subsequently defined as the reciprocal of the last dilution that tests positive. The minimal required dilution (MRD) will be incorporated in the final calculation.
- The ADA attributes will be determined by the following conditions (note that some cases may not be applicable to the current study due to the treatment naïve study population):
- Pre-existing ADAs: antibodies reactive with the study drug present in subjects before treatment. The week 49 timepoint will provide data on pre-existing ADAs to avalglucosidase alfa in the alglucosidase alfa treatment arm.
- Treatment induced ADAs: ADAs developed de novo (seroconversion) following administration of the study drug. If the baseline ADA sample is missing or non-reportable and at least one reportable on-treatment ADA sample is available, the baseline sample will be considered as “negative”.

- Treatment boosted ADAs: Pre-existing ADAs that were boosted at least two titer steps from baseline (ie, 4 fold increase in titers) following administration of the study drug (any time after the first drug administration).

The following kinetics of the ADAs will be analyzed:

- Onset of ADA is defined as the time period (in days) between the first study drug administration and the first instance of treatment induced ADAs.
- Duration of ADA will be calculated as the date of last treatment induced ADA sample minus date of first treatment induced ADA sample + 1.

The following ADA response classifications will be used:

1. Treatment-induced ADA- patients are ADA negative at baseline and have developed an ADA response

    a. Transient ADA response is defined as: 1) Treatment-induced ADA detected only at one sampling timepoint post-baseline (excluding the last sampling time point); or 2) 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
    b. Persistent ADA response is defined as: 1) 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; or 2) 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. The titer values are based on LOTS alglucosidase alfa clinical experience.

        - Low response – if a patient peak titer ≤800 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 was persistently seropositive but titer is 1600-6400 and is positive at final assessment.
        - High response – if a patient was persistently seropositive and titer is ≥12800 and is positive at final assessment.
        - Tolerized – if a patient was persistently seropositive, but negative at the final assessment.
    c. Indeterminate ADA response – if the patient developed ADA at the last time point and all previous samples are ADA negative, therefore cannot determine whether the response will be transient or persistent in duration.

2. Treatment-boosted ADA – patients have pre-existing ADA (positive at baseline or study entry) and have ADA titers boosted to a higher level by a greater than or equal to four-fold increase (i.e., by greater than at least twice the dilution level).
3. Treatment emergent ADA- combination of treatment induced and treatment boosted

2.1.5 Infusion associated reactions (IARs)

Protocol-defined IARs

IARs are defined as AESIs that occur during either the infusion or the observation period following the infusion which are deemed to be related or possibly related to the IMP. At the discretion of the Investigator, AEs occurring after completion of the post-infusion observation period that are assessed as related may also be considered infusion associated reactions (IAR).

Algorithm-defined IARs

An alternative definition of IAR is defined as any treatment-emergent AE meeting one of the following criteria:

a) Event occurs from the start of infusion to the end of infusion plus 24 hours window, and considered related to study drug, ,
b) If AE start date is non-missing but time component is missing, compare AE Start date with infusion start date (date component only) and infusion end date (date component only). If AE Start date is between infusion start date and infusion end date plus 1 day, consider such AE as Algorithm -defined IAR if AE is related to study drug (missing relatedness is assumed related).

2.1.6 Pharmacokinetic endpoints

Pharmacokinetic variables include $C_{max,}$ $AUC_{0\text{-last}}$, total body clearance of a drug from the plasma and volume of distribution, if appropriate.

2.1.7 Pharmacogenetic endpoints

Pharmacogenetic endpoints include mutations within the GAA gene and ACE allele genotype status. The genetic results obtained from this study are not diagnostic and will be used for research purposes only.

2.1.8 Pharmacodynamic endpoints

Pharmacodynamic endpoints include the urinary Hex4 level.

2.1.9 Patient Global Impression of Change

The PGIC items will be completed at Week 49 (the conclusion of the PAP), and will be administered annually in follow-up during the extension period. The PGIC items consist of

4 questions pertaining to overall disease-related symptoms, activities of daily living, as well as mobility and respiratory issues. The items range from -3 (a great deal worse) to 0 (no change) to 3 (a great deal better). The data from this scale will be used to support and validate additional endpoints in the trial.

2.1.10 Quality of life/health economic variables/other endpoints

The following endpoints are included:

- The PDSS and PDIS will be completed via an e diary daily for 2 weeks between visits according to the protocol specified schedules. Patients ≥18 years of age at screening/baseline will complete this assessment through both a 24-hour recall version administered daily during the screening period, as well as a 7-day recall version administered prior to the first infusion. The PDSS and PDIS are both self-administered questionnaires specifically designed to capture the symptoms and impacts pertinent to patients with LOPD. The PDSS contains 12 questions on a scale from 0 (none) to 10 (as bad as I can imagine), while the PDIS contains 15 questions with varying scales implemented depending on question type.
- The R-PAct scale will be completed in selected countries (ie, UK, USA, Canada, Belgium and The Netherlands). Patients whose first language is English or Dutch and who are ≥18 years of age at screening/baseline will complete this assessment. During the treatment period, the assessment will be completed before IMP infusion if possible. The R-PAct is a self-administered questionnaire specifically suited to quantify the effects of Pompe disease on patient's ability to carry out daily activities and their social participation. It consists of 18 items suited to quantify activity limitations, ranging from unable to perform daily life activities (0) to able to perform without difficulty (2) in patients with Pompe disease (26).

These endpoints are exploratory endpoints that are specific to Pompe disease. The PDSS/PDIS are scales recently developed within Sanofi, and the scoring method will be developed following the completion of an external validation study. These will be described and documented in a separate analysis plan.

2.2 DISPOSITION OF PATIENTS

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

Screened patients will be defined as any patient who has signed the informed consent form.

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

Patients treated without being randomized will not be considered as randomized and will not be included in any efficacy population. The efficacy results for the pediatric patients who are enrolled directly in the open-label avalglucosidase alfa long term follow-up phase will be listed.

For patient study status, the total number of patients in each of the following categories will be presented in the PAP clinical study report using summary table:

- Screened patients,
- Screen failure patients and reasons for screen failure (if data is available),
- Nonrandomized but treated patients (if applicable),
- Randomized patients,
- Randomized but not treated patients,
- Randomized and treated patients,
- Number and percentage of patients who did not complete PAP treatment period, with corresponding reasons,
- The total number of patients in each of the following categories will be presented in the ETP clinical study report using summary table.
- Patients who continue into ETP,
- Patients who did not complete ETP,
- Pediatric patients enrolled directly in the open-label avalglucosidase alfa long term follow-up phase
- Number and percentage of patients who did not complete ETP treatment, with corresponding reasons (will be included in final study report only),
- Number and percentage of patients who did not complete ETP follow-up, with corresponding reasons (will be included in final study report only).

For all categories of patients (except for the screened and nonrandomized categories) the summaries will be provided by randomized treatment, and percentages will be calculated using the number of randomized patients as the denominator.

All critical or major deviations potentially impacting efficacy analyses, randomization, and drug-dispensing irregularities, and other major or critical deviations will be summarized in tables giving numbers and percentages of deviations by treatment group.

Additionally, the analysis populations for safety, efficacy, and PK will be summarized in a table by number of patients on the randomized population.

- Efficacy population: modified intent-to-treat (mITT) population/per-protocol population,
- Safety population,
- PK population.

2.2.1 Randomization and drug dispensing irregularities in PAP

Randomization and drug-dispensing irregularities occur whenever:

- 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 based on an incorrect stratum, c) a patient is randomized twice;
  OR,
- A patient is dispensed an IMP kit not allocated by the protocol-defined randomization, such as a) a patient at any time in the study is dispensed a different treatment kit than as randomized (which may or may not contain the correct-as-randomized IMP), or b) a nonrandomized patient is treated with IMP reserved for randomized patients.

Randomization and drug-dispensing irregularities will be monitored throughout the study and reviewed on an ongoing basis.

All randomization and drug-dispensing irregularities will be documented in the clinical study report. Whether any of these constitute a major protocol deviation is deferred to the decision by the clinical team before the database lock. Nonrandomized, treated patients will be described separately.

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

Table 13 - Randomization and drug allocation irregularities

Randomization and drug allocation irregularities
Kit dispensation without IRT transaction
Erroneous kit dispensation
Kit not available
Randomization by error
Patient randomized twice
Stratification error
Patient switched to another site

2.3 ANALYSIS POPULATIONS

The randomized population includes any patient who has been allocated to a randomized treatment regardless of whether the treatment kit is used. The pediatric patients who are enrolled directly in the open-label avalglucosidase alfa long term follow-up phase will be included in the ETP population and will not be included in the randomized population. Listings will be generated based on enrolled population which includes all randomized population and pediatric patients who are enrolled directly in the open-label phase.

2.3.1 Efficacy populations

Patients treated without being randomized will not be considered randomized and will not be included in any efficacy population.

2.3.1.1 Modified intent-to-treat population

The mITT population will include randomized patients who receive at least 1 infusion (partial or total). The mITT population will be used for all efficacy analyses unless otherwise specified. Analyses using the mITT population will be performed according to the treatment arm allocated by randomization, regardless of the actual treatment received.

2.3.1.2 Per-protocol population

The per-protocol (PP) population will consist of mITT patients who meet all of the following criteria:

- Meet all of the inclusion and exclusion criteria (See protocol Section 7.1 and 7.2),
- Received at least 80% of planned # of doses,
- Having a valid FVC (% predicted) assessment at Week 49,
- No major protocol deviations that will potentially impact on primary study objective.

The following criteria have been identified a priori as major deviations that will potentially impact on primary study objective. Additional major deviations maybe identified during the study. These will be reviewed and approved by sponsor prior to database lock and unblinding of treatment assignment.

- Patient receives doses 2 times higher than the protocol planned dose,
- Patient receives treatment different from the randomized assignment during the PAP,
- The treatment assignment is unblinded before the planned unblinding (completion of PAP) not due to safety reasons specified in the protocol.

The reason for excluding patients from the PP population will be summarized by randomized treatment group and presented in data listing.

PP population will be used for sensitivity analysis of the primary efficacy endpoint in PAP only.

2.3.2 Safety population

The safety population will be analyzed according to treatment received. In PAP, safety population includes randomized patients who receive at least 1 infusion in PAP. In ETP, safety population includes patients who receive at least 1 infusion in ETP and includes pediatric patients who are directly enrolled to ETP. Overall safety of avalglucosidase alfa will be based on patients who receive at least 1 infusion during either PAP or ETP.

- Nonrandomized but treated patients will not be part of the safety population in PAP; however, their safety data will be presented separately in PAP.
- Randomized patients for whom it is unclear whether they took the IMP will be included in the safety population as randomized.

2.3.3 Additional analysis populations

The ETP population will consist of patients who receive at least one avalglucosidase alfa dose during the ETP.

