Official Title:	Statistical Analysis Plan: A Phase 3 Double-blind Randomized Study to Assess the Efficacy and Safety of Intravenous ATB200 Co-administered With Oral AT2221 in Adult Subjects With Late-Onset Pompe Disease Compared With Alglucosidase Alfa/Placebo
NCT Number:	NCT03729362
Document Date:	Final SAP v3.0: 14 January 2021

STATISTICAL ANALYSIS PLAN

VERSION: Final SAP v3.0

DATE OF PLAN:

14 January 2021

STUDY TITLE:

A PHASE 3 DOUBLE-BLIND RANDOMIZED STUDY TO ASSESS THE EFFICACY AND SAFETY OF INTRAVENOUS ATB200 CO-ADMINISTERED WITH ORAL AT2221 IN ADULT SUBJECTS WITH LATE-ONSET POMPE DISEASE COMPARED WITH ALGLUCOSIDASE ALFA/PLACEBO

PROTOCOL NUMBER: ATB200-03

Protocol Amendment 3, Dated: 14 August 2020

COMPOUNDS: ATB200 and AT2221 (ie, ATB200/AT2221)

SPONSOR:

Amicus Therapeutics, Inc.
1 Cedar Brook Drive
Cranbury, NJ 08512, USA
Phone: +1 609-662-2000

SIGNATURE PAGE

This document has been prepared and/or reviewed by:

Amicus

Amicus Therapeutics

This document has been reviewed and accepted by:

Amicus

Amicus Therapeutics

Amicus

Amicus Therapeutics

TABLE OF CONTENTS

LIST OF TABLES

1. LIST OF ABBREVIATIONS

Table 1: List of Abbreviations

Abbreviation	Term
6MWD	6-minute walk distance
6MWT	6-minute walk test
AE	adverse event
ANCOVA	analysis of covariance
ATC	Anatomical Therapeutic Chemical classification
CI	confidence interval
CI-MPR	cation-independent mannose 6-phosphate receptor
CK	creatine kinase
COVID-19	coronavirus disease 2019 that is caused by the SARS-Cov-2 virus
ECG	electrocardiogram
eCRF	electronic case report form
EQ-5D-5L	European Quality of Life-5 Dimensions 5 Response Levels
ERT	enzyme replacement therapy
ET	early termination
FVC	forced vital capacity
GAA	human acid α-glucosidase
GSGC	Gait, Stairs, Gowers' maneuver, Chair
Hex4	hexose tetrasaccharide
IAR	infusion-associated reaction
ITT	intent-to-treat population
ITT-LOCF	intent-to-treat population with missing data replaced by the last available value
ITT-OBS	intent-to-treat population that includes all available, observed data without any missing data imputation at Week 52
LOPD	late-onset Pompe disease
LS	least squares
MEP	maximal expiratory pressure
MIP	maximal inspiratory pressure
MMRM	mixed-effect model for repeated measures
mITT	modified intent-to-treat population

Table 1: List of Abbreviations (Continued)

Abbreviation	Term
MMT	manual muscle test
NAb	neutralizing antibody
PDLC	pre-defined limits of change
PGIC	Physician's Global Impression of Change
PK	pharmacokinetic
PP	per-protocol population
PP1	per-protocol population for sensitivity analysis of the 6MWD
PP2	per-protocol population for sensitivity analysis of the percent predicted FVC
PRO	patient-reported outcome
PROMIS	Patient-Reported Outcomes Measurement Information System
PT	preferred term
QMT	quantitative muscle test
rhGAA	recombinant human acid α-glucosidase
R-PAct	Rasch-built Pompe-specific Activity scale
SAP	statistical analysis plan
SGIC	Subject's Global Impression of Change
SD	standard deviation
SNIP	sniff nasal inspiratory pressure
SOC	system organ class
SVC	slow vital capacity
TEAE	treatment-emergent adverse event
TESAE	treatment-emergent serious adverse event
TUG	Timed Up and Go
VC	vital capacity

2. INTRODUCTION

This statistical analysis plan (SAP) describes the planned data derivations, analyses, and statistical methods to be performed for Study ATB200-03. It is based on:

- Protocol for ATB200-03 Amendment 3 dated 14 August 2020
- Amicus standard operating procedure (SOP) ATCRA-SOP-AT-008
- Good Clinical Practices (GCP) and International Conference on Harmonisation (ICH) guidelines E9 (Statistical Principles for Clinical Trials)
- Discussions with Regulatory Authorities

The purpose of this document is to provide details on study populations and on how the variables will be derived, how missing data and intercurrent events will be handled, as well as details on statistical methods to be used to analyze the efficacy and safety data from Study ATB200-03.

The document may evolve over time to reflect protocol amendments, regulatory discussions, and other important changes. However, the final SAP will be finalized, approved by the Sponsor, and placed on file before the database is locked. The data analyses and statistical methods specified in the final SAP supersede those described in the study protocol. Deviations from the final approved plan will be noted in the clinical study report.

Pharmacokinetic (PK) and immunogenicity analyses will be covered in separate analysis plans.

3. STUDY OBJECTIVES

The objectives of this study are to evaluate the efficacy, safety, patient-reported outcomes (PROs), and pharmacodynamics (PD) of ATB200/AT2221 co-administration in subjects with late-onset Pompe disease (LOPD) compared with alglucosidase alfa/placebo. The primary and the secondary objectives are stated below.

3.1. Primary Objective

- To assess the efficacy of ATB200/AT2221 co-administration on ambulatory function, as measured by the 6-minute walk test (6MWT), compared with alglucosidase alfa/placebo

3.2. Secondary Objectives

- To assess the efficacy of ATB200/AT2221 co-administration on pulmonary function, as measured by sitting forced vital capacity (FVC) (% predicted), compared with alglucosidase alfa/placebo
- To assess the efficacy of ATB200/AT2221 co-administration on muscle strength, compared with alglucosidase alfa/placebo
- To assess the efficacy of ATB200/AT2221 co-administration on health-related PROs, compared with alglucosidase alfa/placebo
- To assess the efficacy of ATB200/AT2221 co-administration on motor function, compared with alglucosidase alfa/placebo
- To assess the efficacy of ATB200/AT2221 co-administration on overall clinical impression as assessed by both the physician and subject, compared with alglucosidase alfa/placebo
- To assess the safety, tolerability, and immunogenicity of ATB200/AT2221 co-administration compared with alglucosidase alfa/placebo
- To assess the effect of ATB200/AT2221 co-administration on biomarkers of muscle injury and disease substrate compared with alglucosidase alfa/placebo
- To characterize the population PK of ATB200 and alglucosidase alfa in enzyme replacement therapy (ERT)-experienced subjects using plasma total acid α-glucosidase (GAA) protein level by signature peptide assay and plasma AT2221 concentration
- To characterize the PK of ATB200, alglucosidase alfa, and AT2221 in ERT-naïve subjects using noncompartmental analysis
- To explore the exposure-response relationship for ATB200/AT2221 and alglucosidase alfa/placebo co-administration

3.3. Study Endpoints

3.3.1. Primary Endpoint

The primary efficacy endpoint of this study is the change from baseline to Week 52 in 6-minute walk distance (6MWD) measured in meters, which is the distance walked in the 6MWT.

3.3.2. Key Secondary Efficacy Endpoints

The key secondary efficacy endpoints, in hierarchical order of importance, are:

- Change from baseline to Week 52 in sitting FVC (% predicted)
- Change from baseline to Week 52 in the manual muscle test (MMT) score for the lower extremities
- Change from baseline to Week 26 in 6MWD
- Change from baseline to Week 52 in the total score for the Patient-Reported Outcomes Measurement Information System (PROMIS®) – Physical Function
- Change from baseline to Week 52 in the total score for the PROMIS – Fatigue
- Change from baseline to Week 52 in the total score for the GSGC (Gait, Stairs, Gowers' maneuver, Chair)

Note: See Appendix 2 for further descriptions/definitions of the efficacy endpoints.

3.3.3. Other Secondary Efficacy Endpoints

The other secondary efficacy endpoints are:

- Change from baseline to Week 52 in % predicted 6MWD (where the predicted values are calculated using Enright and Sherrill 1998 reference equations)
- Proportion of subjects improving on both 6MWD and % predicted FVC (where the definition of improvement is provided in Section 8.7).
- Change from baseline to Week 52 in the following variables related to motor function:
  - time to complete the 10-meter walk (ie, assessment of gait) of the GSGC test
  - time to complete the 4-stair climb of the GSGC test
  - time to complete the Gower's maneuver of the GSGC test
  - time to arise from a chair as part of the GSGC test
  - time to complete the Timed Up and Go (TUG) test
- Change from baseline to Week 52 in the following variables related to muscle strength:
  - MMT score for the upper extremities
  - MMT total score

  - Proximal muscles MMT score
  - Quantitative muscle test (QMT) value (kg) for the upper extremities
  - QMT value (kg) for the lower extremities
  - QMT total value (kg)
- Change from baseline to Week 52 in the following variables from PRO measures:
  - Total score for the PROMIS – Dyspnea
  - Total score for the PROMIS – Upper Extremities
  - Total score for the Rasch-built Pompe-specific Activity (R-PAct) Scale
  - European Quality of Life-5 Dimensions 5 Response Levels (EQ-5D-5L) based on the EQ-VAS quantitative score
- Actual value of the subject's functional status (improving, stable, or declining) pertaining to the effects of study drug in the following areas of life at Week 52, as measured by the Subject's Global Impression of Change (SGIC):
  - overall physical wellbeing
  - effort of breathing
  - muscle strength
  - muscle function
  - ability to move around
  - activities of daily living
  - energy level
  - level of muscular pain
- Actual value of the subject's functional status (improving, stable, or declining) at Week 52, as measured by the Physician's Global Impression of Change (PGIC)
- Change from baseline to Week 52 in the following measures of pulmonary function, as follows:
  - Sitting slow vital capacity (SVC) (% predicted)
  - Maximum vital capacity (maximum VC) (% predicted), where maximum VC is defined as the larger of the sitting FVC and the sitting SVC
  - Maximal inspiratory pressure (MIP) (% predicted)
  - Maximal expiratory pressure (MEP) (% predicted)
  - sniff nasal inspiratory pressure (SNIP) (% predicted)

Note: See Appendix 2 for further descriptions of the efficacy endpoints.

3.3.4. Pharmacodynamic Endpoints

Pharmacodynamic endpoints are as follows:

- Change from baseline to Week 52 in serum creatine kinase (CK) level
- Change from baseline to Week 52 in urinary hexose tetrasaccharide (Hex4) level

3.3.5. Pharmacokinetic Endpoints

3.3.5.1. Population Pharmacokinetic Analysis

Population PK from sparse sampling in ERT-experienced subjects (Day 1 and Week 52) and ERT-naïve subjects will be performed. The population PK analysis of total GAA protein and AT2221 concentrations will be provided in a separate PK analysis plan.

3.3.5.2. Pharmacokinetic Sub-analysis (ERT-naïve Subjects)

The ERT-naïve subjects will undergo rich PK sampling to enable a noncompartmental analysis in the naïve subjects. After serial blood sampling in ERT-naïve subjects, PK parameters for plasma total GAA protein concentration will be calculated as follows:

- C_{max}: maximum observed concentration obtained directly from the concentration profile
- t_{max}: time of the first occurrence of C_{max} obtained directly from the concentration profile
- K_{el} or λ_z: the apparent terminal phase elimination rate constant will be estimated by linear regression of logarithmically transformed concentration versus time data; only those data points which are judged to describe the terminal log-linear decline will be used in the regression
- AUC_{0-t}: the area under the plasma drug concentration-time curve from 0 time (predose) to the time of last quantifiable concentration (t), calculated with the linear up/log-down trapezoidal method
- $AUC_{0-\infty}$: the area under the plasma drug concentration-time curve from 0 time (predose) extrapolated to infinite time will be calculated as follows:

  $$AUC_{0-\infty} = AUC_{0-t} + C_t/\lambda_z$$

  where C_t is the last observed quantifiable concentration
- $AUC_{tmax-24h}$: partial area under the plasma drug concentration-time curve from t_{max} to 24 hours post initiation of ATB200 infusion
- $t_{1/2\alpha}$: the alpha-phase terminal elimination half-life, calculated as $\ln(2)/\lambda_z$ from the first 3 or more quantifiable concentrations after C_{max}; estimated for plasma total GAA protein concentration only
- CL_T: total plasma clearance after IV administration, calculated as dose/$AUC_{0-\infty}$

Plasma AT2221 PK parameters will be calculated as follows:

- C_{max}
- t_{max}
- K_{el} or λ_z
- AUC_{0-t}
- $AUC_{0-\infty}$
- $t_{½\beta}$: the beta-phase terminal elimination half-life, calculated as $\ln(2)/\lambda_z$ from the last 3 or more quantifiable concentrations from the beta phase of elimination
- CL_T/F: total plasma clearance from an oral administration, calculated as dose/$AUC_{0-\infty}$
- V_z/F: volume of distribution based on the terminal elimination phase from an oral administration, calculated as dose/($AUC_{0-\infty} \bullet \lambda_z$)

The details of the PK analyses will be described in a separate PK analysis plan.