The pharmacokinetic population will consist of patients from safety population in PAP who have evaluable drug concentration data.

The pharmacodynamics or pharmacogenetic analysis population for the parameter of interest will consist of mITT patients and have evaluable pharmacodynamic data for the parameter of interest.

The ADA evaluable population will consist of patients from safety population who have at least one ADA sample taken post-baseline after drug administration that is appropriate for ADA testing with a reportable result. Patients with missing or non-reportable baseline samples and reportable post-baseline samples will be considered as evaluable.

2.4 STATISTICAL METHODS

Unless otherwise specified, efficacy analyses will use the mITT population, where patients will be considered to be in the treatment group to which they were randomized. Corresponding supportive analysis will be performed for PP population as well. Safety analyses will use the safety population, where patients who receive at least one infusion of avalglucosidase alfa in PAP will be assigned to the avalglucosidase alfa arm for the PAP safety analysis purposes.

For all analyses, the avalglucosidase alfa treatment group will be compared to the alglucosidase alfa group. While the primary comparison of interest is at 12 months (49 weeks), all summary statistics will be computed and displayed by treatment group and each scheduled assessment time as well. Summary statistics for continuous variables will minimally include n, mean, SD, minimum, median, and maximum. For categorical variables, frequencies and percentages will be presented. Graphical displays will be provided as appropriate.

2.4.1 Demographics and baseline characteristics

Continuous data will be summarized using the number of available observations, mean, SD, median, minimum, and maximum for each treatment group. Categorical and ordinal data will be summarized using the number and percentage of patients in each treatment group. Denominators for the percentages will be based on the analysis population used, unless otherwise specified.

Demographics and baseline disease characteristics as described in Section 2.1.1 will be summarized by randomized treatment group and pooled treatment groups in mITT population. Summaries for the safety population will be included in the appendices if the size of the safety

population is different (>10%) from the size of that in the primary analysis population for any treatment group, or if the randomized treatment for any patient is different from the actual treatment received.

Medical history will be summarized by body system and treatment group and overall in mITT population.

Baseline values of the key efficacy parameters such as FVC, 6MWT, MIP, MEP, lower extremity HHD, QMFT and SF-12, will be summarized by randomized treatment group and pooled treatment group in mITT population. Baseline values of additional secondary efficacy parameters will be summarized along with each efficacy analysis. GAA gene and ACE allele genotype status will also be summarized.

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

No specific description of the safety parameters will be provided at baseline. If relevant, the baseline values will be described along with each safety analysis.

2.4.2 Prior or concomitant medications

Prior medications will be defined as medications that are taken in the 30 days before the first infusion of study drug. Concomitant medications will be defined as medications that are taken after the first infusion of study drug. If a drug is started before first infusion of study drug but continues after the first infusion of the study drug, it is considered prior as well as concomitant.

Medications will be summarized by treatment group according to the WHO-DD dictionary, considering the first digit of the anatomic and therapeutic category (ATC) class 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 several times for the same medication. Prior and concomitant mediations will be summarized separately and will use mITT population.

In ETP analyses, prior and concomitant medications will also be summarized for the ETP treatment epoch among the ETP population, as well as for the overall avalglucosidase alfa treatment period among patients who receive at least one dose of avalglucosidase alfa in either PAP or ETP.

2.4.3 Extent of investigational medicinal product exposure and compliance

2.4.3.1 Extent of investigational medicinal product exposure

The extent of study drug exposure will be assessed by the duration of study drug exposure, number of infusions, and amount of dose received. The extent of study drug exposure will be summarized in safety population. Patient year as calculated from the first infusion date to the date of last follow-up visit will be summarized by treatment.

Duration of study drug exposure is defined below:

For PAP

- Duration of IMP exposure is defined as the first dose date in the ETP (or 14 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 ETP,
- Duration of IMP exposure is defined as 14 days after last dose in ETP–first dose date in the ETP, regardless of unplanned intermittent discontinuations. This is defined only for those patients treated in ETP.
- For the overall treatment duration,
- Duration of avalglucosidase alfa total exposure is defined as 14 days after last avalglucosidase alfa dose in study–first avalglucosidase alfa dose date in the study.

The duration of IMP exposure will be summarized descriptively as a quantitative variable (number, mean, SD, median, minimum, and maximum), as well as categorical variable (eg, <3 months, 3-6 months, 6-9 months, 10-12 months, >12 months for PAP).

The cumulative dose information for PAP will be assessed by the total number of infusions received in PAP, as well as total amount of IMP in ml (if this information is available from the database) received in PAP. These data will be summarized descriptively.

The cumulative dose information for avalglucosidase alfa during the overall treatment duration will be assessed by the total number of avalglucosidase alfa infusions received during the overall treatment duration, as well as the total amount of avalglucosidase alfa in ml received during the overall treatment duration. These data will be summarized descriptively by randomized treatment group (Alglucosidase alfa PAP/avalglucosidase alfa ETP vs. avalglucosidase alfa PAP/avalglucosidase alfa ETP), among the safety population.

2.4.3.2 Compliance

Treatment compliance for PAP will be summarized descriptively as quantitative variable using mITT population and randomized treatment group. Compliance is calculated as the total amount of drug actually taken by a patient divided by the total amount of drug expected to be taken multiplied by 100. The number and percentage of patients with noncompliance (missed 2 or more consecutive infusions not for safety cause, or missed ≥20% of total doses in the 49-week PAP), and by reasons will be provided.

Treatment compliance for ETP and overall study period will be summarized similarly.

2.5 ANALYSES OF EFFICACY ENDPOINTS

All efficacy endpoints will be summarized descriptively at each study visit, as well as the last observation available during the PAP. All efficacy analyses will be performed based on the mITT

population. Corresponding analysis will be performed for PP population for the primary efficacy endpoint. If the pure ITT (all randomized patients) population is different from the mITT population, we plan to perform a sensitivity analysis in this population as well to assess the robustness of the results.

2.5.1 Analysis of primary efficacy endpoint in PAP

The primary efficacy endpoint of change from baseline in FVC (% predicted) in upright position to Week 49 will be analyzed in the mITT population using MMRM with change from baseline as the outcome variable. The MMRM model will include the baseline FVC (% predicted) and age (both as continuous variables), gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects.

The analysis will include all post baseline scheduled assessments up to Week 49, regardless of treatment discontinuation status; missing data will not be imputed and will be assumed to be Missing at Random (MAR). Descriptive assessment of MAR assumption and sensitivity analyses to assess the impact of missing data are described in Section 2.5.3.

An unstructured covariance matrix shared across treatment groups will be used to model the within-patient errors. The model will be fitted using restricted maximum likelihood. If the model does not converge with an unstructured covariate matrix, the following covariate matrix structures will be used in the order of heterogeneous Toeplitz (heterogeneous variance, an extension of homogeneous Toeplitz), homogeneous Toeplitz (equal variance and a separate correlation for each level of separation between the time points), heterogeneous AR(1) (heterogeneous variance, an extension of AR(1)), AR(1) (first-order autoregressive, equal variances and exponentially decreasing correlations). The first covariance structure yielding convergence will be used as the primary analysis.

The Kenward-Roger approximation will be used to estimate the denominator degrees of freedom. The difference between treatment groups will be estimated based on least-square means at the Week 49 visit within the MMRM model. A two-sided 95% confidence interval using the estimate and variance of the least-square mean difference will be provided.

Due to the small sample size within each country or region, the primary MMRM will not include region/country as a covariate. Also, since previous alglucosidase alfa studies did not suggest a baseline-by-visit interaction within either study treatment arm, the primary MMRM will not include the baseline-by-visit interaction in the model to improve statistical efficiency.

The primary analyses will be performed after all randomized patients have been followed for at least 12 months (49 weeks) after randomization (or discontinued the study follow-up early).

The primary statistical objective is to test the non-inferiority of avalglucosidase alfa versus alglucosidase alfa at two-sided 5% level of significance. The null and alternative hypotheses based on a non-inferiority margin of 1.1 are described as H_{01} and H_{a1} below:

H_{01}: avalglucosidase alfa–alglucosidase alfa $\leq$-1.1 versus H_{a1}: avalglucosidase alfa–alglucosidase alfa >-1.1

If the lower bound of the two-sided 95% confidence interval for the difference of avalglucosidase alfa-alglucosidase alfa is larger than -1.1, the study will be considered to have met its primary objective and the non-inferiority of avalglucosidase alfa over alglucosidase alfa is established.

After non-inferiority is demonstrated, a test for superiority of avalglucosidase alfa versus alglucosidase alfa will be performed with an overall two-sided 5% level of significance.

The following SAS code will be used for the primary efficacy analysis:

```
ods output Diffs = out1
        LSMeans = out2;
proc mixed data = DATAIN method = reml;
  class TRT VISIT SEX USUBJID;
  model CHG_FVC = BL_FVC SEX  AGE TRT VISIT   TRT*VISIT/ ddfm = kr;
  repeated VISIT/type = un subject = USUBJID  r;
  lsmeans TRT TRT*VISIT/cl diff;
run;
```

where TRT = treatment group, VISIT = (week 13, 25, 37, 49), SEX = (male, female), AGE = age at baseline (continuous), USUBJID = unique patient ID, CHG_FVC is the change from baseline in FVC (% predicted), and BL_FVC is the baseline FVC (% predicted) value.

The same analysis as described above will be performed for the PP population to assess the robustness of the study outcome.

A US FDA-specific exploratory analysis will be conducted using the method described above for those patients in the mITT population 8 years of age or older in order to match the lower age range of patients in the LOTS trial of alglucosidase alfa. This analysis will be used to support US approval.

In the case that both NI and superiority of FVC (% predicted) are demonstrated, a superiority test for 6MWT will be performed with a two-sided alpha of 0.05.

Note that the type I error is controlled for the primary endpoint of FVC (% predicted) in the above testing procedure. Analyses of secondary efficacy endpoints in PAP

The secondary efficacy endpoints for PAP are described in Section 2.1.3.2.

2.5.1.1 Distance walked in 6MWT

2.5.1.1.1 Primary analysis of 6MWT based on distance walked

The primary statistical objective for 6MWT is to test the superiority of avalglucosidase alfa versus alglucosidase alfa at Week 49 in the mITT.

The change from baseline in total distance walked in 6MWT will be analyzed based on the MMRM model, similar to the primary endpoint described in Section 2.5.1, assuming MAR. The

model will include the total distance walked in 6MWT at baseline, baseline FVC (% predicted) and age (as continuous variables), gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects. The difference between treatment groups will be estimated based on least-square means at the Week 49 visit within the MMRM model. A two-sided 95% confidence interval using the least-square mean difference and its variance will be provided. If the lower bound of the two sided 95% confidence interval for the least square mean difference of avalglucosidase alfa–alglucosidase alfa is greater than 0, then the statistical superiority of avalglucosidase alfa over alglucosidase alfa is established.

2.5.1.1.2 Additional supportive analyses for 6MWT

Additional supportive analyses related to the 6MWT endpoint are as follow:

- 6MWT % predicted change from baseline

  The change from baseline in % predicted of the total distance walked in 6MWT will be analyzed separately based on the MMRM model, similar to the primary endpoint described in Section 2.5.1. The model will include the % predicted distance walked in 6MWT at baseline, baseline FVC (% predicted) and age (as continuous variables), as well as gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects. A two-sided 95% confidence interval using the least-square mean difference will be generated in a similar way as the absolute distance in meters.

- Categorical response of 6MWT

  Relative change from baseline in 6MWT for last observation within 49 weeks period for each subject will be categorized into <-15%, [-15%, -10%], [-10, -5%), [-5%, 0), [0, 5%), [5%, 10%), [10%, 15%), and ≥15%. The number and percent of patients within each category will be summarized by treatment group. Treatment difference will be tested with Fisher's exact test.

- Analysis of 6MWT based on subjects who completed the full 6 minute walk by excluding subjects that walked less than 6 minutes.

- Responder analysis

- The derivation of responder criteria for the 6MWT is complicated by the challenges associated with defining MCIDs for this outcome measure (27). Based on Redelmeier (27), 3 responder thresholds are currently proposed for 6MWT endpoint, similar to what was considered in previous LOTS study: 1) 54 meters (the estimated threshold), 2) 37 meters (the lower bound of the 95% CI of the estimated threshold and 3) 30 meters (the difference in 6MWT that result in patients feeling 'a little bit better' (27). In addition, a threshold of 27.5 meters is proposed based on a half of standard deviation of the change from baseline to Week 52 as observed in the previous LOTS study. A response will be defined as having a change from baseline in 6MWT distance of ≥54 meters, 37 meters, 30 meters, or 27.5 meters, respectively, based on Week 49 value. Patients without Week 49 assessment will be treated as non-responders. To compare the relative treatment difference for each of the response variable, an odds ratio (with 95% confidence interval)

for avalglucosidase alfa vs. alglucosidase alfa will be calculated based on logistic regression model with treatment group, adjusting for baseline 6MWT, age, and gender.

- Changes in use of assistive device

  Number and percent of patients who have increased use of walking device (compared to baseline) will be summarized descriptively by study visit and treatment group. Time to increased use of walking device will be summarized using Kaplan-Meier estimate if sufficient number of patients reported increased use of walking device during the PAP.

  When applicable, number and percent of patients with decreased use of assistive device compared to baseline will also be summarized descriptively by study visit and treatment group.

2.5.1.2 MIP and MEP from pulmonary function testing

The change from baseline in MIP % predicted and MEP % predicted will be analyzed separately based on the MMRM model, similar to the primary endpoint described in Section 2.5.1. For each of the endpoints, the model will include the baseline value of the corresponding response variable (either MIP or MEP), age (as continuous variable), gender (male or female), treatment group, visit, and treatment by visit interaction as fixed effects. Due to the correlation between baseline FVC and MIP or MEP, and to avoid potential col-linearity issue, the baseline FVC (used as a randomization stratification factor) is not included in the current MMRM model.

Since the % predicted values can only be calculated for patients with age of 7 or above who have a reference value available, the analyses for MIP % predicted and MEP % predicted will be limited to the mITT patients with age of 8 years or older.

The raw values of MIP and MEP will be summarized descriptively based on mITT.

2.5.1.3 Hand-held dynamometry (lower extremity muscle strength)

Change from baseline in lower extremity muscle strength composite score will be analyzed based on the MMRM model, similar to the primary endpoint described in Section 2.5.1. The model will include summary HHD lower extremity score at baseline, baseline FVC (% predicted) and age (as continuous variables), gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects.

As a supportive analysis, change from baseline in % predicted of the summary score for the lower extremity muscle strength will be analyzed based on the MMRM model, similar to the primary endpoint described in Section 2.5.1. The model will include the % predicted of the summary score at baseline, baseline FVC (% predicted), and age (as continuous variables), gender (male or female), treatment group, visit, baseline HHD % predicted score-by-visit interaction and treatment-by-visit interaction as fixed effects. Similar analyses on the % predicted for each muscle group will be conducted.

Since % predicted values are only available for all muscle groups for patients age 6 or older, the analysis on % predicted HHD lower extremity summary will be performed in mITT patients with an age of 6 or older.

Raw score as well as the % predicted scores for each muscle group will also be summarized descriptively by treatment group and study visit.

In order to evaluate quality of observed changes in muscle strength measured with HHD, a sub-group analysis of HHD data will be performed of data confirmed to be collected by same assessor and with the patient in the same position as baseline and Week 49. HHD outcomes will include all muscle groups and UE and LE composites.

2.5.1.4 Quick Motor Function Test

Change from baseline in total score of QMFT will be analyzed based on the MMRM model in mITT, similar to the primary endpoint described in Section 2.5.1. The model will include total score at baseline, baseline FVC (% predicted), age (as continuous variables), gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects.

Each individual item of the QMFT will be summarized descriptively.

2.5.1.5 12-item short form health survey

Change from baseline in PCS and MCS will be analyzed separately based on the MMRM model, similar to the primary endpoint described in Section 2.5.1. The model will include baseline score (PCS or MCS), baseline FVC (% predicted), and age (as continuous variables), gender (male or female), treatment group, visit, and treatment-by-visit interaction as fixed effects. Descriptive summary of change over time and shift from baseline will be performed as well.

The analyses will be conducted based on the mITT patients who are of age ≥18 at screening/baseline.

2.5.2 Multiplicity issues

A sequential test strategy for the primary efficacy endpoint and key secondary endpoints will be used to control Type 1 error rate at 5% in a strong sense. Testing will proceed according to the following order and will stop if there is a non-significant comparison:

1. The primary efficacy endpoint of FVC (% predicted) will be tested for NI of avalglucosidase alfa versus alglucosidase alfa first.
2. If NI is demonstrated, the superiority of the avalglucosidase alfa versus alglucosidase alfa in FVC (% predicted) will be tested with the same overall 5% significance level.
3. If the superiority of avalglucosidase alfa versus alglucosidase alfa is demonstrated on the primary efficacy endpoint, the hypothesis testing for the secondary efficacy endpoints will proceed according to the following order:
    a) Change from baseline to Week 49 in total distance walked in 6MWT (a superiority test with a two-sided alpha of 0.05),

b) Change from baseline to Week 49 in % predicted of MIP (a superiority test with a two-sided alpha of 0.05),

c) Change from baseline to Week 49 in % predicted of MEP (a superiority test with a two-sided alpha of 0.05),

d) Change from baseline to Week 49 in summary score of lower extremity strength by HHD (a superiority test with a two-sided alpha of 0.05).

Analyses of additional secondary endpoints will be considered for supportive purposes.

2.5.3 Sensitivity analyses for the primary efficacy endpoint in PAP

All sensitivity analyses will be performed using the mITT population, unless otherwise specified.

2.5.3.1 Sensitivity analyses to assess the impact of missing data

The primary analysis model of MMRM for FVC (% predicted) change from baseline is based on the assumption that missing data are MAR. To assess this assumption, missing data patterns will first be summarized descriptively by study visit and treatment. Then, the efficacy profile of those patients with missing FVC (% predicted) at each study visit will be presented by treatment group, as well as the reason for missing at the end of follow-up.

Specifically, the following descriptive summaries will be provided:

- Mean (SE) change from baseline by time of dropout (Week 13, 25, 37, 49, or no dropout [or >49 weeks]) within each treatment arm, as well as both treatment groups combined.
- Mean (SE) change from baseline by completion status (completers vs. dropouts) and further by reason for dropout among those patients who dropped within each treatment arm, as well as both treatment groups combined. The categories maybe combined or modified based on clinical reason before the study is unblinded.

The following sensitivity analyses using tipping-point method will be conducted to address potential deviation of MAR assumption. The tipping point approach is like a progressive stress-testing to assess how severe departures from MAR must be in order to overturn conclusions from the primary analysis. If implausible departures from MAR in order to change the results from statistically significance to statistically insignificance, the results will be considered to be robust to the departure from MAR assumption. We will then be more confident in the results obtained based on statistical methods with the MAR assumptions in MMRM. Reference-based imputation is not planned since they may bias towards equivalence and thus is not conservative for NI trial. Additional sensitivity analyses may be added as appropriate if the efficacy profile summaries described above suggestive of NMAR.

Missing data will be imputed 1000 times to generate 1000 complete datasets with the SAS Proc MI procedure. Since in general, the missing pattern will not be monotone, a two-step approach will be used and will be performed by treatment group:

- Step 1: Intermittent (non-monotone) missing data will be imputed first based on the missing-at-random (MAR) assumption and a multivariate joint Gaussian imputation

model using Markov chain Monte Carlo (MCMC) method within each treatment arm,

- Step 2: The remaining, monotone missing data will be imputed using the sequential regression method.

MAR-based imputations for monotone missing data will be generated using sequential regression multiple imputation, where a separate regression model is estimated for imputation of FVC (% predicted) at each time point. Each regression model will include explanatory variables of treatment, baseline FVC (% predicted), age, gender, FVC (% predicted) values at all previous visits.

The model adds a sensitivity parameter, δ, to the imputed values to reflect the difference in mean change from baseline in FVC (% predicted) between patients with missing data and patients with observed data. A sensitivity parameter δavalglucosidase alfa will be introduced to reflect violation of MAR for the avalglucosidase alfa group. The final imputed values for the avalglucosidase alfa group will be the imputed values from previous two steps-δavalglucosidase alfa. The imputed values for the alglucosidase alfa will remain the same, ie, the sensitivity parameter for alglucosidase alfa, δaa, is set to 0. For each given δavalglucosidase alfa value, a MMRM with FVC (% predicted) change from baseline as the outcome variable and the baseline FVC (% predicted), age (both as continuous variables), gender (male or female), treatment group, visit, and treatment by visit interaction as covariates will be fit. The results from the 1000 analyses will be combined using Rubin's formulae and the 95% confidence interval will be constructed. The testing will be repeated over a range of plausible values for δavalglucosidase alfa, and the tipping points with which the non-inferiority is not achieved will be identified separately. This analysis will be performed only when the non-inferiority is demonstrated in the primary analysis.

2.5.3.2 Sensitivity analyses with alternative model or different distribution assumption

The following analyses will be performed to address potential deviation from the primary model assumption

1. An analysis of covariance including baseline FVC, age, gender, treatment group as covariates will be performed for the endpoint of change from baseline in FVC % predicted at Week 49. For the patients who are known to start alternative treatment prior to Week 49, the last value prior to initiation of an alternative treatment will be used in the analysis. For other patients who prematurely discontinued prior to Week 49 but with unknown alternative treatment information, their last value prior to dropout will be used in the analysis.
2. A Wilcoxon-Mann-Whitney (WMW) will be performed for change from baseline in FVC (% predicted) at Week 49. Missing FVC (% predicted) at Week 49 will be imputed by baseline or last assessment (within PAP) for the subject, whichever is worse. This analysis will assess the robustness of the results when the normality assumption which underlies the MMRM model is deviated.
3. Change from baseline in FVC (% predicted) will be analyzed with MMRM model including the covariates specified in the primary model (baseline FVC, age, gender, treatment group, visit, and treatment-by-visit interaction as fixed effects).