3.3.5.3. Immunogenicity Endpoints

Immunogenicity endpoints are:

- Total anti-drug antibodies (ADAs), including titers
- Neutralizing antibodies (NAbs):
  - Inhibition of recombinant human acid α-glucosidase (rhGAA) binding to cation-independent mannose-6-phosphate receptor (CI-MPR)
  - Inhibition of rhGAA-mediated hydrolysis of 4-MU-glucoside
  - Inhibition of rhGAA-mediated hydrolysis of glycogen
- Anti-rhGAA cross-reactive with alglucosidase alfa
- Anti-rhGAA-specific Immunoglobulin E (IgE)

The details of the immunogenicity analyses will be described in a separate document: "Modeling and Simulation Plan (MSP): Population Pharmacokinetics and Immunogenicity Analyses for ATB200/AT2221 in Patients with Pompe Disease Enrolled in Study ATB200-03."

3.3.6. Safety Endpoints

The safety profile of ATB200/AT2221 will be characterized using incidence of treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), and adverse events (AEs) leading to discontinuation of study drug, frequency and severity of immediate and late infusion-associated reactions (IARs), and any abnormalities noted in other safety assessments. The impact of immunogenicity to ATB200 and alglucosidase alfa on safety and efficacy will also be described. The immunogenicity analyses will also be performed by ERT status subgroups (ERT-naïve and ERT-experienced).

4. STUDY DESIGN

4.1. Summary of Study Design

This is a Phase 3 double-blind, randomized, parallel-group, multicenter, international study of ATB200/AT2221 in adult subjects with LOPD compared with alglucosidase alfa/placebo. Subjects enrolled will be either ERT-experienced or ERT-naïve.

The study consists of a 12-month double-blind treatment period during which eligible subjects are randomly assigned in a 2:1 ratio to receive either ATB200/AT2221 co-administration or alglucosidase alfa/placebo co-administration.

Subjects who complete 12 months of treatment in this study will have the option to participate in an open-label extension (OLE) study (ATB200-07) to receive ATB200/AT2221 under a separate protocol.

4.2. Definition of Study Drugs

Study drugs defined in this protocol are ATB200/AT2221 co-administration and alglucosidase alfa/placebo combination. The study drugs will be administered every 2 weeks (QOW).

The dose for ATB200 as well as alglucosidase alfa will be 20 mg/kg, and both are administered via infusion. The dose for AT2221 will be 260 mg (4 x 65 mg oral capsules) for subjects who weigh ≥ 50 kg, and 195 mg (3 x 65 mg oral capsules) for subjects who weigh ≥ 40 to < 50 kg. The AT2221 and placebo are to be taken orally 1 hour prior to the start of the infusion.

4.3. Sample Size and Power Considerations

The primary endpoint for this study is the change from baseline to Week 52 in the 6MWD.

Using a 2-group t-test with a 1-sided significance level of 0.025 and a 2:1 randomization ratio, a total of 99 subjects (66 subjects in the ATB200/AT2221 group and 33 subjects in the alglucosidase alfa/placebo group) would yield approximately 90% power to detect a clinically meaningful standardized effect size of 0.7 between the 2 groups in a superiority test for the primary endpoint. This calculation was performed using nQuery 8©.

Assuming a 10% dropout rate (after randomization), a total of approximately 110 subjects were planned to be randomized to ensure 99 evaluable patients.

Details of the sample size justification are described in Appendix 1.

4.4. Randomization and Stratification

Although this is a multicenter trial, the randomization was not stratified by center. The following 2 risk factors were identified as key prognostic factors, and therefore randomization was stratified according to these:

1. Baseline 6MWD (categorized as: 75 to < 150 meters, 150 to < 400 meters, ≥ 400 meters)
2. ERT status (categorized as: ERT-experienced versus ERT-naïve)

The randomization schedule was generated and administered centrally by Almac Clinical Technologies (Almac), independent of the sponsor's project team. This centralized block

randomization procedure was used to balance the above risk factors, (1) to reduce bias and increase the precision of statistical inference, and (2) to allow various planned and unplanned subset analyses. The randomization ratio was 2:1 (ATB200/AT2221 to alglucosidase alfa/placebo), fixed. Approximately 30 ERT-naïve subjects were planned to be enrolled.

4.5. Clinical Assessments

Study assessments were to be performed in accordance with the Schedule of Assessments of the ATB200-03 protocol.

4.5.1. Impact of COVID-19 Pandemic

The emergence of the coronavirus disease 2019 (COVID-19) during the conduct of study and the related travel restrictions and quarantines resulted in some subjects experiencing missed infusions, missed assessments, delayed visits, and early withdrawals. These could potentially alter the subject's response to treatment, confound the assessment of the 6MWD and FVC, and impact the analyses.

Assessment and infusion visits that were missed due to COVID-19 related reasons were to be recorded as protocol deviations in the electronic data capture (EDC) and the reasons attributed to COVID-19. Infusions that were missed near the scheduled assessments visits at Weeks 26, 38, and 52 were to be examined on a case-by-case basis to determine whether catch-up infusions were needed, and if so, how many catch-up infusions were needed, prior to rescheduling the visits to perform make-up assessments. As a result, subjects who missed visits due to COVID-19 related reasons could be enrolled longer than 12 months in the study due to delayed visits.

As a general rule, a subject missing at least 2 consecutive infusions just prior to the Week 52 assessments (or between Week 38 and Week 52) due to COVID-19 was required to receive make-up infusions prior to the EOS assessments. The make-up infusions required 4 consecutive infusions to be administered regardless of the number of infusions missed. For example: if a subject missed 3 consecutive infusions at Week 42, 44 and 46, then the subject was required to receive make-up infusions at Week 48, 50, 52 and 54, and the EOS assessments were to be conducted thereafter, prior to Week 56. A subject missing 5 consecutive infusions was to be withdrawn from the study.

Analyses will be performed to assess whether the number of subjects with missing data due to COVID-related reasons are comparable between the 2 groups (at Week 52 and across the study).

5. INTERIM ANALYSES

There is no interim analysis planned for this study.

6. GENERAL CONSIDERATIONS FOR DATA ANALYSES AND HANDLING

6.1. Analysis Populations

The total number of subjects screened, number who screen-failed, and the number randomized (including subjects randomized after rescreening) will be tabulated.

6.1.1. Screen Failures

Subjects who signed an informed consent form (ICF) but failed to qualify for the study and were therefore not randomized, will be considered screen failures.

Information on screen failures will be presented in a data listing.

6.1.2. Safety Population

The safety population is defined as all subjects who received at least 1 dose of study drug (ATB200/AT2221 or alglucosidase alfa/placebo). This population will be used in the assessment and reporting of safety data. Subjects will be analyzed according to the actual treatment received.

6.1.3. All Randomized Population

The 'All Randomized' population consists of all subjects who were randomized, regardless of whether they took a dose of study drug. This population will be used only for subject accounting purposes, and it will be analyzed according to the planned treatment groups.

6.1.4. ITT Population

The intent-to-treat (ITT) population consists of all randomized subjects who received at least 1 dose of study drug. This population will be analyzed according to the planned treatment groups. The ITT population will be used for additional baseline and demographic summaries.

For the purpose of efficacy analysis, the ITT population has been characterized in 2 different ways (ITT-observed [OBS] and ITT-last observation carried forward [LOCF]) (see Section 8.2 and Section 6.3.2 for further descriptions).

6.1.4.1. ITT-OBS

The ITT-OBS is the ITT population using all available, observed data without imputation for missing post-baseline data. That is, missing data at Week 52 and at other visits are not replaced.

The ITT-OBS population is identical to the ITT population as defined above. The mixed-effect model for repeated measures (MMRM) analysis will be performed on the ITT-OBS population.

6.1.4.2. ITT-LOCF

The ITT-LOCF is the ITT population with missing data replaced with the last available value from post-baseline results. That is, the LOCF replaces missing data at Weeks 26, 38, and 52 with the last available endpoint value. Where applicable, the observed baseline result will be used to replace a missing post-baseline result at the Week 12 visit. Imputed baseline value cannot be used to replace missing Week 12 result. The missing value at Week 52 is replaced with the last

available value from the subject in the study. This can be the value from the early termination (ET)/end of study (EOS) visit if available. If not available, the last available value from prior post-baseline visits (Week 38, Week 26, or Week 12, whichever is available) will be used to replace the missing value at Week 52.

Example for visits earlier than Week 52: Suppose the subject only missed the assessment at Week 12 but had assessments at some other visits after Week 12 (ie, Week 26 and/or Week 38 and/or Week 52). Then the missing assessment at Week 12 is replaced with the last available value on/prior to Week 12 (eg, the observed baseline value). If the subject missed the Week 12 visit due to discontinuation/withdrawal from the study and had follow-up assessments at the ET/EOS visit, then the last available value from the ET/EOS visit will be used to replace the missing Week 12 value. This EOS value is then carried forward to replace missing values at Weeks 26, 38, and 52 as the subject had withdrawn from the study since Week 12.

Note: The observed baseline result can only be used to replace missing data at Week 12, where applicable (ie, if the subject has other post-baseline data). It will not be carried forward to replace missing values at Week 26, 36 or 52. In addition, imputed baseline values cannot be used to replace missing post-baseline values.

6.1.5. Modified ITT Population

The modified intent-to-treat (mITT) population is a subset of the ITT population consisting of all randomized subjects who took at least 1 dose of study drug and have both baseline and at least 1 post-baseline assessment for the 6MWD. This population will be analyzed according to the planned treatment groups. The mITT population will be used for subject accounting purposes.

6.1.6. Per-protocol Population

The per-protocol population (PP) is a subset of the mITT population consisting of any mITT subjects who do not have pre-specified protocol deviations that are considered important, as they potentially may have altered the subject's response to study treatment or confounded the study assessments, thereby impacting the analyses of the primary endpoint (6MWD) and the FVC. This population will be analyzed according to the actual treatment received.

The PP population will be characterized separately for sensitivity analyses for the 6MWD and % predicted FVC. The per-protocol population 1 (PP1) will be used for the sensitivity analyses of the 6WMD, while per-protocol population 2 (PP2) will be used for sensitivity analyses of the % predicted FVC.

Pre-specified important protocol deviations leading to exclusion from the PP1 and PP2 populations include those in Table 2. These populations will be determined and finalized before locking the database.

Table 2: Pre-specified Important Protocol Deviations Leading to Exclusion from Per-protocol Populations (PP1 and PP2)

Item	Important Protocol Deviation Description (Excluded from PP1: for 6MWD)	Important Protocol Deviation Description (Excluded from PP2: for FVC)
1.	Subject did not have sufficient treatment exposure (ie, subject did not have at least 80% compliance with study treatment)	Subject did not have sufficient treatment exposure (ie, subject did not have at least 80% compliance with study treatment)
2.	Subject did not have the primary endpoint (6MWD) assessment at the Week 52 visit (including catch-up assessment or delayed visit) due to COVID-19 pandemic or other reasons	Subject did not have FVC assessment at the Week 52 visit (including catch-up assessment or delayed visit) due to COVID-19 pandemic or other reasons
3.	Subject was discontinued from the study for missing at least 5 infusions	Subject was discontinued from the study for missing at least 5 infusions
4.	Subject withdrew consent and/or discontinued from the study	Subject withdrew consent and/or discontinued from the study
5.	Subject was unblinded (accidentally or otherwise) before database lock	Subject was unblinded (accidentally or otherwise) before database lock
6.	Subject was inconsistent in the use/non-use of assisted devices during 6MWT	

Abbreviations: 6MWD = 6-minute walk distance; 6MWT = 6-minute walk test; FVC = forced vital capacity; PP1 = per-protocol population 1 is used for the sensitivity analyses of the 6WMD; PP2 = per-protocol population 2 (PP2) is used for the sensitivity analyses of the percent predicted FVC

6.1.7. Completers

The Completers population consists of subjects who completed the Week 52 visit and have both 6MWD and FVC assessments at the Week 52 visit (including make-up assessments or delayed visits). Subjects who completed the Week 52 visit and were able to do only 1 6MWT (instead of 2) at Week 52 will still be considered as completers. This population will be used for subject accounting purposes.