4. The FVC (% predicted) will be analyzed with a Linear Mixed Effects model which includes fixed effects of age (as continuous variables), gender, treatment, time (in years) and the treatment * time interaction; as well as subject specific random intercept and random slope. The model will be fitted using restricted maximum likelihood estimation with the PROC MIXED procedure in SAS**.** Comparison between treatment groups will be made by testing and estimating the contrast of the treatment * time interaction in the model. All FVC (% predicted) values from baseline through Week 49 (based on actual assessment date) will be included for this analysis. This analysis model will be appropriate if the FVC (% predicted) over time is approximately linear.

2.5.3.3 Sensitivity analyses with respect to the constancy assumption

It is expected that the 2 populations from LOTS trial and current trial are generally similar. Sensitivity analyses are planned for key prognostic and predictive factors to demonstrate that the results are robust to small imbalances. The constancy assumption requires that the effect of the active comparator of alglucosidase alfa relative to placebo in the current non-inferiority trial is similar to the effect that was observed in the historical LOTS trial. The constancy assumption will be empirically checked by comparing change from baseline of FVC (% predicted) at Week 49 between the current trial and the historical trial. If constancy assumption violation is suspected based on clinical judgement, an exploratory analysis using the covariate-adjustment regression model approach proposed by Nie and Soon (2) will be performed to evaluate the impact of population difference between the historical trial and the current trial on the degree of constancy assumption violation. An ANCOVA model will be fitted to LOTS data to generate a predictive model for the outcome of change from baseline in FVC (% predicted) at Week 49 as a function of the following covariates: treatment, age, gender, race, duration of disease, baseline FVC, baseline 6MWT, respiratory support device use at baseline, and treatment interaction with gender, baseline FVC, respiratory support device use at baseline based on LOTS trial results. If the difference of the estimated relative effect of control over placebo between the current trial population and the historical LOTS trial population exceeds the non-inferiority margin, the constancy assumption may be violated. In this case, covariate-adjustment fixed margin may be performed by using α_1=0.2 and α_2=0.05 in the approach proposed by Nie and Soon (2) to adjust for population difference. Note that this approach assumes that violation of constancy assumption is caused by known, imbalanced, predictive baseline characteristics. In reality, violation of constancy assumption can also be caused by unknown baseline characteristics or post-baseline characteristics. In that case, the covariate-adjustment may not be able to make correct adjustment.

2.5.3.4 Additional supportive analyses for the primary endpoint

2.5.3.4.1 Responder analysis

A responder analyses based on relative change from baseline in FVC (% predicted) will be performed. Response will be defined as having a relative change from baseline of ≥5%, 10% or 15% increase from baseline at Week 49 respectively. Patients without Week 49 value will be considered as non-responders.

2.5.3.4.2 Changes in respiratory support

Number and percent of patients with new use or increased use of respiratory device (compare to baseline) will be summarized descriptively by treatment group at Week 13, 25, 37 and 49. Time to increased use of respiratory support will be summarized using Kaplan-Meier estimate if sufficient number of patients reported such events during the PAP. Similarly, time to decreased use of respiratory support will be summarized using Kaplan-Meier estimate if sufficient number of patients reported such events during the PAP. Patients without event will be censored at the last visit within PAP.

2.5.3.4.3 Correlation between FVC (% predicted) change and other efficacy parameters

Correlation between FVC (% predicted) change, 6MWT change, and other secondary efficacy endpoints plus patient reported outcomes, HHD upper extremity muscle strength, MIP, MEP, as well as safety endpoints such as inhibitory antibody status and peak antibody titer will be explored graphically. Corresponding summary tables will be generated as appropriate.

2.5.4 Subgroup analyses for the primary efficacy endpoint and key secondary efficacy endpoint in PAP

Subgroup analyses for the primary efficacy endpoint of FVC % predicted and key secondary efficacy endpoint 6MWT will be performed. They will include age group (<18 years, ≥18 years but <45 years, and ≥45 years old), gender, baseline FVC groups (<55%, and ≥55%), race, and ethnicity. Due to the limited power in these subpopulations, treatment difference within the subgroup will be difficult to interpret. Therefore, the focus of this analysis is on assessment of interaction between subgroup covariate and the treatment. If p-value for the interaction term is less than 0.1, nature of interaction will be explored to determine if it quantitative interaction or qualitative interaction.

The following three MMRM models for the primary efficacy endpoint will be performed:

- MMRM model to include baseline FVC (% predicted, as continuous), age (as continuous), gender, treatment group, visit, treatment-by-gender interaction, and treatment-by-visit interaction as fixed effects. This will be used to assess the gender interaction with the treatment.
- MMRM model to include baseline FVC (% predicted, as continuous), age (<18 years old and ≥18 years old), gender, treatment group, visit, treatment-by-age interaction, and treatment-by-visit interaction as fixed effects. This will be used to assess the age interaction with the treatment. The sample size for age <18 maybe too small for this analysis.
- MMRM model to include baseline FVC (% predicted, as continuous), age (as continuous), gender, treatment group, visit, treatment-by-FVC (categorical) interaction, and treatment-by-visit interaction as fixed effects. This will be used to assess the baseline FVC interaction with the treatment.

The same estimation method, covariance matrix and degree of freedom calculation method as that of the primary analysis will be employed here. The p-value for interaction effect will be provided. A two-sided 95% confidence interval within each subgroup for the least-square mean difference between treatment groups will be provided within the framework of these models.

Additional exploratory subgroup analyses will be performed as well, based on baseline covariates such as region (US vs. non-US), baseline use of walking device, baseline 6MWT, duration of disease, use of walking device, use of respiratory support, and ACE genotype. These data will be summarized descriptively by treatment group within the subgroups. The cutoff based on population median will be used for the continuous variables.

2.5.5 Additional efficacy analyses including PRO endpoints

All other exploratory efficacy endpoints will be summarized descriptively and analyzed using a similar method for the primary efficacy endpoint described in Section 2.5.1. For PDSS/PDIS, following the scoring method and algorithm, results will be summarized by descriptive statistics, MMRM, and responder analysis. The analyses will be based on mITT population.

Analyses of efficacy data in extension treatment (ETP)

Unless otherwise specified, ETP efficacy analyses will use the mITT population, where patients will be considered to be in the treatment group to which they were randomized.

All efficacy endpoints will be summarized descriptively at each study visit, as well as the last observation available during the study. For study visits in ETP (beyond week 49), both changes from baseline and changes from ETP baseline (week 49) will be presented.

2.5.5.1 Analyses of efficacy data in ETP at PAP database lock

At PAP database lock, all available efficacy data for those patients who have entered ETP will be summarized descriptively. Graphical displays will be provided when appropriate.

Long-term effect of avalglucosidase alfa will be assessed among patients who receive avalglucosidase alfa throughout the blinded treatment phase, as well as the long-term follow-up phase. Descriptive statistics and graphical display will be provided to assess trend over time on all efficacy endpoints.

2.5.5.2 Analysis of the effect of switching from alglucosidase alfa to avalglucosidase alfa

Two analyses are planned to evaluate effect of avalglucosidase alfa in treatment experienced patients.

In the first analysis, a piecewise linear mixed effect model for % FVC (predicted) and 6MWT will be fitted for all patients who switched from alglucosidase alfa in PAP to avalglucosidase alfa in ETP. This analysis will be conducted to address a comment from CHMP in 2015. The model will include time as a continuous effect but assumed separate pre and post-switch slope for the time variable, with random intercept and slope at subject level. Similar piecewise linear model will be

generated for the patients in the avalglucosidase alfa arm who continue on the open label avalglucosidase alfa treatment. Model assumption on linearity of trend before and after 49 weeks will be assessed and the model maybe modified if there is good rational to justify an alternative model.

Due to potential period effect (patients may have FVC decline over time as the disease progresses), the estimated rate of change in % FVC (predicted) before and after switch may not be representative of the treatment effect of avalglucosidase alfa directly. The possibility of period effect will be assessed based on patients who were randomized to avalglucosidase alfa from study initiation without switching.

In order to support the label for the broader study population, an additional analysis will be conducted in % FVC (predicted) comparing the value from Week 49 to Week 97, by using a within-patient test in the treatment group who switched from alglucosidase alfa in PAP to avalglucosidase alfa in ETP. The change in % FVC (predicted) from Week 49 to Week 97 and the associated 95% CI will be calculated from the MMRM with fixed term for visit and random term for subject. This MMRM analysis will be based on % FVC (predicted) data from Week 49 and subsequent visits in ETP. Since database lock is planned once all data in PAP are observed, not all patients would have the week 97 visit by the cutoff date of PAP database lock. Thus, the majority of missing data at visits in ETP including Week 97 are expected to be attributable to an administrative reason that can be reasonably assumed missing at random. As explained by Brown and Prescott (28), within-subject correlations across the time points in MMRM allow observations at each time point to influence estimates of treatment effects at every other time point, and under missing at random MMRM improves efficiency than using the data from Week 49 and Week 97 visits only. In addition, summary statistics in % FVC (predicted) change from Week 49 will be provided by the randomized treatment groups at all ETP visits (including at Week 121) along with 95% CI for the within-subject comparison as appropriate. Summary statistics include the number of patients whose data are not missing, the sample mean with a 95% CI, the standard deviation, median, minimum, and maximum. The 95% CI at each visit will be constructed using only the data at that visit.

2.6 ANALYSES OF SAFETY DATA

In PAP analysis, the safety analyses will be carried out with patients by the actual treatment received, irrespective of the treatment the patient has been randomized to.

- In ETP analysis and overall evaluation of avalglucosidase alfa safety, the safety analyses will be carried out by the following 4 groups:
- Group 1 includes data in PAP and ETP among subjects who received avalglucosidase alfa in PAP and ETP,
- Group 2 includes data in ETP among subjects who received alglucosidase alfa in PAP and received avalglucosidase alfa in ETP,
- Group 3 includes data in PAP and/or EPT among subjects who received avalglucosidase alfa in that period,

- Group 4 includes data in PAP and/or EPT among pediatric subjects who received avalglucosidase alfa in that period.

General 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 the last available value before the start of first infusion in PAP or ETP, except possibly ECG parameters. The complete definitions of baseline are listed in Section 2.1.1.
- The potentially clinically significant abnormality (PCSA) values are defined as abnormal values considered medically important by the Sponsor according to predefined criteria/thresholds based on literature review and defined by the Sponsor for clinical laboratory tests, vital signs, and ECG (PCSA version dated May 2014 [Appendix A]).
  - PCSA criteria will determine which patients had at least 1 PCSA during the treatment epoch, taking into account all evaluations performed during the period, including nonscheduled or repeated evaluations.
  - The number of all such patients will be the numerator for the on-treatment PCSA percentage; the denominator will be the number of patients assessed for that given parameter in the treatment epoch in the safety population.
- The analysis of the safety variables will be essentially descriptive and no inferential testing is planned.

2.6.1 Analyses of adverse events

Generalities

The primary focus of adverse event reporting will be on treatment-emergent adverse events. Pretreatment and posttreatment 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, treatment-emergent, or post-treatment. 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 or post-treatment. Details on classification of adverse events with missing or partial onset dates are provided in Section 2.10.3.

Adverse event summaries will include number (n) and percentage (%) of patients experiencing an adverse event. The denominator for computation of percentages is the number of patients in the specific population within each treatment group. The number of events will be included in some summaries as well whenever appropriate.

Unless otherwise specified, sorting order will follow the internationally agreed SOC order, and further by decreasing number of events in PTs within SOCs in avalglucosidase alfa arm. When

more than one PT has same number of events, the order of presentation will be alphabetical in PTs.

Multiple occurrences of the same event in the same patient will be counted only once in the tables within a treatment phase.

Analysis of all treatment-emergent adverse events in PAP

The following treatment-emergent adverse event summaries will be generated for the safety population.

- An overview of treatment-emergent adverse events in PAP. Number (%) of patients will be provided by treatment group to include:
  - Treatment-emergent adverse events,
  - Serious treatment-emergent adverse events,
  - Treatment-emergent adverse events related to study drug,
  - Protocol-defined IARs,
  - Algorithm-defined IARs,
  - Treatment-emergent adverse events leading to death,
  - Treatment-emergent adverse events leading to permanent treatment discontinuation.
- All treatment-emergent adverse events during PAP by primary SOC, HLGT, HLT, and PT, showing number (%) of patients with at least 1 treatment-emergent adverse event sorted by the SOC internationally agreed order. The other levels (HLGT, HLT, PT) will be presented in alphabetical order.
- All treatment-emergent adverse events during PAP by primary SOC and PT sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC for the avalglucosidase alfa arm.
- All treatment-emergent adverse events during PAP presented by PT, sorted by decreasing incidence of PT in the avalglucosidase alfa arm.
- All treatment-emergent adverse events regardless of relationship during PAP by primary SOC, and PT,
- All treatment-emergent adverse events during PAP by maximal severity (ie, mild, moderate, or severe), presented by primary SOC and PT,
- Most common treatment-emergent adverse events during PAP by primary SOC and PT sorted by the internationally agreed SOC order and by decreasing incidence of PTs within each SOC for the avalglucosidase alfa arm. The most frequent TEAE are defined as those preferred terms with incidence of ≥5% in the avalglucosidase alfa treatment arm.
- Treatment-emergent adverse events with potential difference between treatment arms during PAP, by primary SOC and PT. The events with potential difference are defined as those preferred terms with difference in incidence rate of >10% between arms.

- For group 3 only, number (% per 1 year) of patients experiencing TEAEs by time of onset will be presented by SOC and PT.

Analysis of all treatment emergent serious adverse event(s) in PAP

- All treatment-emergent SAEs during PAP by primary SOC, and PT,
- All treatment-emergent SAEs regardless of relationship and related to IMP during PAP, by primary SOC, and PT.
- SAE listings will be provided as well.

Analysis of all treatment-emergent adverse event(s) leading to treatment discontinuation in PAP

- Treatment-emergent adverse events leading to treatment discontinuation during PAP, by primary SOC and PT,
- Treatment-emergent adverse events leading to treatment discontinuation during PAP presented by PT, sorted by decreasing incidence of PT in the avalglucosidase alfa arm,
- Listing of treatment-emergent adverse events leading to treatment discontinuation during PAP, including details on dose, severity, relationship, and outcome, etc.

Analysis of adverse events with AESIs in PAP

- Treatment emergent AESIs during PAP, by primary SOC and PT, showing the number (%) of patients, sorted by decreasing incidence of PT within the avalglucosidase alfa arm,
- Treatment emergent protocol-defined IARs during PAP, by primary SOC and PT, presented by overall number (%), as well as by start time of IAR in relation to infusion time (0-3 hrs, >3-24, >24-72, >72 hrs),
- Treatment emergent algorithm-defined IARs during PAP, by primary SOC and PT, presented by overall number (%), as well as by start time of IAR in relation to infusion time (0-3 hrs, >3-24, >24-72, >72 hrs),
- Treatment emergent protocol-defined IARs during PAP, by PT and worst severity, presented by overall number (%), as well as by start time of IAR in relation to infusion time (0-3 hrs, >3-24, >24-72, >72 hrs),
- Treatment emergent algorithm-defined IARs during PAP, by PT and worst severity, showing presented by overall number (%), as well as by start time of IAR in relation to infusion time (0-3 hrs, >3-24, >24-72, >72 hrs).

Analysis of all treatment-emergent adverse events in ETP

Similar analyses of the adverse events will be performed for the ETP as the PAP.

In addition, overall analysis of adverse events for the patients who receive avalglucosidase alfa, will be performed.

Analysis of pretreatment and post-treatment adverse events

- All pretreatment adverse events by primary SOC and PT,
- All adverse events with onset during the post treatment period of PAP or ETP by primary SOC and PT,
- All posttreatment SAEs with onset during the post treatment period of PAP or ETP by primary SOC and PT.

Anaphylactic/hypersensitivity reactions and immune-mediated reactions

A comprehensive programming search of AEs which meet the Standard MedDRA Query (SMQ) criteria for hypersensitivity and anaphylactic reaction will be used to identify adverse events that potentially are associated with symptoms of anaphylactic and hypersensitivity reaction. Results of this search will be provided in summary tables and in a by patient listing by group. The most recent version of MedDRA SMQ will be used at the time of analysis. A medical review of these cases will be performed.

A listing of potential immune mediated reactions will be provided. Search criteria will include but not limited to the MedDRA Preferred Terms of glomerulonephritis, nephrotic syndrome, proteinuria, hematuria, vasculitis, serositis, myocarditis, skin lesion, skin necrosis, arthralgia, arthritis, myalgia, arthropathy, lymphadenopathy, serum sickness, type III immune complex mediated reaction and influenza-like symptoms. A medical review of these cases will be performed.

The exposure adjusted incidence rate (EAIR) for selected TEAEs will be calculated as number of patients with the specific TEAEs divided by patient years. For the calculation of EAIR, for patients with event, the patient year is calculated as time from first treatment infusion to the time of first event; for patients without event, it is calculated as time from first treatment infusion to the last administration + 4 weeks (28 days). The rate per 100 patient years will be displayed for all the categories described below.

The EAIRs will be calculated for the following TEAEs:

- Overview of TEAEs (Including patients with any TEAE, patients with any treatment-emergent SAE, patients with any TEAE potentially related to avalglucosidase alfa, patients with any treatment-emergent severe AE, patients with any TEAE leading to death, patients with any TEAE leading to permanent treatment discontinuation). SOC and PT will not be displayed in this table;
- TEAEs by primary SOC and PT.

2.6.2 Evaluation of ADA on Relevant Safety Parameters

To evaluate the effect of ADA and NAb on AEs, the number and percentage of patients experiencing any TEAEs, any treatment-emergent SAEs, or any IARs, hypersensitivity (narrow SMQ), and anaphylaxis (narrow SMQ) will be presented by the following sub-categories: 1. ADA status (ever positive, always negative) 2. ADA peak titer category (always negative, peak titer

100-800, 1600-6400, ≥12,800) 3. ADA response type (always negative, transient response (if occur), and persistent responses subcategories: low response, intermediate response, high ADA response and tolerized at defined timepoints. 4. Correlation of frequency of IAR and ADA peak titer. 5. Neutralizing antibody status (always negative, ever positive). Pertinent groups will be assessed to address both treatment naïve and switch patients.

2.6.3 Deaths

The following summaries of deaths will be generated for the safety population.

- Number (%) of patients who die by study period (on-study, PAP treatment epoch, ETP treatment epoch, post treatment epoch) and reasons for death,
- Deaths in nonrandomized patients or randomized but not treated patients,
- Treatment-emergent adverse events leading to death (death as an outcome on the adverse event case report form page as reported by the Investigator) by primary SOC, and PT. Corresponding listing will be provided as well.

2.6.4 Analyses of laboratory variables

The summary statistics (including number, mean, median, SD, minimum and maximum) of all laboratory variables (laboratory values and changes from baseline) will be calculated for each visit or study assessment (baseline and each postbaseline time point, by treatment group. This section will be organized by biological functions.

The number (%) of patients with laboratory abnormalities of all laboratory variables, when applicable, will be summarized for each visit or study assessment. Comparisons of the worst laboratory abnormalities during PAP with the baseline will be presented for each laboratory parameter, when applicable. Similar analyses will be conducted combining both PAP and ETP together, but only for the patients in avalglucosidase alfa group.

The incidence of PCSAs (list provided in Appendix A) at any time during the treatment-emergent adverse event period will be summarized by biological functions and treatment groups.

- The PCSA analyses will be conducted for PAP primarily. Similar analyses will be conducted combining both PAP and ETP together, but only for the patients in avalglucosidase alfa group.

Drug-induced liver injury

The liver function tests, namely AST, ALT, alkaline phosphatase, and total bilirubin, are used to assess possible drug-induced liver toxicity. The following analyses will be performed:

- Time to onset of the initial ALT or AST elevation (>3 × ULN) and total bilirubin elevation (>2 × ULN), whichever comes first will be analyzed using Kaplan Meier estimates by study cohort and treatment arm, if necessary.
- A graph of distribution of peak values of ALT versus peak values of total bilirubin (in logarithmic scale or in the scale of x ULN if appropriate) will also be presented. The graph

will be divided into 4 quadrants with a vertical line corresponding to 3 x ULN for ALT and a horizontal line corresponding to 2 x ULN for total bilirubin if necessary. A similar graph will be provided for peak values of AST versus peak values of total bilirubin.

- Listing of possible Hy's law cases identified by treatment group (eg, patients with any elevated ALT or AST >3 x ULN, and associated with an increase in bilirubin ≥2 x ULN) with ALT, AST, alkaline phosphatase and total bilirubin if necessary.
- Summary of the incidence of liver-related adverse events by treatment group if necessary. The selection of preferred terms will be based on the hepatic disorder SMQ.
- Additional similar analyses will be performed for the ETP and combined PAP and ETP period.