A subject who withdraws/discontinues from the study and returns for the end of study or early termination visit is not a completer and does not belong in the Completers population.

6.1.8. PK Population

The PK population consists of those randomized subjects who have been exposed to at least 1 dose of the study drug (ATB200/AT2221 or alglucosidase alfa/placebo) and have at least 1 PK assessment. For ERT-naïve subjects who undergo serial PK sampling, noncompartmental PK analysis is described in Section 3.3.5.2. For all other subjects who undergo sparse PK sampling, population PK analysis will be described in a separate Modeling and Simulation Plan.

6.2. Derived and Transformed Data

6.2.1. Study Day

The study has no Study Day 0. The date of first dose of study drug (First Dose Date) is labeled as "Day 1" for analysis purposes; it will serve as the reference start date.

Total duration of treatment will be calculated as the difference between the dates of last and first dose of study drug plus 1.

If the date of interest occurs on or after the first dose/randomization date, then study day will be calculated as (date of interest – date of first dose) + 1. If the date of interest occurs prior to the first dose/randomization date, then study day will be calculated as (date of interest – date of first dose).

6.2.2. Baseline Definition

In general, baseline is defined as the last non-missing measurement prior to the administration of the first dose of study drug, unless otherwise specified. It is understood that baseline may be different for different efficacy and safety endpoints.

- For the 6MWD, FVC, and SVC, the average of the last 2 values obtained on or prior to the first dose date will be used as the baseline value.
- For other functional assessments (such as the MMT, QMT, TUG, GSGC, PROMIS instruments, EQ-5D-5L (quantitative EQ-VAS score and descriptive health status), and R-PAct scale), the screening results will serve as the baseline values.
- For all other pulmonary function tests (MIP, MEP, and SNIP), baseline will be the last available value obtained on or prior to the first dose date.
- For safety data (including all lab assessments, physical examinations, ECG, vital signs, weight, etc), the last available result on or prior to the first dose date will be used as the baseline value (without considering the time of assessment on Day 1). If this value is missing, then the screening value will be used instead.

6.2.3. Baseline Age

Calculation of subject's age in years is based on the date of informed consent, using the following formula: Age (year) = FLOOR ([date of informed consent – date of birth] / 365.25), where the FLOOR ( ) function returns the integer part of the result.

6.2.4. Week 52/ET Assessments

Subjects who discontinue study drug or who are withdrawn from the study are required to return for an EOS or ET visit. The muscle strength tests, pulmonary function tests, motor function tests, PRO assessments, and pharmacodynamic endpoints (Hex4 and CK) performed at the ET visit will be used as the final values at the Week 52 visit for statistical analysis. The 6MWT is to be performed twice at the Week 52/ET visit, and the average of the 2 test values will be used as the final value at Week 52 for statistical analysis.

Note: As a result of the COVID-19 related policies and restrictions, the Week 52 visit (and/or the ET/EOS visit) may be delayed. Even if that happens, the delayed visit assessments will still be used for analyses. The make-up assessments at Week 52 will be used as the Week 52 results. In addition, some subjects were only able to have 1 assessment of 6MWD at Week 52. For these subjects, the single assessment will still be used as the Week 52 value.

6.2.5. Change and Percent Change from Baseline

Change from baseline is calculated as:

Value at a visit – Baseline value.

Percent change from baseline is calculated as:

100 * (Value at a visit – Baseline value) / Baseline value.

6.2.6. Analysis Visit Windows

6.2.6.1. Visit Windows

Analysis visit windows will be used for all by-visit analyses, as provided for in Table 3 and Table 4.

Table 3: Visit Windows for Functional Assessments, PROs, ECG, and CPE (as applicable)

Visit	Week	Target Day	Analysis Visit Window
Visit 1	Week -2[a]	Day -15/-14	Screening/Baseline for efficacy
Visit 1.10[b]	< Week 0	Day -10	Randomization
Visit 6	Week 12	Day 85	Day 2 – 134
Visit 7	Week 26	Day 183	Day 135 – 225
Visit 8	Week 38	Day 267	Day 226 – 316
Visit 9 (EOS/ET)	Week 52	Day 365	> Day 316
Visit 10 (F/U)[c]	Week X.10	N/A	N/A

Abbreviations: AE = adverse event; CPE = complete physical exam; ECG = electrocardiogram; EOS = end of study; ET = early termination; F/U = follow-up; IRT = Interactive Response Technology; N/A = not applicable; PRO = Patient-reported outcome

[a] Denotes Day -15/-14 in the screening period.

[b] Denotes the randomization day. Randomization is carried out via the IRT system and subjects do not need to be present. Hence, 'randomization day' is not a scheduled visit in the study protocol. Using the 'Visit 1.10' designation can help with data programming for analysis purposes.

[c] Denotes follow-up visits occurring at Week X.10 (AEs and concomitant medications are to be collected up to 30 days after last dose date).

Table 4: Visit Windows for Vital Signs, Lab Measurements, and BPE (as applicable)

Visit	Week	Target Day	Analysis Visit Window
Visit 1	Week -2[a]	Day -15/-14	Screening/Baseline for efficacy
Visit 1.10[b]	< Week 0	Day -10	Randomization
Visit 2	Week 0[c]	Day 1	Baseline for safety
Visit 3	Week 2	Day 15	Day 2 – 22
Visit 4	Week 4	Day 29	Day 23 – 36
Visit 5	Week 6	Day 43	Day 37 – 64
Visit 6	Week 12	Day 85	Day 65 – 134
Visit 7	Week 26	Day 183	Day 135 – 225
Visit 8	Week 38	Day 267	Day 226 – 316
Visit 9 (EOS/ET)	Week 52	Day 365	> Day 316
Visit 10 (F/U)[d]	Week X.10	N/A	N/A

Abbreviations: AE = adverse event; BPE = brief physical exam; EOS = end of study; ET = early termination; F/U = follow-up; IRT = Interactive Response Technology; N/A = not applicable; PRO = patient-reported outcome

[a] Denotes Day -15/-14 in the screening period.

[b] Denotes the randomization day. Randomization is carried out via the IRT system and subjects do not need to be present. Hence, 'randomization day' is not a scheduled visit in the study protocol. Using the 'Visit 1.10' designation can help with data programming for analysis purposes.

[c] Denotes study visit Day 1, at which baseline safety assessments are performed.

[d] Denotes follow-up visits occurring at Week X.10 (AEs and concomitant medications are to be collected up to 30 days after last dose date).

6.2.6.2. Impact of COVID-19 on the Visit Windows

Due to the COVID-19 related policies and restrictions, sites were allowed to perform make-up assessments at a later date when feasible. This resulted in some delayed visits, and some of this may appear in the Study Data Tabulation Model (SDTM) datasets as unscheduled visits or multiple assessments at some specified visits.

Delayed visits and make-up assessments that go beyond the stated visit windows will be remapped to the planned study visit. For example, the visit window for Week 26 is Day 135 to Day 225, and the window for Week 38 is Day 226 to Day 316. Suppose a subject had a COVID-19 related situation leading to a delay in the Week 26 assessment such that the delayed visit / make-up assessment for Week 26 occurs on Day 261 (which is 36 days outside of the planned visit window for Week 26, and it's inside the Week 38 visit window), followed by the Week 38 assessment on Day 300 (which is right within the Week 38 visit window). Here, the delayed assessment occurring on Day 261 will be remapped to the Week 26 visit (which was the intended visit).

6.2.7. Multiple Assessments in a Visit Window

6.2.7.1. For Efficacy Data

In general, for all post-baseline efficacy and other functional by-visit analyses, if multiple assessments (from both scheduled and unscheduled visits) occur in the same visit window, the visit dates should first be examined to see if they are part of the COVID-19 related delayed visits or make-up assessments that require visit windows to be adjusted or remapped (as discussed in Section 6.2.6.2) in order for the data to be in the appropriate study visits.

- If they are part of COVID-19 related delayed visits or make-up assessments, then the assessments should be handled in accordance with the handling of delayed visits and make-up assessments in Section 6.2.4 and the visit window remapping described in Section 6.2.6.2.
- If they are not part of COVID-19 related delayed visits or make-up assessments that require visit window adjustment per Section 6.2.4 and Section 6.2.6.2, then the usual visit window rules apply.

That is, if they are not COVID-19 related (or even if they are COVID-19 related but do not require any visit window adjustment or remapping), then for multiple assessments in the same visit window, the value closest to the planned target visit day will be used for analysis at that visit. If two assessments are equidistant from the target day, or if two assessments occur on the exact same day but at different times, the last value in chronological order will be used.

6.2.7.2. For Safety Data

For all post-baseline safety by-visit analyses, if multiple assessments (from both scheduled and unscheduled visits) occur in the same visit window, the value closest to the planned target visit day will be used for that visit (unless these are COVID-19 related delayed visits for which visit window adjustment or remapping is required). If two assessments are equidistant from the target day or occur on the same day but at different times, the last assessment in chronological order will be used (unless these are COVID-19 related delayed visits for which visit window adjustment or remapping is required).

Note: See Section 6.2.6.2 regarding the handling of the impact of COVID-19 related delayed visits and assessments on the analysis visit windows.

6.3. Handling of Missing Data for Key Endpoints

6.3.1. Imputation for Missing Baseline Values

All available subjects in the ITT population have baseline values for both 6MWD and FVC. Hence, no baseline imputation is needed for these 2 key efficacy endpoints.

For the multi-component/multi-item endpoints (such as GSGC and MMT), a subject could be missing either specific items or the entire assessment. If the baseline value is partially missing (subject is missing only specific items at baseline), the average value using all subjects with non-missing values for that item across the 2 treatment groups combined will be used to replace the missing item score. If the baseline total score is completely missing, the average score using all subjects with non-missing total scores for the endpoint across the 2 treatment groups

combined will be used to replace the missing baseline total score. See Section 8.8.4 for the handling of missing baseline values for PROs including PROMIS instruments.

6.3.2. Imputation for Missing Post-baseline Values

For this study, subjects who permanently discontinued the assigned study drug are subsequently discontinued from the study. However, these subjects are required to return for the EOS/ET visit. Efficacy assessments collected at the EOS/ET visit will be used to replace the missing endpoint value at Week 52 in the ITT-LOCF population. That is, the last available observation will be carried forward to replace the missing value at Week 52 (see Section 6.1.4.2).

The MMRM analysis is the primary analysis for the primary endpoint (6MWD), and it will be performed as an ITT-OBS (observed cases) analysis. That is, all available, observed values will be used without imputation, and missing data at Week 52 are not replaced.

6.4. Handling of Intercurrent Events

Intercurrent events are defined as events that occur after treatment initiation and either preclude observation of the primary endpoint or affect its interpretation (eg, discontinuation of treatment, terminal events such as death, etc), per ICH E9 R1.

The potential intercurrent events for this study are listed in Table 5 together with the proposed strategy and handling of such events for the mixed-effect model repeated measures (MMRM) analysis and the analysis of covariance (ANCOVA).

Table 5: Accounting for Intercurrent Events

Intercurrent Event	Treatment Strategy for Handling Event	Handling of Event
Intervention for AEs, SAEs and IARs[a]; Incorrect treatment assigned; Prespecified important major protocol deviations that can potentially affect the assessment of 6MWD.	Treatment policy strategy (based on ITT principle): • If subject continues in the study and dosing is resumed	For the ITT-OBS population: Value of interest (eg, 6MWD result observed at a given visit) will be used, regardless of whether the intercurrent event occurred. Note: No imputation will be performed for missing values at Week 52 (and other post-baseline visits). This is for the MMRM analysis. For the ITT-LOCF population: Value of interest will be used, regardless of whether the intercurrent event occurred. Note: Missing values at Week 52 (and other post-baseline visits) will be imputed with the last available post-baseline assessment value. This is for the ANCOVA analysis. Example: Compare "Treat A + Use of prohibited concomitant medication as needed" versus "Treat B + Use of prohibited concomitant medication as needed."
	• If subject permanently discontinues from study drug (and is withdrawn from the study)	For the ITT-OBS population: Observed value at time of discontinuation will be used. Note: No imputation will be performed for missing values at Week 52 and other post-baseline visits. Data after study discontinuation will be considered missing and will be handled by the MAR assumption in the MMRM analysis. For the ITT-LOCF population: The last available assessment value will be used (regardless of whether the intercurrent event occurred). This setup is for the ANCOVA analysis.