2.6.5 Analyses of vital sign variables

The summary statistics (including number, mean, median, SD, minimum and maximum) of all vital signs variables (central laboratory values and changes from baseline) will be calculated for each scheduled visit or study assessment (baseline, each postbaseline time point) by treatment group.

The incidence of PCSAs (list provided in Appendix A) based on worst value at any time during the treatment-emergent adverse event period will be summarized by treatment groups.

The PCSA analyses will be conducted for both PAP and ETP separately. Similar analyses will be conducted for combined PAP and ETP period.

2.6.6 Analyses of electrocardiogram variables

The summary statistics (including number, mean, median, SD, minimum and maximum) of all ECG variables (laboratory values and changes from baseline) will be calculated for each visit or study assessment (baseline, each postbaseline time point) by treatment group.

The incidence of PCSAs (list provided in Appendix A) based on worst value at any time during the treatment-emergent adverse event period will be summarized by treatment groups.

The PCSA analyses will be conducted for both PAP and ETP separately. Similar analyses will be conducted for combined PAP and ETP period.

2.6.7 Analyses of physical examinations

The summary statistics of all physical examination variables will be summarized descriptive by treatment group if appropriate.

2.6.8 Analyses of Immunogenicity parameters

2.6.8.1 ADA incidence and characterization

ADA status: ADA seroconversion is classified as always negative and ever positive. Baseline ADA status will be reported as negative or pre-existing ADA at initiation of treatment as well as at drug switch after PAP period.

The following ADA incidence will be summarized descriptively for each treatment group during the PAP and baseline will be reassessed at the beginning of ETP for the alglucosisdase alfa treated patients that switch to avalglucosidase alfa.

- Treatment emergent ADA incidence, defined as
  - 100 x (treatment boosted + treatment induced ADA positive patients)/(number of evaluable patients),
- Treatment induced ADA incidence, defined as,
  - 100 x (treatment induced ADA positive patients)/(number of evaluable patients with ADA negative at baseline),
- Treatment boosted ADA incidence, defined as
  - 100 x (treatment boosted ADA positive patients)/(number of evaluable patients with ADA positive at baseline),

2.6.8.2 Duration of ADA

The kinetics and duration of the immune responses will be analyzed as follow:

- Onset time of ADA will be analyzed descriptively using summary statistics minimum, Q1, median, Q3 and maximum.
- Duration of ADA will be analyzed descriptively using summary statistics minimum, Q1, median, Q3 and maximum. It will only be calculated for the patients with at least two ADA positive samples. The median duration and the quartiles will be reported.

2.6.8.3 ADA titers

- ADA peak titer, last titer, and geometric mean titer will be summarized. Graphs of ADA titer over time and boxplots of the highest post-baseline ADA titer will be provided. The analyses will be performed for PAP, ETP and combined PAP and ETP periods. Both treatment naïve and switch patients will be evaluated separately.

2.6.8.4 ADA Response Classification

Response type classification will be provided separately for treatment naïve and switch patients as follows:

- The number and percent of transient ADA response will be summarized descriptively. This will be performed for the patients with at least two post baseline samples where the last sampling timepoint is negative.
- The number and percent of persistent ADA response and its subclassifications will be summarized descriptively. This will be performed for the patients with at least two post baseline samples which are separated by at least 16 weeks.
- The number and percent of indeterminate ADA response will be summarized descriptively. This will be performed for the patients with at least one post baseline sample.

2.6.8.5 Neutralizing ADA

- Incidence of neutralizing antibodies (inhibition of enzyme activity and inhibition of enzyme uptake) will be reported for both treatment naïve and switch patients
- Onset time of neutralizing ADA will be analyzed descriptively using summary statistics minimum, Q1, median, Q3 and maximum.
- Duration of neutralizing ADA will be analyzed descriptively using summary statistics minimum, Q1, median, Q3 and maximum. It will only be calculated for the patients with at least two neutralizing ADA positive samples.

2.6.8.6 Cross-reactivity evaluation

Cross reactivity will be assessed at Week 25 and Week 49 during the PAP. Patients from the avalglucosidase alfa treatment arm who are positive for anti-avalglucosidase alfa antibodies will be tested to determine if the antibodies cross-react with alglucosidase alfa and patients from the alglucosidase treatment arm who are positive for anti- alglucosidase alfa antibodies will be tested to determine if the antibodies cross-react with avalglucosidase alfa at these timepoints. ADA titer data for the cross-reactivity assessment will be analyzed descriptively by treatment group using summary statistics minimum, Q1, median, Q3 and maximum to both molecules. A patient listing of titers will also be provided. In addition, patients from the avalglucosidase alfa treatment arm who are positive for anti-avalglucosidase alfa antibodies will be further evaluated by immunodepletion to determine cross-reactivity. The percentage of patients with cross-reactive antibodies will be provided.

2.6.9 Association of ADA with PK

The following analysis will be considered: 1. Within Subject level AUC change from baseline to end of the double-blind treatment period will be plotted by peak titer category. Gender will be separated by different color. 2. Between subject comparison of ADA-positive vs. ADA negative. Plot of AUC at baseline compared to timepoints where full PK assessment is available. Patients will be evaluated by titer categories based on titer category at the time of PK assessment. Patients will be separated based on treatment groups. 3. Summary table to include AUC and % change from Day 1 at each scheduled visit by peak titer categories. 4. Data will be assessed for switch patients as appropriate.

2.6.10 Association of ADA and PD marker

Hex-4 is a clinically relevant PD marker that is related to the drug's mechanism of action. Summary table of number (%) of patients with elevated urinary hex-4 by titer value at the specified visit (eg, peak titer categories) over time will be provided. Within-subject changes from baseline in PD marker will be evaluated with respect to the changes in the status and/or titer value of ADA/NAb (neutralizing antibodies).

2.6.11 Association of ADA with selected efficacy

Correlation analysis will be performed between immunogenicity (ADA titers, response categories and neutralizing ADA) and FVC, 6MWT, (raw and percent predicted), MIP, MEP.

2.7 SUMMARY OF PHARMACOKINETIC DATA

Pharmacokinetic parameters will be summarized using descriptive statistics by study weeks. Plasma concentration data will be summarized using descriptive statistics by treatment weeks. The population pharmacokinetic results will be documented in a separate report.

2.8 ANALYSES OF PHARMACODYNAMICS ENDPOINT DATA

Creatinine-normalized urine Hex4 levels will be summarized using descriptive statistics at each scheduled study visit. Observed measurements as well as the change from baseline will be summarized. If a linear trend in the change of urine Hex4 levels is observed, a longitudinal model may be employed to model change from baseline over time. In addition, 95% confidence intervals of changes will be presented.

Correlation between PK endpoints and urine Hex4 levels will be explored as appropriate.

Similar analyses will be carried out to explore the relationship between other PD endpoints, biomarkers, efficacy assessments, and exploratory endpoints.

2.9 ANALYSES OF QUALITY OF LIFE/HEALTH ECONOMICS VARIABLES

The 18 questions for the R-PAct will be scored according to the 0 (unable to perform) to 2 (able to perform without difficulty) Likert scale and will have a descriptive analysis by treatment group.

Scores for the PDSS and PDIS will be calculated separately for each questionnaire, as well as a composite score for both questionnaires together. Note that these are exploratory endpoints recently developed, and the scoring method for these scales is still under development. These will be finalized prior to study unblinding and described in a separate document.

2.10 DATA HANDLING CONVENTIONS

2.10.1 General conventions

In general, the baseline value is defined as latest value prior to the start of first infusion in PAP or ETP. In the case an assessment performed on the same date as the first infusion date, but it is impossible to determine the evaluation time relative to first infusion start time, the evaluation time will be assumed to be following the protocol-defined schedule.

2.10.2 Data handling conventions for primary and secondary efficacy variables

The baseline value for the efficacy variables is the last non-missing measurement prior to the first infusion of the study drug. The ETP baseline value for the efficacy variables is the last non-missing measurement prior to the first infusion of avalglucosidase alfa in ETP.

If the baseline value is missing, the patient will be excluded from the MMRM efficacy analyses under MAR assumption. The missing data handling for the key efficacy endpoints are described in Section 2.1.3.

2.10.3 Missing data handling in data presentation

In general, missing baselines will not be imputed. The following approaches are default methods for missing data handling in summary tables.

- Categorical data at baseline will be summarized for each treatment group using counts (n) and percentages (%). Denominator will be the analysis population specified for the summary, unless otherwise specified. Missing data may be presented as a separate category.
- 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 IMP is equal to the date of last administration reported on drug exposure case report form page. If all the infusion dates are missing, then the duration is missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of missing/partial dates for adverse events or concomitant medications

Missing or partial adverse event onset dates and times will be imputed so that if the partial adverse event onset date/time information does not indicate that the adverse event started prior to treatment or after the treatment-emergent adverse event period, the adverse event will be classified as treatment-emergent. No imputation of adverse event end dates/times will be

performed. These data imputations are for categorization purpose only and will not be used in listings. No imputation is planned for date/time of adverse event resolution.

No imputation for 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 or concomitantly, it will be considered a prior, concomitant, and post-treatment medication.

Handling of missing or partial birth date for calculation of age

Complete missing of birth date is not expected for the study. However, there could be rare situations in which the birth day or month is not provided for the subject due to confidentiality. In this case, the missing day will be imputed as 15th of the month, and missing month will be imputed as June as default. The adjustment maybe needed in the case of conflicting with study dates. These adjustments should be very rare and will be documented in the data specification document.

Handling of adverse events when date and time of first investigational medicinal product administration is missing

When the date and time of the first IMP administration is missing, all adverse events that occurred on or after the day of randomization should be considered as treatment-emergent adverse events. The exposure duration should be kept as missing.

The last dose intake should be clearly identified in the case report form and should not be approximated by the last returned package date.

Handling of missing assessment of relationship of adverse events to investigational medicinal product

If the assessment of the relationship to IMP is missing, then the relationship to IMP in the frequency tables is considered as possibly related, but no imputation should be done at the data level.

Handling of missing severity of adverse events

If the severity is missing for 1 of the treatment-emergent occurrences of an adverse event, the maximal severity on the remaining occurrences will be considered. If the severity is missing for all the occurrences, a “missing” category will be added in the summary table.

Handling of potentially clinically significant abnormalities

If a patient has a missing baseline he will be grouped in the category “normal/missing at baseline.”

For PCSAs with 2 conditions, one based on a change from baseline value or a normal range and the other on a threshold value, with the first condition being missing, the PCSA will be based only on the second condition.

For a PCSA defined on a threshold and/or a normal range, this PCSA will be derived using this threshold if the normal range is missing; eg, for eosinophils the PCSA is >0.5 GIGA/L or > ULN if ULN ≥0.5 GIGA/L. When ULN is missing, the value 0.5 should be used.

Measurements flagged as invalid by the laboratory will not be summarized or taken into account in the computation of PCSA values.