Table 5: Accounting for Intercurrent Events (Continued)

Intercurrent Event	Treatment Strategy for Handling Event	Handling of Event
Withdrawn at subject's own request (eg, subject withdrew consent); Withdrawn based on investigator's opinion; Withdrawn for persistent noncompliance[b]; Withdrawn due to SAEs or COVID-19 related reasons. Withdrawn due to IARs	Treatment policy strategy (based on ITT principle) Here, subject is permanently discontinued from study drug (and is withdrawn from the study).	For the ITT-OBS population: The observed value available at time of discontinuation will be used. No imputation will be performed for missing values at Week 52 and other post-baseline visits. Data after study discontinuation will be considered missing and will be handled by the MAR assumption in the MMRM analysis; For the ITT-LOCF population: The last available assessment value will be used, regardless of whether the intercurrent event occurred. Note: This setup is for the ANCOVA analysis.

Abbreviations: 6MWT = 6-minute walk test; AE = adverse event; IAR = infusion-associated reaction; ITT = intent-to-treat; MAR = missing at random
Note: A subject who permanently discontinues study drug is subsequently withdrawn from this study.
[a] Intervention for AEs and IARs includes concomitant medications, dose modifications, interruptions, or withdrawal.
[b] Withdrawn for persistent noncompliance with study drug, study requirements, or failure to return to the study site for infusion visits. Eg, subject missed 5 or more infusions; or subject's calculated compliance value < 80%.

6.4.1. Missing AE Attributes

Adverse events with missing attributes will be imputed with the worst possible outcome, for summary purposes. For example, if an AE is missing relationship to study drug, it will be imputed as being 'related' to the study drug; if an AE is missing the severity level, it will be imputed as being 'severe.'

Nonetheless, in the data listings, the actual observed values will be reported. That is, if the response is missing, it will show as missing in the data listings.

6.4.2. Missing Start and Stop Dates for Prior and Concomitant Medication

Missing start and stop dates for prior and concomitant medications will be imputed in accordance with the rules described in Section 9.4.1.

6.4.3. Missing Start and Stop Dates for Adverse Events

Missing start and stop dates for AEs will be imputed in accordance with the rules described in Section 9.1.6.

7. CHARACTERISTICS OF STUDY POPULATION

7.1. Subject Disposition

For each analysis population (Screened, All Randomized, ITT, ITT-OBS, ITT-LOCF, mITT, Safety, PP, Completers, and PK), the study enrollment and completion status will be summarized by the number and percent of subjects who completed the study, who discontinued the study early for each treatment group, and for all subjects combined. The number of subjects who were randomized following rescreening will be included in the disposition table.

The number and percent of subjects, by the reason for discontinuing the study early, will also be presented for each treatment group and for all subjects combined. The categories classifying the subjects' study completion status are derived from the end of study electronic case report form (eCRF) page and include categories such as "Lost to Follow-Up," "Adverse Event," and "Death." The reasons for early study termination will likely provide some insight into how the trial was conducted and some understanding of the potential dropout pattern.

Study Completion/Termination Status:

The completion status of subjects at the end of the study will be categorized and summarized by treatment group with respect to the number of subjects who:

- Completed the study (based on the 'Yes' checked box on the eCRF)
- Discontinued the study prematurely

Note that the 'Yes' checked box on the eCRF does not necessarily indicate that the subject belongs to the Completers population. The Completers population is defined in Section 6.1.7. The reasons for early study termination will be summarized as recorded on the eCRF (eg, AE, protocol violation, administrative, lost to follow-up, death, COVID-19 related, etc).

7.2. Screen Failures

The number of subjects who failed screening will be presented in an individual subject data listing, which will include ERT status (ERT-naïve or ERT-experienced) and other demographic data.

7.3. Protocol Deviations

Protocol deviations data are collected and consolidated into a deviation log, and transferred to Amicus for review on an ongoing basis throughout the course of the study to identify trends and determine the type (eg, inclusion/exclusion criteria, study procedures) and the category (eg, minor or major) of each deviation. The number and percent of subjects with protocol deviations will be presented for each treatment group and summarized by deviation type (major or important). A listing of subjects with all protocol deviations (minor, major, and important) will be provided.

7.4. Demographic and Baseline Characteristics

Demographic and baseline characteristics will be summarized with descriptive statistics defined in Section 8.1 by treatment group for the ITT, mITT, and safety populations. Demographic and baseline characteristics to be presented include:

- Gender; race; ERT status (naïve or experienced); subjects using assistive devices at baseline; subjects with a history of falls (Yes, No); and subjects with a history of IARs (Yes, No) will be summarized by treatment group and overall using number and percentage as described in Section 8.1.
- The continuous variables height (cm), weight (kg), CK, Hex4, ALT, AST, and body mass index (BMI) (kg/m^2) will be summarized using descriptive statistics: number of observations (n), mean, standard deviation (SD), median, first quartile (Q1), third quartile (Q3), minimum, and maximum as described in Section 8.1.

Other baseline variables to be summarized will include:

- Age at informed consent date (in years), as a continuous variable;
  - Categorized as: ≥ 18 to < 35 years, ≥ 35 to < 50 years, ≥ 50 to < 65 years, ≥ 65 years
- Age at diagnosis (in years) as a continuous variable;
  - Categorized as: < 18 years, ≥ 18 to < 35 years, ≥ 35 to < 50 years, ≥ 50 years
- Age at first ERT dose in years (for subjects who are ERT-experienced only) as a continuous variable;
  - Categorized as: < 18 years, ≥ 18 to < 35 years, ≥ 35 to < 50 years, ≥ 50 years
- ERT duration (for ERT-experienced subjects) as a continuous variable, calculated as: ROUND([date of randomization – date of first dose of ERT] / 365.25);
  - Categorized as: 2 to < 3 years, 3 to < 5 years, ≥ 5 years.

For partial dates involving age at diagnosis and age at first dose of ERT, missing Day will be imputed as '01,' and missing Month will be imputed as 'JANUARY.'

Other baseline characteristics include the baseline 6MWD (that is, the average of the 2 screening values), and the corresponding % predicted 6MWD, baseline FVC and other pulmonary function parameters, baseline motor function tests, muscle strength tests, and PROs.

The total number of subjects enrolled by country and by region will be presented by treatment group and overall. The 3 specified regions are: North/South America, Europe, and Asia Pacific. The demographic and baseline characteristics summary will also be presented by ERT status (naïve vs experienced).

The Pompe disease history information will be summarized for subjects who are ERT-naïve, ERT-experienced, have baseline 6MWD categorized as: 75 to < 150 m, 150 to < 400 m, and ≥ 400 m, based on the descriptive statistics defined in Section 8.1, for the ITT, and safety populations.

A listing of baseline data will be provided.

7.5. Listing of Subject Inclusion and Exclusion Criteria

For screen failures, a listing for all subjects will be provided showing which specific eligibility criteria were not met.

7.6. Medical History and Medical Conditions Present at Entry

The medical history (captured on the eCRF) will be coded using version 20.1 or later of the Medical Dictionary for Regulatory Activities (MedDRA) into system organ class and preferred term. The frequency of subjects with medical history by system organ class and preferred term will be presented on the safety population.

7.7. Handling the Use of Assistive Device During the 6MWT

Subjects using a permitted assistive device during the Screening/Baseline 6MWT should use the same assistive device for all follow-up 6MWT assessments in the study. If the subject added a new device, changed the device at certain visits, or did not use the assistive device, then that will be considered a protocol deviation as the change can affect the primary endpoint and the results of these assessments. See Section 6.1.6 and Section 6.4 for the listing of pre-defined major protocol deviations and the listing of intercurrent events.

7.8. Prior and Concomitant Medications and Nondrug Therapies

Prior medications and nondrug therapies are defined as any medication or nondrug therapy taken or occurring before the first dose of study drug. Concomitant medications and nondrug therapies include medications and nondrug therapies with start dates (and times) on or after the first dose of study drug, medications and nondrug therapies with onset dates (and times) prior to the first dose of study drug without a stop date, or medications and nondrug therapies with a stop date (time) after the first dose of study drug. Medications starting after the last dose of study drug will not be considered as concomitant but will be flagged as post medications in the listing of medications. Nondrug therapies include physiotherapy and occupational therapy, procedures, surgery, or use of assistive devices.

Prior and concomitant medications are coded to indication-specific Anatomical Therapeutic Chemical classification (ATC) and preferred name using the World Health Organization Drug Dictionary (WHO DDE, September 2013 or later).

Prior and concomitant medication use will be summarized by level 4 ATC and preferred name using the number and percentage of subjects by treatment group. Medications will be sorted alphabetically by ATC and preferred name within ATC, and in decreasing order of the total number of subjects who took each medication within each ATC class. Subjects with multiple occurrences of a medication in ATC and preferred name will only be counted once within each ATC and preferred name. In addition, the total number of subjects to ever take any concomitant medications will be presented. Since medications are coded to ATC by indication, preferred names may appear under multiple ATCs.

A listing of all medications, both prior to infusions and concomitant, and nondrug therapies will be presented. Investigator verbatim descriptions as well as coded terms will be included in the listings. Any abbreviations and codes will be clearly explained on each page of the listing. The listing will be sorted by treatment group and subject identifier, and will include level 4 ATC,

preferred name, reported name, dose, route of administration, dosing frequency, start date, end date, indication, and period of medication (prior only, concomitant only, prior and concomitant), and nondrug therapy. Pre-medications (which are a subset of all medications) will be listed in a separate data listing, and this listing will include a flag indicating whether the specified medication is a new pre-medication and whether there were changes in the subject's pre-medication (eg, changed to a different pre-medication change in dose of the pre-medication) prior to infusions.

Prior and concomitant medications and nondrug therapies will be summarized by treatment group.

7.8.1. Non-diagnosis Related Prior Medication History

Non-diagnosis related medications are medications taken for conditions other than the one under study. These medications will be identified, and a supportive listing will be provided to include the unique data associated with the medication start and stop dates, whether the medication is ongoing, dose, frequency of administration, route of administration, trade name, generic name, and why it was prescribed.

7.8.2. Prior Diagnosis Related Medication History

Prior diagnosis related medication history, if available, will be flagged in the listing discussed above in Section 7.8.1.

7.9. Baseline Physical Examination

Overall assessment of baseline physical exam (categorized as normal, abnormal, or not performed) will be summarized by number and percent, together with post-baseline data.

7.10. Baseline Vital Signs

Baseline vital signs including systolic and diastolic blood pressures (mmHg), heart rate (bpm), and body temperature (°C) will be summarized using descriptive statistics, together with post-baseline data.

7.11. Baseline Laboratory Data

Baseline laboratory results will be summarized with descriptive statistics, together with the post-baseline lab data. See Section 9.2 for analysis of clinical laboratory data.

7.12. Baseline Primary and Secondary Efficacy Evaluations

The results from the primary and secondary efficacy evaluations obtained at baseline will be summarized with appropriate descriptive statistics, together with post-baseline data.

8. ANALYSIS OF EFFICACY

8.1. General Considerations

The Statistical Analysis System (SAS©) software version 9.4 (or the latest version at the time of the analysis) and R software will be used for all statistical procedures and analyses.

In general, where basic summary statistics are needed, continuous variables will be summarized using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum); categorical variables will be summarized using number and percentage. For basic summaries involving the change from baseline, a 95% confidence interval (CI) for the mean difference will also be provided.

All inferential statistical tests for the primary and key secondary efficacy endpoints will be 1-sided and will be performed at the alpha level of 0.025, unless otherwise specified.

8.2. Testing Statistical Assumptions and Alternative Analysis

Estimated treatment effects such as the least squares (LS) means estimate from linear mixed models (eg, MMRM) and linear model (eg, ANCOVA) are known to be robust against moderate to high deviations from the normality assumption.

The underlying assumptions for the primary efficacy model will still be checked. The MMRM analysis will be used as the primary efficacy analysis for the 6MWD regardless of the results of the normality checks, as the MMRM model is believed to be robust against deviations from normality (Bartlett 2014). The nonparametric randomization-based covariance analysis described in Section 8.5.1.2 will be performed as a sensitivity analysis only if the normality assumption is notably violated with the Shapiro-Wilk test p-value < 0.01. The MMRM analysis assumes that data are missing at random (MAR). For this study, a lot of effort has been (and will continue to be) made to follow up all subjects so as to try to reduce the number of dropouts in the study. Despite all this, it is still possible to have subjects who could not do the 6MWD and/or FVC assessment due to COVID-19 related reasons. In view of the potential impact this can have on the efficacy assessments and analysis, the following analyses are considered:

- The MMRM analysis will be used as the primary efficacy analysis for the 6MWD, based on the ITT-OBS population.