Handling of ADA titer with missing or non-numerical values

If the ADA titer is reported as "< value", then the actual value is imputed as this value. For example, "<100" will be imputed as 100. A negative ADA status will be assumed as a value of 0 (will be excluded when geometric mean of the group needs to be calculated).

2.10.4 Study day calculation

Based on study protocol, study Day 1 is the date of first study infusion.

Study day for a given assessment is defined as the assessment date - date of first study infusion + 1 if the assessment date is on or after Day 1, or assessment date - date of first study infusion if the assessment date is before Day 1.

If a randomized patient is not treated (first infusion date is missing), study Day 1 is the date of randomization.

2.10.5 Windows for time points

The visit windows for efficacy endpoints specified in Table 16 are defined in following table. For analysis purpose, a larger window (as shown in the last column of Table 14) will be used. This will be applicable for those analyses requiring visit information.

Table 14 - Study visit nominal visit days and windows for efficacy endpoints

Scheduled Visit	Visit window defined in protocol in study days	Visit window for analysis purpose in study days
Screening/Baseline	Day -14 to -1	(-28, the day before 1st infusion)
Week 13	Day 85 ±14	(the day after 1st infusion, Day 85 +41)
Week 25	Day 169 ±14	Day 169 (-42, +41)
Week 37	Day 253 ±14	Day 253 (-42, +41)
Week 49	Day 337 ±14	(Day 337 -42, the day prior to the 1st ETP treatment or Day 337 + 41 if 1st ETP treatment is missing)
Week 61	Day 421 ±14	(the day after the 1st ETP treatment, Day 421 +41)
Week 73	Day 505 ±14	Day 505 (- 42, +83)
Week 97	Day 673 ±14	Day 673 (-84, +83)
Week 121	Day 841 ±14	Day 841 (-84, +83)

Scheduled Visit	Visit window defined in protocol in study days	Visit window for analysis purpose in study days
Week 145	Day 1009 ±14	Day 1009 (-84, +83)
Week 169	Day 1177 ±14	Day 1177 (-84, +83)
Week 193	Day 1345 ±14	Day 1345 (-84, +83)
Week 217	Day 1513 ±14	Day 1513 (-84, +83)
Week 241	Day 1681 ±14	Day 1681 (-84, +83)
Week 265	Day 1849 ±14	Day 1849 (-84, +83)

The analysis visit will be defined by comparing the actual visit date with the nominal (or target) date and the corresponding analysis window from Table 14 . If more than one non-missing values are assigned to the same visit, then the one closest to the target date will be used in the by-visit analysis. Multiple values on the same date will be averaged first. If two assessments are on different date but equal distance from the target date, the later date will be used.

For the safety analyses (laboratory, ECG and vital signs), the nominal visit will be used for the by-visit type of analyses.

2.10.6 Unscheduled visits

Unscheduled visit measurements of laboratory data, vital signs, and ECG will not be included in the by-visit summaries, but will be used for computation of baseline, worst values, and summary of PCSAs.

2.10.7 Pooling of centers for statistical analyses

Investigation of the effects of geographic regions may be performed on an exploratory basis.

3 INTERIM ANALYSIS

No interim analyses are planned during the double-blinded PAP. After the primary analysis for PAP, interim analyses may be performed during the open-label extension period to provide additional information for regulatory purpose. However, those additional analyses after PAP will be considered for supportive purpose.

4 DATABASE LOCK

The database lock for the primary efficacy analyses is planned to be approximately 30 days after last patient last visit in PAP.

The final database lock for the study is planned to be approximately 30 days after the last patient last visit in ETP.

5 SOFTWARE DOCUMENTATION

All summaries and statistical analyses will be generated using SAS Version 9.4 or higher.

6 REFERENCES

1. Lachmann R, Schoser B. The clinical relevance of outcomes used in late-onset Pompe disease: can we do better? Orphanet J Rare Dis. 2013;8:160.

2. Lei Nie, Guoxing Soon. A covariate-adjustment regression model approach to noninferiority margin definition, Statist Med. 2010;29:1107-113.

3. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. Eur Respir J. 2012;40(6):1324-43.

4. Biomedical Systems. Gneral Pulmonary Software Specifications and User Requirements. Sponsor: Sanofi-Genzyme. Protocol: EFC14028.

5. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med. 2002;166(1):111-7.

6. Gibbons WJ, Fruchter N, Sloan S, Levy RD. Reference values for a multiple repetition 6-minute walk test in healthy adults older than 20 years. J Cardiopulm Rehabil. 2001;21(2):87-93.

7. Geiger R, Strasak A, Treml B, Gasser K, Kleinsasser A, Fischer V, et al. Six-minute walk test in children and adolescents. J Pediatr. 2007;150(4):395-9.

8. Montes J, McDermott MP, Martens WB, Dunaway S, Glanzman AM, Riley S, et al. Six-Minute Walk Test demonstrates motor fatigue in spinal muscular atrophy. Neurology. 2010;74(10):833-8.

9. Evans JA, Whitelaw WA. The Assessment of Maximal Respiratory Mouth Pressures In Adults. Respir Care. 2009;54(10):1348-59.

10. Wilson SH, Cooke NT, Edwards RH, Spiro SG. Predicted normal values for maximal respiratory pressures in caucasian adults and children. Thorax. 1984;39(7):535-8.

11. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. Phys Ther. 1996;76(3):248-59.

12. The National Isometric Muscle Strength (NIMS) Database Consortium. Muscular weakness assessment: Use of normal isometric strength data. Arch Phys Med Rehabil. 1996;77:1251-5.

13. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. Arch Phys Med Rehabil. 1997;78:26-32.

14. Beenakker EA, van der Hoeven JH, Fock JM, Maurits NM. Reference value of maximum isometric muscle force obtained in 270 children aged 4-16 years by hand-held dynamometry. Neuromuscular Disord. 2001;11:441-6.

15. Bäckman E, Odenrick P, Henriksson KG, Ledin T. Isometric muscle force and anthropometric values in normal children aged between 3.5 and 15 years. Scand J Rehabil Med. 1989;21(2):105-14.

16. van Capelle CI, van der Beek NA, de Vries JM, van Doorn PA, Duivenvoorden HJ, Leshner RT, et al. The quick motor function test: a new tool to rate clinical severity and motor function in Pompe patients. J Inherit Metab Dis. 2012;35(2):317-23.

17. Ware J Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996 Mar;34(3):220-33.

18. Quality Metric Incorporated. User's manual for the SF-12v2 health survey, third edition.

19. Angelini C, Semplicini C, Tonin P, Filosto M, Pegoraro E, Sorarù G, et al. Progress in enzyme replacement therapy in glycogen storage disease type II. Ther Adv Neurol Disord. 2009;2(3):143-53.

20. Russell DJ, Rosenbaum PL, Avery LM, Lane M. Gross motor function measure (GMFM-66 & GMFM-88) user's manual. Cambridge University Press. 2002.

21. Palisano RJ, Rosenbaum P, Bartlett D, Livingston MH. Content validity of the expanded and revised Gross Motor Function Classification System. Dev Med Child Neurol. 2008;50(10):744-50.

22. Mathiowetz V, Wiemer DM, Federman SM. Grip and pinch strength: Norms for 6- to 19-year-olds. Am J Occup Ther. 1986 Oct;40(10):705-11.

23. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-36.

24. Varni JW, Burwinkle TM, Seid M. The PedsQL TM 4.0 as a school population health measure: feasibility, reliability, and validity. Qual Life Res. 2006;15(2):203-15.

25. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care. 2001;39(8):800-12.

26. van der Beek NA, Hagemans ML, van der Ploeg AT, van Doorn PA, Merkies IS. The Rasch-built Pompe-specific activity (R-PAct) scale. Neuromuscul Disord. 2013;23(3):256-64.

27. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the Six Minute Walk test in chronic lung disease patients. Am J Respir Crit Care Med. 1997;155(4):1278-82.

28. Brown H, Prescott R. Applied Mixed Models in Medicine, 2nd Edition. John Wiley & Sons, May 1, 2006.

29. Sanofi Research and Development Quality Document. Analysis and reporting of safety data from Clinical Trials through the Clinical Study Report. Version 3. 2014.

7 LIST OF APPENDICES

Appendix A Potentially clinically significant abnormalities criteria

Table 15 - PCSA abnormalities (29)

Measures	Adult Criteria	Pediatric Criteria
Liver function tests		
ALT	>3 x ULN >5 x ULN >10 x ULN >20 x ULN	≥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	≥3 x ULN ≥5 x ULN ≥10 x ULN ≥20 x ULN
Alkaline Phosphatase	>1.5 x ULN	≥1.5 x ULN
Total Bilirubin	>1.5 x ULN >2 x ULN	≥1.3 x ULN
ALT and Total Bilirubin	ALT >3 x ULN and Total Bilirubin >2 x ULN	ALT ≥3 x ULN and Total Bilirubin ≥2 x ULN
Hematology		
White Blood Cell (WBC)	<3.0 GIGA/L (non-Black), <2.0 GIGA/L (Black), ≥16.0 GIGA/L	Birth/0 to 27 days old (Neonates) <4.0 GIGA/L >25.0 GIGA/L 28 days/1 month to 23 months old (Infants) <4.0 GIGA/L >20.0 GIGA/L 24 months/2 years to <6 years old (Children) >3.0 GIGA/L >16.0 GIA/L

Measures	Adult Criteria	Pediatric Criteria
		6 to <12 years old (Children) <5.0 GIGA/L >17.0 GIGA/L 12 to 16/18 years old (Adolescents) <4.5 GIGA/L >13.5 GIGA/L
Lymphocytes	>4.0 GIGA/L	Birth/0 to 27 days old (Neonates) <1.2 GIGA/L >17.0 GIGA/L 28 days/1 month to 23 months old (Infants) <2.0 GIGA/L >13.5 GIGA/L 24 months/2 years to <6 years old (Children) <1.0 GIGA/L >9.5 GIGA/L 6 to <12 years old (Children) <1.0 GIGA/L >8.0 GIGA/L 12 to 16/18 years old (Adolescents) <0.6 GIGA/L >6.0 GIGA/L

Measures	Adult Criteria	Pediatric Criteria
Neutrophils	<1.5 GIGA/L (non-Black) <1.0 GIGA/L (Black)	Birth/0 to 27 days old (Neonates) <4.0 GIGA/L (1 day old) <1.5 GIGA/L (2 – 7 days old) <1.25 GIGA/L (>7 day – 1 month old) >1 ULN 28 days/1 month to 23 months old (Infants) <1.0 GIGA/L (1 – 3 months) <1.2 GIGA/L (3 – 24 months) >1 ULN 24 months/2 years to <6 years old (Children) <1.2 GIGA/L >1 ULN 6 to <12 years old (Children) <1.2 GIGA/L >1 ULN 12 to 16/18 years old (Adolescents) <1.2 GIGA/L >1 ULN
Monocytes	>0.7 GIGA/L	
Basophils	>0.1 GIGA/L	
Eosinophils	>0.5 GIGA/L or >ULN if ULN ≥0.5 GIGA /L	>0.5 GIGA/L Or >ULN if ULN >0.5 GIGA/L