  The MMRM analysis uses all available data without imputation for missing data, and the amount of missing data for 6MWD at Week 52 is expected to be small (< 10%). Under these conditions, the MMRM is preferable to the ANCOVA (White, Carpenter et al. 2012; Siddiqui, Hung, et al. 2009).

- The ANCOVA model remains the main analysis for the % predicted FVC and other continuous endpoints, based on the ITT-LOCF population, unless otherwise specified.

The MMRM is a likelihood-based method that does not require a complete dataset; it uses all available data across visits, allowing it to account for differences in variability between visits; and it can model the within-subject variability. These features of the MMRM may be important now in view of the COVID-related interruptions at study sites resulting in delayed visits and make-up assessments, with potentially high between-visit variability. This can also be helpful

given that this study enrolled both ERT-naïve and ERT-experienced subjects who potentially differ in variability.

The validity of the MAR assumption in this study will be assessed by examining the amount and reason of the missing data and assessing whether they are balanced between the treatment groups. If the amount of missing data is small (eg, less than 10% of the total sample size), and/or are balanced between the treatment groups by reason of the missing data, then the impact of the missing data will be expected to be small.

In addition, a 'control group imputation' algorithm based on data sampled from the control group (alglucosidase alfa/placebo) will be implemented to replace missing data at the post-baseline visit. For example: For each instance of a missing 6MWD result at Week 52 in either treatment group (ie, active control or new treatment), a random sample of 15 values will be selected (with replacement) from the available set of data in the control group at Week 52. The average of these 15 values will then be used to replace the missing value. These steps will be carried out until all missing data at Week 52 have been replaced. The planned analysis (eg, MMRM) will then be performed on these newly imputed data. Consistency in the results with the original analysis (that is, the treatment effects are in the same direction, and the conclusions are similar) will suggest that the MAR assumption is valid.

8.3. Statement of the Null and Alternative Hypotheses

8.3.1. Primary Null and Alternative Hypotheses

The primary endpoint is the change from baseline to Week 52 in 6MWD, measured in meters. The primary endpoint analysis is based on a comparison of the change in 6MWD between ATB200/AT2221 ('new treatment') and alglucosidase alfa/placebo ('control').

H_0: the new treatment is less effective than the control, with respect to the primary efficacy endpoint

H_A: the new treatment is more effective than the control, with respect to the primary efficacy endpoint

The primary null hypothesis will be tested at the 1-sided significance level of 0.025, unless otherwise stated.

8.3.2. Key Secondary Null and Alternative Hypotheses

There are 6 key secondary endpoints in hierarchical order of importance, as follows:

- Change from baseline to Week 52 in sitting FVC (% predicted)
- Change from baseline to Week 52 in the MMT score for the lower extremities
- Change from baseline to Week 26 in 6MWD
- Change from baseline to Week 52 in the total score for the PROMIS – Physical Function
- Change from baseline to Week 52 in the total score for the PROMIS – Fatigue
- Change from baseline to Week 52 in GSGC total score

The null and alternative hypotheses for superiority testing of all the key secondary endpoints are as follows:

H_{0i}: the new treatment is less effective than the control, with respect to the i^{th} key secondary efficacy endpoint

H_{Ai}: the new treatment is more effective than the control, with respect to the i^{th} key secondary efficacy endpoint

8.4. Multiple Comparisons and Multiplicity

The primary analysis is based on a single inter-group comparison of the primary efficacy endpoint between the new treatment and control at the 1-sided significance level of 0.025. The study has 6 key secondary efficacy endpoints which are ordered. The family-wise Type I error rate (FWER) for the statistical tests of the primary and key secondary endpoints will be controlled at 0.025. The p-value for comparing the primary and each key secondary endpoint will be calculated irrespective of its statistical significance.

To strongly control the FWER at this level, a gate-keeping approach will be utilized in which each family of statistical tests (family of primary and family of key secondary) will be conducted in a sequential manner. The test for the primary endpoint will be conducted first at the 0.025 significance level and, if significant, the ordered key secondary endpoints will be similarly tested at the 0.025 significance level. The statistical significance of the key secondary endpoints will be interpreted following a hierarchical testing order, each at the 1-sided alpha level of 0.025. If at any point the null hypothesis fails to be rejected, then that comparison and any other comparison below it cannot be claimed as successful.

The hypothesis tests included in the other secondary endpoints are not controlled, and each will be tested at the 1-sided alpha level of 0.025.

8.5. Analysis of the Primary Efficacy Endpoint

8.5.1. Primary Efficacy Analysis

The primary efficacy endpoint (change from baseline to Week 52 in 6MWD) will be analyzed using an MMRM model to compare between treatment and control on the ITT-OBS population.

For this analysis, the dependent variable is the change from baseline to all post-baseline visits. Independent variables in the model are the fixed, categorical effects of treatment, time (ie, visit), treatment-by-time interaction, ERT status, and gender, as well as the fixed, continuous covariates of baseline 6MWD, baseline age, baseline weight, and baseline height. SAS PROC MIXED will be used for the analysis, with restricted maximum likelihood estimation (REML) and unstructured within-subject covariance structure. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom and adjust standard errors for the tests of fixed effects. If this model fails to converge, a compound symmetry (CS) covariance structure will be used instead. The OBSMARGINS (OM) option in the LSMEANS statement will be specified to allow for the use of the observed marginal distributions of the categorical variable rather than assuming a balance among the levels of the categorical variable. The E option will also be specified to check for estimability of LS means.

All the relevant estimates from this model, including LS means, standard errors (SE), treatment differences in LS means, and the corresponding 95% CIs will be estimated for each time point. The main variable to be estimated (ie, the estimand) in the primary comparison is the difference between ATB200/AT2221 and alglucosidase alfa/placebo in mean change in 6MWD from baseline to Week 52, regardless of whether intercurrent events have occurred. The significance test will be based on the treatment comparison of LS means at Week 52 and a p-value will be presented for this time point only. Estimated LS means (± SE) for the change from baseline by treatment group will be plotted over time.

Exploratory Interaction Tests:

The study is not powered to detect any interaction. However, the following interactions will be explored.

Time-based interactions (ie, time-by-baseline 6MWD, time-by-baseline age, time-by-baseline weight, time-by-ERT status) will be examined separately (one at a time) in separate MMRM models to understand the changes over time.

Similarly, treatment-by-covariate interactions (treatment-by-ERT status, treatment-by-baseline 6MWD, treatment-by-age, treatment-by-gender, treatment-by-weight, and treatment-by-height) will be explored. Each interaction term will be explored using a separate MMRM model that includes the full set of covariates of model terms (time, treatment-by-time interaction, gender, baseline 6MWD [continuous], baseline age [continuous], baseline weight [continuous], and baseline height [continuous]) as well as the added interaction term. The treatment-by-covariate interaction terms with a 2-sided $p < 0.10$ will be further examined clinically. Forest plots will be used to present the subgroups defined by the baseline covariates.

Secondary analyses of the primary endpoint and % predicted FVC (including supportive and sensitivity analyses) are described in Section 8.5.1.1 and Section 8.5.1.2. Analysis of the key secondary endpoints are described in Section 8.6, other secondary endpoints in Section 8.8, and subgroup analyses in Section 8.9. In all analyses involving the use of the LSMEANS, the OM option (as described above) will be used where applicable. If the LS means are not estimable with the OM option but they are without it, then the analysis without the OM option will be used. This will be footnoted accordingly.

8.5.1.1. Supportive Analyses for the Primary Efficacy Endpoint and % Predicted FVC

The following supportive analyses will be performed:

Table 6: Supportive Analyses for the 6MWD and % Predicted FVC

	Primary Endpoint (6MWD)	% Predicted FVC
1	ANCOVA at Week 52 on ITT-LOCF	MMRM analysis on ITT-OBS
2	Rate of change in 6MWD based on 12-month slopes	Rate of change in % predicted FVC based on 12-month slopes

Abbreviations: 6MWD = 6-minute walk distance; ANCOVA = analysis of covariance; FVC = forced vital capacity; ITT-LOCF = intent-to-treat population with missing data replaced by the last available value; ITT-OBS = intent-to-treat population that includes all available, observed data without any missing data imputation at Week 52; MMRM = mixed-effect model for repeated measures

In addition, note that as a supportive analysis, each of the other key secondary endpoints will be analyzed using an MMRM analysis on the ITT-OBS population as described in Section 8.3.2.

Details for the supportive analyses are as follows:

- MMRM analysis: This is based on the ITT-OBS population, similar to what has been described for the primary endpoint analysis in Section 8.5.1. There is no need to explore the treatment-by-covariate interaction tests in using MMRM as supportive analysis for the key secondary endpoints.
- Rate of change using ANCOVA on the ITT-OBS population based on the 12-month slopes of the 6MWD assessments (ie, change in 6MWD per month), where the slope is estimated for each subject using a simple linear regression of the observed values. The mean slopes are then compared between the 2 treatment groups adjusting for the baseline value (continuous) and ERT status, as well as baseline age, gender, baseline height, and baseline weight. All the relevant LS estimates (including the LS means for each treatment group, LS mean difference, 95% CI for the LS mean difference, SE, and p-value for comparing the 2 treatment groups) will be provided.

8.5.1.2. Sensitivity Analysis for the Primary Endpoint and % Predicted FVC

The following sensitivity analyses will be considered for the primary endpoint and % predicted FVC.

Table 7: Sensitivity Analyses for the 6MWD and % Predicted FVC

	Primary Endpoint (6MWD)	**First Key Secondary Endpoint (% Predicted FVC)**
1	Nonparametric randomization-based covariance analysis on ITT-LOCF at Week 52 (if Shapiro-Wilk p-value < 0.01)	Nonparametric randomization-based covariance analysis on ITT-LOCF at Week 52 (if Shapiro-Wilk p-value < 0.01)
2	MMRM analysis incorporating the handling of intercurrent events	ANCOVA incorporating the handling of intercurrent events
3	ANCOVA on the per-protocol population 1 (PP1)	ANCOVA on the per-protocol population 2 (PP2)
4	ANCOVA on ITT-OBS, using the % predicted 6MWD	ANCOVA on ITT-LOCF for the subset of subjects with baseline % predicted FVC value < 85%, vs ≥ 85%; MMRM analyses in ITT-OBS for the subset of subjects with baseline % predicted FVC value < 85%, vs ≥ 85%
5	Randomization test applied to test statistics from the ANCOVA model	Randomization test applied to test statistics from the ANCOVA model

Table 7: Sensitivity Analyses for the 6MWD and % Predicted FVC (Continued)

	Primary Endpoint (6MWD)	First Key Secondary Endpoint (% Predicted FVC)
6	ANCOVA based on ITT-LOCF where externally studentized residuals that are large (eg, > 3) in magnitude (absolute value) are considered as outliers and are removed from the analysis	ANCOVA based on ITT-LOCF where externally studentized residuals that are large (eg, > 3) in magnitude (absolute value) are considered as outliers and are removed from the analysis

Abbreviations: 6MWD = 6-minute walk distance; ANCOVA = analysis of covariance; FVC = forced vital capacity; ITT-LOCF = intent-to-treat population with missing data replaced by the last available value; ITT-OBS = intent-to-treat population that includes all available, observed data without any missing data imputation at Week 52

Details of the sensitivity analyses are as follows:

- Nonparametric randomization-based covariance analysis (nonparametric ANCOVA): This analysis is based on Koch, Tangen et al. 1998 and LaVange, Durham et al. 2005. The SAS/IML® (Interactive Matrix Language) implementation of this procedure via the NPARCOV3 macro (Zink and Koch 2012) will be used. For this method, the ERT status (naïve or experienced) will be specified as STRATA, and baseline value, baseline age, gender, height, and weight will be specified as COVARIATES. The categorical variable (ie, gender) will be recoded as a numeric variable. No missing values are allowed in this macro. Due to the potential for smaller ERT-naïve stratum size, the weighted estimates will be combined across strata prior to covariance adjustment. The covariance matrix calculated under the alternative hypothesis will be used for statistical test to compare between the treatment groups. The relevant results including estimates of the treatment difference, SE, p-value for the treatment comparison, and the 95% confidence interval for the treatment difference will be provided. This is an ITT-LOCF analysis at Week 52, unless otherwise specified.
  - As stated in Section 8.2, this nonparametric method will be used as a sensitivity analysis for the primary endpoint only if the normality assumption for the primary MMRM analysis model is notably violated with the Shapiro-Wilk test p-value < 0.01.
  - This sensitivity analysis will also be used for the % predicted FVC if the normality assumption for the ANCOVA model on the ITT-LOCF population is violated with the Shapiro-Wilk test p-value < 0.01.
  - This nonparametric method will be used to compare the biomarker endpoints at Week 52 between the treatment groups based on the ITT-LOCF population.
- ANCOVA that incorporates the handling of the intercurrent events: This model will be adjusted for the baseline value (continuous) and ERT status, as well as baseline age, gender, baseline height, and baseline weight. See Section 6.4 for the list of intercurrent events.
- ANCOVA applied to the PP population: This model will be adjusted for the baseline value (continuous) and ERT status, as well baseline age, gender, baseline height, and baseline weight to compare between the 2 treatment groups.

- ANCOVA for % predicted 6MWD: ANCOVA using the change from baseline to Week 52 in % predicted 6MWD, where the % predicted 6MWD is calculated using the Enright and Sherrill 1998 reference equations. This analysis will be performed on observed data (ITT-OBS) and the model will be adjusted for baseline 6MWD (as a continuous variable) and ERT status to compare between the 2 treatment groups. The baseline age, gender, height, and weight will not be included as covariates in this ANCOVA analysis because they are already incorporated in the calculation of the % predicted 6MWD values.
- ANCOVA for the subset of subjects with baseline % predicted FVC value < 85% and the subset with baseline value ≥ 85%. This analysis will be performed on the ITT-LOCF population.
- MMRM analysis for the subset of subjects with baseline % predicted FVC value < 85% and the subset with baseline value ≥ 85%. This analysis will be performed on the ITT-OBS population.
- Randomization test applied to ANCOVA: Analysis using a randomization test applied to test statistics obtained from the ANCOVA model adjusted for baseline 6MWD (as a continuous variable), ERT status, baseline age, gender, baseline height, and baseline weight to compare between the 2 treatment groups. The p-value for comparing between the 2 treatment groups will be provided. This analysis will be performed on the ITT-LOCF population.

8.6. Analysis of the Key Secondary Endpoints

Key secondary endpoints will be analyzed according to the hierarchical order specified in Section 8.3.2 using a stepwise closed testing procedure to control the overall Type 1 error rate.

Each key secondary endpoint will be analyzed using an ANCOVA model on the ITT-LOCF population. The model will be adjusted for the baseline value (as a continuous covariate) and ERT status (ERT-naïve versus ERT-experienced), as well as baseline age, gender, baseline height, and baseline weight to compare between the 2 treatment groups. The OBSMARGINS (OM) option in the LSMEANS statement will be specified to allow for the use of the observed marginal distributions of the categorical variable rather than assuming a balance among the levels of the categorical variable. The E option will also be specified to check for estimability of LS means. If the LS means are not estimable with the OM option but they are without it, then the analysis without the OM option will be used.

All the relevant estimates from the ANCOVA (including LS means for each treatment group, LS mean difference, standard error of the estimate, 95% CI for the LS mean difference, and p-value for comparing between the 2 treatment groups) will be provided.

The treatment-by-covariate interactions (ie, treatment-by-ERT status, treatment-by-baseline value, treatment-by-age, treatment-by-gender, treatment-by-weight, and treatment-by-height) will be explored. Each interaction term will be explored using a separate ANCOVA model that includes the full set of covariates (ie, treatment, baseline value (as a continuous variable), ERT status, baseline age, gender, height, and weight) as well as the interaction term. Interaction terms with a 2-sided $p < 0.10$ will be further examined clinically. Forest plots will be used to present

the subgroups defined by the baseline covariates. A plot of the LS mean (± SE) change from baseline over time will be provided for each key secondary endpoint.

As a supportive analysis, each of the key secondary endpoints will be analyzed using an MMRM analysis on the ITT-OBS population. See further description in Section 8.5.1.1.

For the % predicted FVC, additional supportive and sensitivity analyses will be performed as specified in Section 8.5.1.1 and Section 8.5.1.2, respectively.

If the superiority tests based on the ANCOVA analysis using ITT-LOCF specified for 6MWD and % predicted FVC fail, an informal determination for non-inferiority (NI) for the % predicted FVC will be made based on the 2-sided 95% CI for the treatment difference (new treatment – control). That is, if the lower bound of the 95% CI is equal to or greater than -1.1% (which is the assumed NI margin), then NI would be deemed to have been met.

8.7. Composite Subject-level Response Analysis

As a supplementary analysis to the primary and key secondary endpoints for the study, and to support the interpretation of further clinical benefit, an analysis using a composite subject-level response will be performed based on a multi-domain responder index (MDRI). This MDRI will be defined to capture the totality of evidence using the 3 relevant clinical outcomes, 6MWD, FVC (% predicted), and MMT lower extremity score. The MDRI approach provides a framework for presenting the totality of data combining the 3 components which measure motor function, pulmonary function, and muscle strength, which are the domains that are impacted in individuals with LOPD. A clinically meaningful threshold for within-subject changes in each clinical outcome and the categorization of each outcome are provided in Table 8.

Table 8: Threshold for Defining Clinically Meaningful Within-subject Change

Clinical Outcome	Declining	Stable	Improving
6MWD[a]	< -6%	-6% to < +6%	≥ +6%
% predicted FVC[b]	< -3%	-3% to < +3%	≥ +3%
MMT lower extremity score[a]	< -7%	-7% to < +7%	≥ +7%

[a] Threshold is based on the percentage change from baseline.
[b] Threshold is based on the change from baseline.

The above thresholds are consistent with published MCID (minimal clinically important differences) values for comparable instruments in similar disease populations (Schrover, Evans et al. 2017; du Bois, Weycker et al. 2011; Baschung Pfister, de Bruin et al. 2018).

The ordinal composite response across all 3 ternary outcomes is defined as follows:

- 3 = significant improvement: subject is either improving on all 3 outcomes, or is improving on both 6MWD and % predicted FVC;
- 2 = moderate improvement: subject is improving on either 6MWD only, or % predicted FVC only, or both 6MWD and MMT lower extremity, or both % predicted FVC and MMT lower extremity;
- 1 = minor: subject is improving on MMT lower extremity score only;

- 0 = no improvement: subject is not improving on any of the 3 clinical outcomes.

The main interest for this analysis is the inter-group comparison. There is no intra-subject comparison for the MDRI approach.

This method defines a hierarchical ordinal categorical response variable, for which a continuation logit model (Agresti 2002) will be used to perform separate logistical regression analyses adjusting for baseline risk factors including ERT status (ERT-naïve or ERT-experienced), duration of prior alglucosidase alfa (only for ERT-experienced subjects), gender, and baseline age.

The results of these separate logistic regression analyses are then pooled to test the overall null hypothesis that treatment with ATB200/AT2221 is less effective than alglucosidase alfa/placebo with respect to either the positive composite response (ie, significant improvement) or the negative response (ie, minor/no improvement). The inter-group comparison will be based on the combined analysis of comparisons of the positive response rates of significant improvements and the negative response rates of minor/no improvement.

Additional details regarding the separate logistic regression analyses and test against the overall null hypothesis are described below.

The "significant improvement" response category is considered as positive response, the "moderate improvement" category is considered as stable response, and "minor/no improvement" is considered as negative response. Treatment (ie ATB200/AT2221) is expected to have high positive responses and low negative responses, while the control (ie, alglucosidase alfa/placebo) is expected to have high negative responses and low positive responses). Separate logistic regression models are used to analyze and test (1) positive responses vs. not positive responses (where "not positive responses" includes the "moderate improvement" and "minor/no improvement" categories; (2) negative responses vs. stable responses, conditional on subjects who are not positive responders.

The respective hypotheses for Test 1 and Test 2 are stated as follows: H_{10}: $\delta_1 \leq 0$ vs. H_{1A}: $\delta_1 > 0$ and H_{20}: $\delta_2 \leq 0$ vs. H_{2A}: $\delta_2 > 0$, where δ_1 and δ_2 are the effect sizes for comparing treatment to control for Test 1 and Test 2 respectively.

Let the logistic regression model for Test 1 be:

$$\text{logit}(p_{1,Positive}) = \ln\frac{p_{1,Positive}}{p_{1,Not\ pos}} = \beta_{1,0} + \beta_{1,1}\text{TRT} + \beta_{1,2}\text{ERT} + \beta_{1,3}\text{ DUR} + \beta_{1,4}\text{ SEX} + \beta_{1,5}\text{ AGE}$$

and Test 2 be:

$$\text{logit}(p_{2,Negative}) = \ln\frac{p_{2,Neg}}{p_{2,Stable}} = \beta_{2,0} + \beta_{2,1}\text{TRT} + \beta_{2,2}\text{ERT} + \beta_{2,3}\text{ DUR} + \beta_{2,4}\text{ SEX} + \beta_{2,5}\text{ AGE}$$

For any coefficient β, the Wald statistic is given by $Wald = b/se_b$, where b is the estimate for β and se_b is the standard error for β, and the Wald statistic is approximately normally distributed. Therefore, in testing the null hypothesis about the treatment effect, say $\beta_{1,2} = 0$ (for Test 1), the test statistic for treatment effect is: $(b_{1,2} - 0)/se_b = b_{1,2}/se_b$.

Let $Z_1 = b_1/se_{b1}$ and $Z_{2/1} = b_{2/1}/se_{b2/1}$ be the test statistics for testing treatment effect in Test 1 and Test 2 respectively, where Z_1 and $Z_{2/1}$ are standardized to have standard normal distributions (so that each has mean = 0 and variance = 1).

The inter-group comparison is based on the combined analysis of the positive response rates of significant improvements and the negative response rates of minor/no improvement. The overall null and alternative hypothesis (H_0: $H_{10} \cap H_{20}$) is tested against the alternative hypothesis (H_A: $H_{1A} \cup H_{2A}$) based on the combination test statistic:

$Z = \frac{w_1}{w_1+w_2} Z_1 + \frac{w_2}{w_1+w_2} Z_{2/1}$, where w_1 and w_2 are weights (constants) for the random variables Z_1 and $Z_{2/1}$ respectively (Rpatel 2019). The weights, w_i, represent the information fraction in test i. The expected value of Z is given by:

$E(Z) = \frac{w_1}{w_1+w_2} E(Z_1) + \frac{w_2}{w_1+w_2} E(Z_{2/1}) = 0$, under H_0.

Assuming independence of Z_1 and $Z_{2/1}$, the variance of Z is given as:

$$Var(Z) = \left(\frac{w_1}{w_1+w_2}\right)^2 Var(Z_1) + \left(\frac{w_2}{w_1+w_2}\right)^2 Var(Z_{2/1})$$
$$= \left(\frac{w_1}{w_1+w_2}\right)^2 + \left(\frac{w_2}{w_1+w_2}\right)^2$$

Inference is based on the standardized score Z_W, where

$$Z_W = \frac{\frac{w_1}{w_1+w_2} Z_1 + \frac{w_2}{w_1+w_2} Z_{2/1}}{\left[\left(\frac{w_1}{w_1+w_2}\right)^2 + \left(\frac{w_2}{w_1+w_2}\right)^2\right]^{1/2}}$$

follows the standard normal distribution under the null hypothesis. A 1-sided p-value for testing against the overall null hypothesis H_0 is given by $p = 1 - \Phi(z_w)$, where $\Phi(.)$ is the cumulative distribution function of a random variable that follows the standard normal distribution.

8.8. Analysis of Other Secondary Endpoints

8.8.1. Motor Function Test

Motor function test endpoints (including % predicted 6MWD time to complete TUG, and time to complete individual GSGC components) will be summarized by treatment group and visit. The change from baseline to Week 52 will be analyzed and compared between the 2 treatment groups using an ANCOVA model on the ITT-LOCF population, similar to that for the key secondary endpoints. No supportive and/or sensitivity analyses will be performed for these motor function tests.

8.8.2. Muscle Strength Tests

Muscle strength test endpoints (including MMT and QMT) will be summarized by treatment group and visit for the lower extremity score, upper extremity score, total MMT score, and the proximal muscle group score. In addition, the change from baseline to Week 52 for each endpoint measure will be analyzed and compared between the 2 treatment groups using an

ANCOVA model on the ITT-LOCF population, similar to that for the key secondary endpoints. A plot of the LS mean (± SE) change from baseline over time will be provided by treatment group.