Measures	Adult Criteria	Pediatric Criteria
Hemoglobin	Males: ≤115 g/L (≤7.14 mmol/L), ≥185 g/L (≥11.48 mmol/L) Females: ≤95 g/L (5.9 mmol/L), ≥165 g/L (10.24 mmol/L) Decrease from Baseline: ≥20 g/L (1.24 mmol/L)	Birth/0 to 27 days old (Neonates) <86 mmol/L or 12.0 g/dL or any decrease > 0.31 mmol/L or 2 g/dL 28 days/1 month to 23 months old (Infants) <1.40 mmol/L or 9.0 g/dL or any decrease > 0.31 mmol/L or 2 g/dL 24 months/2 years to <16/18 years old (Children, Adolescents) <1.55 mmol/L or 10.0 g/dL or any decrease > 0.31 mmol/L or 2 g/dL
Hematocrit	Males : ≤0.37 v/v, ≥0.55 v/v Females : ≤0.32 v/v, ≥0.5 v/v	Birth/0 to 27 days old (Neonates) <0.39 l/l or 40% >0.61 l/l or 47% 28 days/1 month to 23 months old (Infants) <0.29 l/l or 29% >0.42 l/l or 42% 24 months/2 years to <16/18 years old (Children, Adolescents) <0.32 l/l or 32% >0.47 l/l or 47%
RBC	≥6 TERA/L	
Platelets	<100 GIGA/L ≥ 700 GIGA/L	<100 GIGA/L >700 GIGA/L
ECG – PCSA criteria		
HR	<50 bpm <50 bpm and decrease from baseline ≥20 bpm <40 bpm <40 bpm and decrease from baseline ≥20 bpm <30 bpm <30 bpm and decrease from baseline ≥20 bpm	Birth/0 to 27 days old (Neonates) ≤90 bpm and decrease from baseline≥20 bpm ≥190 bpm and increase from baseline ≥20 bpm 28 days/1 month to 23 months old (Infants) ≤80 bpm and decrease form baseline

Measures	Adult Criteria	Pediatric Criteria
		≥20 bpm
	>90 bpm	≥175 bpm and increase from baseline
	>90 bpm and increase from baseline ≥20bpm	≥20 bpm
	>100 bpm	24 months/2 years to <6 years old (Children)
	>100 bpm and increase from baseline ≥20bpm	≤75 bpm and decrease from baseline
	>120 bpm	≥20 bpm
	>120 bpm and increase from baseline ≥20 bpm	≥140 bpm and increase from baseline
		≥20 bpm
		6 to <12 years old (Children)
		≤50 bpm and decrease from baseline
		≥20 bpm
		≥120 bpm and increase from baseline
		≥20 bpm
		12 to 16/18 years old (Adolescents)
		≤50 bpm and decrease from baseline
		≥20 bpm
		≥120 bpm and increase from baseline
		≥20 bpm
PR	>200 ms	Birth/0 to 27 days old (Neonates) ≥120 ms
	>200 ms and increase from baseline ≥25%	28 days/1 month to 23 months old (Infants) ≥140 ms
	> 220 ms	24 months/2 years to <6 years old (Children) ≥160 ms
	>220 ms and increase from baseline ≥25%	6 to <12 years old (Children) ≥170 ms
	> 240 ms	12 to 16/18 years old (Adolescents) ≥180 ms
	> 240 ms and increase from baseline ≥25%	
QRS	>110 ms	Birth/0 to 27 days old (Neonates) ≥85 ms
	>110 msec and increase from baseline ≥25%	28 days/1 month to 23 months old (Infants) ≥85 ms
	>120 ms	24 months/2 years to <6 years old (Children) ≥95 ms
	>120 ms and increase from baseline ≥25%	6 to <12 years old (Children) ≥100 ms
		12 to 16/18 years old (Adolescents) ≥110 ms
QTc (either QTcF or QTcB)	Absolute values (ms	Birth/0 to <12 years old (Neonates, Infants, Children)
		Absolute values (ms)

Measures	Adult Criteria	Pediatric Criteria
	>450 ms	Borderline: 431 – 450 ms
	>480 ms	Prolonged*: >450 ms
	>500 ms	Additional: ≥500 ms
		AND
	Increase from baseline	Increase from baseline
	30-60 ms	Borderline: Increase from baseline 30 – 60 ms
	>60 ms	Prolonged*: Increase from baseline >60 ms
		12 to 16/18 years old (Adolescents)
		Absolute values (ms)
		Borderline: 431 – 450 ms (Boys);451 – 470 ms (Girls)
		Prolonged*: >450 ms (Boys);>470 ms (Girls)
		Additional: ≥500 ms
		AND
		Increase from baseline
		Borderline: Increase from baseline30 – 60 ms
		Prolonged*: Increase from baseline >60 ms
*QTc prolonged and ΔQTc >60 ms are the PCSA to be identified in individual subjects/patients listings.		
Clinical chemistry		
Creatinine	≥150 µmol/L (Adults)	Birth/0 to <6 years old (Neonates, Infants, Children)
	≥30% increase from baseline	>53 µmol/L or 0.6 mg/dL
	≥100% increase from baseline	6 years to <12 years old (Children)
		≥90 µmol/L or 1.1 mg/dL
		12 years to 16/18 years old (Adolescents)
		≥132 µmol/L or 1.5 mg/dL
Blood Urea Nitrogen	≥17 mmol/L	Birth/0 to 27 days old (Neonates)
		≥4.3 mmol/L or 12 mg/dl
		28 days/1 month to 16/18 years old (Infants, Children, Adolescents)
		≥6.4 mmol/L or 18 mg/dl
Chloride	<80 mmol/L	≤80 mmol/L
	>115 mmol/L	≥115 mmol/L

Measures	Adult Criteria	Pediatric Criteria
Sodium	≤129 mmol/L	≤129 mmol/L
	≥160 mmol/L	≥150 mmol/L
Potassium	<3 mmol/L	Birth/0 to 27 days old (Neonates)
	≥5.5 mmol/L	≤3.0 mmol/L
		≥7.0 mmol/L
		28 days/1 month to 23 months old (Infants)
		≤3.5 mmol/L
		≥6.0 mmol/L
		24 months/2 years to 16/18 years old (Children, Adolescents)
		≤3.5 mmol/L
		≥5.5 mmol/L
Glucose		
Hypoglycemia	≤3.9 mmol/L and <LLN	<2.7 mmol/L
Hyperglycemia	≥11.1 mmol/L (unfasted);	≥7 mmol/L (fasted after >12 hours of fast);
	≥7 mmol/L (fasted)	≥10.0 mmol/L (unfasted)
Albumin	≤25 g/L	
Vital signs		
Heart rate	≤50 bpm and decrease from baseline ≥20 bpm	Birth/0 to 27 days old (Neonates)
	≥120 bpm and increase from baseline ≥20 bpm	≤90 bpm and decrease from baseline ≥20 bpm
		≥190 bpm and increase from baseline ≥20 bpm
		28 days/1 month to 23 months old (Infants)
		≤80 bpm and decrease form baseline ≥20 bpm
		≥175 bpm and increase from baseline ≥20 bpm
		24 months/2 years to <6 years old (Children)
		≤75 bpm and decrease from baseline ≥20 bpm
		≥140 bpm and increase from baseline ≥20 bpm
		6 to <12 years old (Children)
		≤50 bpm and decrease from baseline ≥20 bpm
		≥120 bpm and increase from baseline ≥20 bpm
		12 to 16/18 years old (Adolescents)
		≤50 bpm and decrease from baseline ≥20 bpm

Measures	Adult Criteria	Pediatric Criteria
		≥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	Birth/0 to 27 days old (Neonates) ≤60 mmHg and decrease from baseline ≥20 mmHg ≥85 mmHg and increase from baseline ≥20 mmHg 28 days/1 month to 23 months old (Infants) ≤70 mmHg and decrease from baseline ≥20 mmHg ≥98 mmHg and increase from baseline ≥20 mmHg 24 months/2 years to <6 years old (Children) ≤70 mmHg and decrease from baseline ≥20 mmHg ≥101 mmHg and increase from baseline ≥20 mmHg 6 to <12 years old (Children) ≤80 mmHg and decrease from baseline ≥20 mmHg ≥108 mmHg and increase from baseline ≥20 mmHg 12 to 16/18 years old (Adolescents) ≤90 mmHg and decrease from baseline ≥20 mmHg ≥119 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	Birth/0 to 27 days old (Neonates) ≤34 mmHg and decrease from baseline ≥10 mmHg ≥50 mmHg and increase from baseline ≥10 mmHg 28 days/1 month to 23 months old (Infants) ≤34 mmHg and decrease from baseline ≥10 mmHg ≥54 mmHg and increase from baseline ≥10 mmHg 24 months/2 years to <6 years old (Children) ≤34 mmHg and decrease from baseline ≥10 mmHg ≥59 mmHg and increase from baseline ≥10 mmHg 6 to <12 years old (Children) ≤48 mmHg and decrease from baseline ≥10 mmHg ≥72 mmHg and increase from baseline ≥10 mmHg 12 to 16/18 years old (Adolescents) ≤54 mmHg and decrease from baseline ≥10 mmHg ≥78 mmHg and increase from baseline ≥10 mmHg

Appendix B Schedule of the efficacy assessment

Table 16 - Schedule of efficacy assessment

		PAP				ETP						
	Screnning/Baseline	**W13**	**W25**	**W37**	**W49**	**W61**	**W73**	**W97**	**W121**	**W145**	**Every 24 weeks**	**End of treatment**
PFT	X	X	X	X	X	X	X	X	X	X	X	X
6MWT	X	X	X	X	X	X	X	X	X	X	X	X
Hand-held dynamometry	X	X	X	X	X	X	X	X	X	X		
GMFM-88/GMFCS	X	X	X	X	X	X	X	X	X	X		
QMFT	X	X	X	X	X	X	X	X	X	X		
GSGC	X	X	X	X	X	X	X	X	X	X		
SF-12	X	X	X	X	X	X	X	X	X	X	X	X
EQ-5D-5L	X	X	X	X	X	X	X	X	X	X	X	X
PedsQL	X	X	X	X	X	X	X	X	X	X	X	X
PDSS/PDIS	X	X	X	X	X	X	X	X	X	X	X	X
R-PAct	X	X	X	X	X	X	X	X	X	X	X	X

Signature Page for
efc14028-16-1-9-sap

Approve & eSign	[Name] Clinical 15-Apr-2020 14:31:45 GMT+0000

Approve & eSign	[Name] Clinical 15-Apr-2020 14:39:43 GMT+0000