A supportive analysis for the MMT lower extremity (right/left hip flexion, right/left hip abduction, right/left knee flexion, and right/left knee extension) will be conducted. For this analysis, the number of subjects with changes in score of ≤ -3, -2, -1, 0, 1, 2, or ≥ 3 (from baseline to the Week 52 visit) will be tabulated for each body part, and then summed across all 8 body parts for each treatment group. For this analysis, subjects must have tested on the same muscles (right or left) at both baseline and post-baseline visit in order to determine the point change. The summed changes in scores by categories across the 8 body parts will then be summarized (by number and percentage) for each treatment group. No inferential comparisons will be performed. Similar summaries will be performed for the body parts associated with upper extremity (right/left shoulder abduction, right/left shoulder adduction, right/left elbow extension, and right/left elbow flexion), overall MMT (lower extremity and upper extremity combined), and the proximal muscle group.

No supportive or sensitivity analysis will be performed for the QMT scores.

8.8.3. Other Pulmonary Function Tests

The other pulmonary function test endpoints (ie, % predicted SVC, maximum VC [% predicted], % predicted MIP, % predicted MEP, and SNIP [% predicted]) will be summarized by treatment group and visit. In addition, the change from baseline to Week 52 for each endpoint will be analyzed and compared between the 2 treatment groups using an ANCOVA model on the ITT-LOCF population, similar to that for the key secondary endpoints. No sensitivity or supportive analyses will be conducted for these.

A plot of the LS mean (± SE) change from baseline over time will be provided for the pulmonary function test (PFT) endpoints by treatment group. In addition, the difference between sitting FVC and supine FVC results will be summarized by visit.

The % predicted SVC will be calculated as: (observed SVC results / predicted value)*100, where the predicted value is obtained from the corresponding VCMax predicted value in the dataset

8.8.4. Patient-reported Outcomes

The total score and change from baseline in total score will be summarized for the R-PAct scale, EQ-5D-5L (ie, based on the quantitative EQ-VAS score), and PROMIS instruments. The change from baseline to Week 52 in the total score and the EQ-VAS score will be analyzed using an ANCOVA model on the ITT-LOCF population, similar to that for the key secondary endpoints. A plot of the LS mean (± SE) change from baseline over time will be provided for each of these PROs by treatment group. The EQ-5D categorical responses and the changes in the responses will be summarized by visit for each of the 5 dimensions.

Handling Baseline Missing Data for PROs:

For the PROs, missing baseline total scores will be handled similarly to that described for the multi-item endpoints in Section 6.3.1. That is, a missing baseline total score will be imputed with the average of all non-missing baseline total scores (from both treatment groups combined). If

the subject is only missing specific item(s) at baseline, the average of all non-missing values for that specific item (from both treatment groups combined) will be used.

Note: a missing baseline total score (or a missing individual item score) will only be imputed if the subject has at least 1 non-missing post-baseline total score (or individual item score).

Handling Post-baseline Missing Data for PROs:

For any PRO with a total score based on multiple items/questions, a missing post-baseline total score at a given visit will be handled using the "50% prorating rule" described below:

- If ≥ 50% of items are available, the total score will be calculated and prorated as the average of the non-missing items multiplied by the total number of items expected (Haywood, Garratt et al. 2004);
- If < 50% of items are available, the total score will not be calculated, and it will be set to missing (in the ITT-OBS population). For the ITT-LOCF population, missing data will be imputed as described in Section 6.1.4.1 and Section 6.3.2.

For the SGIC, the responses for each of the 8 items will be summarized by treatment group and visit. Additionally, the response scale for each item at Week 52 will be divided into 3 categories (improving, stable, or declining) which reflect the functional status, and these will be summarized by treatment group.

For the PGIC, a summary of the response score will be provided by treatment group and visit.

8.8.5. Proportion Improving on Both 6MWD and % Predicted FVC

The proportion of subjects improving on both 6MWD and % predicted FVC (where definition of improvement is provided in Section 8.7) will be compared between the 2 treatment groups using a logistic regression model adjusting for baseline age, ERT status, gender, baseline height, and baseline weight. The number and percent of subjects improving will be presented for each treatment group, and the p-value for comparing between the 2 treatment groups and the 95% confidence interval for the treatment difference will be calculated from the logistic regression model.

8.8.6. Analysis of Pharmacodynamic Endpoints

For the pharmacodynamic endpoints/biomarkers (CK and urinary Hex4), the change from baseline and percent change from baseline at each visit will be summarized by treatment group. A line plot over time of the mean ± SE for the change and percent change from baseline will be provided by treatment group. The change from baseline to Week 52 will be compared between treatment groups using the nonparametric randomization-based covariance analysis described in Section 8.5.1.2. Analysis will be performed on the ITT-LOCF population.

8.8.7. Immunogenicity

This study will assess anti-rhGAA antibodies and NAbs in subjects treated with ATB200/AT2221 and alglucosidase alfa. The effect of immunogenicity results on PK, efficacy, safety, and PD will be explored. Further details will be provided in the MSP mentioned in Section 3.3.5.3 for the analysis of immunogenicity and PD data.

- Total ADA, including titers, will be assessed using the ATB200 assay, regardless of what treatment arm the subject is in, because the assay is set up for use with either treatment.
- NAbs (inhibition of rhGAA-mediated hydrolysis of 4-MU glucoside, inhibition of rhGAA-mediated hydrolysis of glycogen, and inhibition of rhGAA binding to CI-MPR (for ATB200 analysis only) will be assessed using ATB200 in the assays, regardless of the treatment arm, because the assays are set up for use with either treatment (with the exception of the CI-MPR assay).
- CI-MPR will be handled slightly differently as follows: All samples will be analyzed in the CI-MPR assay regardless of treatment arm, but data will be reported/used for further analysis only for subjects treated with ATB200 (the assay cannot be set up for use with alglucosidase alfa).
- GAA cross-reactivity (anti-rhGAA antibodies cross-reactive to alglucosidase alfa) is also included in the protocol and will be assessed regardless of treatment arm (Screen/confirm analysis only; ie, no titers will be assessed with the GAA assay).
- Anti-rhGAA IgE will be assessed as needed regardless of treatment arm.

8.9. Subgroup Analyses

Subgroup analyses by age group, gender, and race (using all available categories) will be performed for the primary endpoint and the change from baseline to Week 52 in % predicted FVC.

In addition, the following subgroups will be analyzed for the primary endpoint and the change from baseline to Week 52 in % predicted FVC, unless otherwise specified:

- ERT status (categorized as: ERT-experienced versus ERT-naïve)

  This subgroup analysis will apply to the primary and key secondary endpoints, % predicted SVC, as well as for the overall adverse event (AE) overview table, summary of treatment-emergent AEs by system organ class and by preferred term, and PD endpoints.
- Baseline 6MWD (categorized as: 75 to < 150 meters, 150 to < 400 meters, ≥ 400 meters)

  Note that for this subgroup analysis, categories may be collapsed depending on the actual number of subjects in each category, eg, 75 to < 150 meters and 150 to < 400 meters may be combined into 75 to < 400 meters.
- Baseline 6MWD < *median value*, baseline 6MWD ≥ *median value*
- Baseline FVC < *median value*, baseline FVC ≥ *median value*
- Regions (North/South America, Europe, Asia Pacific)
- ERT duration categorized as: 2 to < 3, 3 to < 5, and ≥ 5 years
- History of IARs (Yes, No)

All subgroup analysis results will be presented by forest plots. The p-values for the treatment-by-subgroup interactions will be presented together in a tabular form.

Note that subgroup analyses will not be performed if the size of the specified subgroup category is small and the analyses will not be meaningful.

9. ANALYSIS OF SAFETY

Safety will be evaluated through the analyses of AEs, clinical safety laboratory data, vital signs, physical examination, electrocardiogram (ECG), and concomitant medications.

Safety analyses will be based on the safety population. Safety data will be summarized using counts and percentages for categorical data and using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum) for continuous data. No formal statistical tests will be performed.

Infusion-associated reactions in the ATB200/AT2221 arm will be compared in frequency to the alglucosidase alfa/placebo arm in this study.

Adverse events will be coded by system organ class (SOC) and preferred term (PT) using MedDRA version 20.1 or higher.

Concomitant medications and nondrug therapies will be summarized by treatment group. Concomitant medications and nondrug therapies will be coded using WHO DDE, March 2017 or later.

9.1. Treatment-emergent Adverse Events

Summary of adverse events will be based on TEAEs. A TEAE is defined as any event that started or changed in the severity on or after the first dose of study drug.

For AEs occurring on Day 1, if the start time of the event is missing or after the administration of the first dose of study drug, the event will be considered a TEAE.

9.1.1. Adverse Event Reporting Period

In general, TEAEs that occur more than 30 days from the last dose of study drug will not be counted.

- For subjects who permanently discontinue study drug prematurely or who complete the treatment period and do not continue or enter the optional OLE study, ATB200-07, their AEs will be tabulated up to 30 days after the last treatment visit in Study ATB200-03. Their TEAE tabulation period (days) is calculated as:

  (last dose date of study drug + 30 days) – first dose date + 1.

- For subjects who are confirmed to have a positive result for anti-rhGAA antibodies at the time of discontinuation and up to the 30-day safety follow-up visit, they will continue to have follow-up immunological testing for up to 12 months (ie, at 6 months and 12 months) after the last dose date of study drug, but the AE collection will be up to 30 days from the time of study discontinuation.

  In order to generate summary results for the clinical study report (CSR), the summary will use only the events (ie, AEs and immunogenicity) that occurred within the first 30 days after the last dose of study drug in Study ATB200-03. The additional potential immunogenicity events from Month 1 to Month 12 post the date of last dose of study drug will be reported as an addendum to the ATB200-03 CSR.

9.1.2. Summary of Treatment-emergent Adverse Events

The overall summary of TEAEs table will be presented by treatment group and for all treated subjects using the number and percentage of subjects. The following events will be summarized:

- Any treatment-emergent adverse events (any TEAE)
- Treatment-emergent serious adverse events (TESAEs)
- TEAEs leading to study drug discontinuation
- TEAEs related to study drug (ie, treatment related TEAEs)
- TEAEs related to study drug and leading to study drug discontinuation
- TESAEs leading to study drug discontinuation
- TESAEs related to study drug (ie, treatment related TESAEs)
- TESAEs related to study drug and leading to study drug discontinuation
- Deaths

In addition, summary tables will be presented for each of the individual bulleted items listed above.

The number and percentage of subjects who experienced TEAEs will be presented by SOC and by PT within SOC for each treatment group. The table by SOC and PT will be repeated, one with the PTs within SOC sorted according to descending order of the ATB200/AT2221 group, and another the PTs sorted alphabetically within the SOCs. A subject will be counted only once within the same SOC and within the same PT.

The cumulative number and percentage of subjects with TEAEs will be summarized by severity (mild, moderate, severe) for each treatment group, and by relationship to study drug ('related' vs 'not related') for each treatment group. If a subject experienced more than 1 TEAE within different severity or relationship categories within the same SOC/PT, only the worst case (worst severity and related TEAE) will be reported.

All AEs will be provided in a listing that will include the subject identifier, SOC, PT, and the reported term, date of onset and study day, date of resolution and study day, the seriousness, the severity, the relationship to each individual study drug (for TEAEs), duration, the action taken, and the outcome. All TEAEs will be flagged in the listing.

Listings will also be generated for deaths, SAEs, and discontinuations due to TEAEs.

9.1.3. TEAEs Reported as Infusion-associated Reaction

For any AE, the reporting investigator is requested to enter on the AE eCRF whether the AE is deemed to be an infusion-associated reaction and determine the relationship to each of the study drug components (ATB200, AT2221, alglucosidase alfa, and placebo).

To further characterize TEAEs considered by the reporting investigator to be IARs, a summary table of TEAEs reported as IARs will be provided. The number and percentage of subjects who experienced IARs will be presented by SOC and by PT within SOC for each treatment group. A subject will be counted only once within the same SOC and within the same PT.

Summary of IARs will also be presented by severity, and for relationship to each treatment component (ATB200, AT2221, alglucosidase alfa, and placebo) as well as for the co-administration. Additional summary tables/plots will be provided to describe the clinical characteristics of the IARs by treatment group, including the event rates, changes in severity over time, and time to first IAR (see Section 9.4.1). The IAR summary will also be presented by prior history of IARs (yes/no), and by ERT status (experienced/naïve).

A listing of treatment-emergent IARs, serious IARs, IAR-related deaths, and IARs leading to study drug discontinuation will be provided. A listing displaying the summary of number of infusions and infusions associated with IARs by each individual subject who had an IAR will also be provided. For each subject, this listing provides subject ID, number of infusions received, number of infusions associated with IARs, percentage of infusions associated with IARs, the TEAEs that are IARs (IAR-TEAEs) occurring in each of the time intervals (see Section 9.1.4.2 and Section 9.1.4.3), and the event severity of the IARs corresponding to (or associated with) each infusion.

A summary of IAR-TEAEs by subjects with IARs, and the number of subjects with 1, 2, 3, 4 to 6, 7 to 10, 11 to 19, and ≥ 20 IARs will be summarized for each treatment group. A summary of the number of infusions and IARs associated with the infusions will be provided for all subjects as well as only for subjects who had IARs.

9.1.4. Additional Analysis of IARs

Additional summary tables/plots will be provided to describe the clinical characteristics of the IARs by treatment group, including the event rates, changes in severity over time, and time to first IAR.

9.1.4.1. Analysis of Time to First IAR-TEAE

This calculation is only performed for subjects who have IARs. For these subjects, the time (in weeks) from the first dose of study drug to the onset of the first TEAE that is an IAR (ie, IAR-TEAE) will be calculated for each subject as:

> Time to first IAR-TEAE (Weeks) = (date of the first IAR-TEAE – date of first dose of study drug + 1) / 7.

The calculated times will also be presented in a subject listing.

9.1.4.2. Analysis of IARs by Onset Intervals

The IARs will also be analyzed by onset intervals categorized as follows:

- 1 to 3 days, 4 to 7 days, 8 to 14 days, 15 to 28 days, 29 days to ≤ 3 months, > 3 months to ≤ 6 months, > 6 months to ≤ 12 months, and > 12 months.

The event rates and changes in severity over time (over these intervals) will be summarized for each treatment group. This analysis will further describe the clinical characteristics of IARs and help determine whether the IARs disappear over time or appear late.

9.1.4.3. Analysis by Time After Infusion Start

IARs and anaphylactic reactions will be analyzed by the time after infusion start. Based on protocol descriptions of immediate-type IARs and late-type infusion reactions, the following time after infusion start will be used:

- 0 to < 2 hours, 2 to < 4 hours, 4 to < 6 hours, 6 to < 12 hours, 12 to < 24 hours (1 day), 1 to < 4 days (96 hours).

The event rates and changes in severity over time (over these intervals) will be summarized to help describe the clinical characteristics of IARs and anaphylactic reactions.

9.1.5. Describing Relationship to Study Drug

For each AE entered on the AE eCRF, the reporting investigator was asked to assess relationship to both infusion drug (ATB200 or alglucosidase alfa) and capsule (AT2221 or placebo). The relationship to each study drug will be presented by 2 categories:

- "related," which includes possibly related, probably related, and definitely related, as reported by the investigator
- "not-related," which includes unrelated and unlikely related, as reported by the investigator

To describe the overall relationship to study drug, two approaches will be presented:

1. Based on the pooling into one designation of the two individual relationships to ATB200 or alglucosidase alfa and AT2221 or placebo
    a. If the two categories of the individual relationships are discordant (ie, "related" to one and "not related" to the other), the pooled designation will be considered "related"
    b. If the two categories of the individual relationships are concordant (ie, "related" to both or "not related" to both), the pooled designation will be concordant with the individual categories
2. Based on the assessment of causality to the individual component, yielding these possible combinations
    a. "Related" to ATB200 – "related" to AT2221
    b. "Related" to ATB200 – "not related" to AT2221
    c. "Not related" to ATB200 – "related" to AT2221
    d. "Not related" to ATB200 – "not related" to AT2221
    e. "Related" to alglucosidase alfa – "related" to placebo
    f. "Related" to alglucosidase alfa – "not related" to placebo
    g. "Not related" to alglucosidase alfa – "related" to placebo
    h. "Not related" to alglucosidase alfa – "not related" to placebo

9.1.6. Missing AE Onset Date, Severity, and Relationship

9.1.6.1. Missing or Partial AE Onset or Medication Start Date

Complete dates are very important to correctly describe safety events such as TEAEs and designation of unique AE occurrences. When dates are completely missing or partially missing, they will be imputed.

The following algorithm will be used to impute partial or missing AE start dates:

1. If Year is not missing and is after the year of first application of study drug:

   If Month is missing, then Month will be imputed as January.

   If Day is missing, then Day will be imputed as the first of the month.

2. If Year is not missing and is the same as the year of the first application of study drug:

   If Month is missing, then impute the Month as the month of the first application of study drug.

   If Day is missing but Month is the same as the month of first application of study drug, then impute Day as the first day of study drug application.

   If Day is missing but Month is after the month of first application of study drug, then impute Day as the first of the month.

3. If Year is missing, then impute the year as the year of the first application of study drug:

   If Month is missing, then impute the Month as the month of the first application of study drug.

   If Day is missing, then impute the Day as the day of the first application of study drug.

4. If the start date is completely missing, but the AE is either ongoing or the stop date is after the first date of drug application, then impute the start date as the first application of study drug.

5. If using the above rules, the stop date is before the start date, then leave the start date missing and assume that the AE is treatment-emergent for the purpose of the analysis.

No imputations will be applied to AE stop dates.

9.1.6.2. Missing or Partial Medication Stop Date

References to year and month are the year and month of the stop date:

1. If year and month are known and study medication stopped during that month and year, use the stop date of study medication.

2. If year and month are known and informed consent was provided during that month and year, use the date of informed consent.

3. If year and month are known and study medication stopped after the date of informed consent and not in the month that medication stopped, use the last day of the month.

4. If year and month are known and are prior to the month of informed consent, use the first day of the month.
5. If only year is known and study medication stopped during that year, use the stop date of study medication.
6. If only year is known and study medication stopped after that year, use December 31st of that year.
7. If only year is known and study medication stopped prior to that year, use the first day of the year.
8. Should any of the previous stop dates be created before a start date, either a complete date or an imputed one, use the (imputed) start date instead of the date that would otherwise be created.
9. Otherwise, if stop date is unknown, leave as missing.

9.1.6.3. Missing Severity

If severity is missing for any AE/TEAE, then its severity will be classified as "severe" in the summary tables.

9.1.6.4. Missing Relationship

If the assessment of relationship of the TEAE to study drug is missing, then it will be classified as "related" in the summary tables.

9.2. Clinical Safety Laboratory Data

Laboratory data will be presented using the Système International (SI) unit. Descriptive statistics for clinical safety laboratory data will be presented by treatment group and by visit, starting at the Baseline visit. Analyses will be based on the safety population. Change from baseline and percent change from baseline to post-baseline visit will also be presented. No formal statistical comparisons will be performed. Data will be attributed to visits according to the visit windows in Table 4.

Summary tables by treatment group will be presented for each category of data separately. Routine clinical laboratory data will include hematology, serum chemistry, and urinalysis. Quantitative laboratory test results will be summarized using descriptive statistics (n, mean, SD, median, Q1, Q3, minimum, and maximum). Line plots (for ALT and AST) over time will be provided by treatment group using the mean ± SE for the change from baseline and percent change from values. Qualitative tests (eg, some urinalysis assessments) will be categorized accordingly. The set of laboratory parameters included in each table will correspond to those requested in the study protocol.

The percentage of subjects with specific treatment-emergent laboratory values that meet pre-defined limits of change (PDLC) will be summarized for these laboratory parameters.

The incidence rates of PDLC values will be compared descriptively between treatment groups, without any statistical testing.

If a subject has even 1 post-baseline value that meets the PDLC criteria for any lab parameter, then all values of that parameter for that subject will be listed.

Table 9: Pre-defined Limits of Change Criteria for Clinical Safety Laboratory Data

Laboratory Test	Parameter for ANY Value and LAST Value
Chemistry	
Albumin	Composite: < LLN and > 25% decrease from BL
Bilirubin	Composite: > ULN and > 25% increase from BL
	Absolute Value: > 2X ULN
Bicarbonate	Absolute Value: < 16 mEq/L
Calcium	Composite: > ULN and > 10% increase from BL
Phosphorus	Composite: < LLN and > 25% decrease from BL
	Composite: > ULN and > 25% increase from BL
Potassium	Composite: < LLN and > 15% decrease from BL
	Composite: > ULN and > 15% increase from BL
Sodium	Composite: < LLN and decrease > 5 mEq/L or more from BL
	Composite: > ULN and increase > 5 mEq/L or more from BL
Uric Acid	Male: absolute value ≥ 624.6 μM
	Female: absolute value ≥ 505.6 μM
	Composite: < LLN and > 25% decrease from BL
	Composite: > ULN and > 25% increase from BL
ALT	Absolute Value: > 3X ULN
	Absolute Value: > 5X ULN
	Absolute Value: > 8X ULN
AST	Absolute Value: > 3X ULN
	Absolute Value: > 5X ULN
	Absolute Value: > 8X ULN
ALT > 3X ULN and Tbili > 2X ULN	Composite: ALT > 3X ULN and Tbili > 2X ULN (with the Tbili elevation > 2 X ULN within 30 days of the ALT elevation > 3X ULN)
AST > 3X ULN and Tbili > 2X ULN	Composite: AST > 3X ULN and Tbili > 2X ULN (with the Tbili elevation > 2X ULN within 30 days of the AST elevation > 3X ULN)
CK	Absolute CK ≥ 2X ULN or 2X baseline value

Table 9: Pre-defined Limits of Change Criteria for Clinical Safety Laboratory Data (Continued)

Laboratory Test	Parameter for ANY Value and LAST Value
Hematology	
Hemoglobin	Male: absolute value ≤ 11.5 g/dL
	Female: absolute value ≤ 9.5 g/dL
	Change: ≥ 2 g/dL decrease from BL
	Change: ≥ 2 g/dL increase from BL
Platelets	Absolute value: ≤ 75 x 10^9/L
	Absolute value: ≥ 700 x 10^9/L
	Composite: > ULN and increase > 25% from BL
	Composite: < LLN and > 25% decrease from BL
White Blood Count	Composite: < LLN and > 25% decrease from BL
	Composite: > ULN and > 50% increase from BL
Eosinophils	Absolute value: > 10%

Abbreviations: BL = baseline; LDH = lactate dehydrogenase; LLN = lower limit of normal; PDLC = pre-defined limit of change; SGOT (AST) = serum glutamic oxaloacetic transaminase (aspartate aminotransferase); SGPT (ALT) = serum glutamic pyruvic transaminase (alanine aminotransferase); Tbili = total bilirubin; ULN = upper limit of normal

Supportive data listings for all laboratory results will be generated by laboratory type (hematology, serum chemistry, and urinalysis). Subject data listings will include the type of visit (eg, scheduled test, retest, or unscheduled), age, sex, laboratory test, test units, laboratory test result, and the laboratory standard normal ranges adjusted as appropriate for age and sex, if available. Laboratory values outside of the normal range will be flagged in the data listings.

Results of urine pregnancy tests will be presented in a listing only.

9.3. Vital Signs, Physical Findings, and Other Observations Related to Safety

9.3.1. Vital Signs

Descriptive statistics for vital signs will be presented by treatment group for systolic blood pressure, diastolic blood pressure, heart rate, and body weight at each visit starting at baseline. Change from baseline to post-baseline visit will be presented. No formal statistical comparisons will be performed. Data will be attributed to visits as described in Section 6.2.6.

The incidence rates of PDLC for vital sign by parameter will be summarized (altogether, not by visit).

The incidence rates will be presented by treatment group. There will be no formal statistical testing.

A listing will present all values for a subject and vital sign parameter if at least 1 post-baseline value for that subject and parameter meets the PDLC criterion.

Table 10 presents criteria for categorizing the change from baseline in vital signs as PDLC.

Table 10: Pre-defined Limits of Change Criteria for Vital Signs

Vital Sign	Criteria
Pulse rate	≥ 120 bpm at any time post-dose and ≥ 15 bpm increase from baseline at any time post-dose
	≤ 50 bpm at any time post-dose and ≥ 15 bpm decrease from baseline at any time post-dose
Systolic blood pressure	≥ 180 mm Hg at any time post-dose and ≥ 20 mm Hg increase from baseline at any time post-dose
	≤ 90 mm Hg at any time post-dose and ≥ 20 mm Hg decrease from baseline at any time post-dose
Diastolic blood pressure	≥ 105 mm Hg at any time post-dose and ≥ 15 mm Hg increase from baseline at any time post-dose
	≤ 50 mm Hg at any time post-dose and ≥ 15 mm Hg decrease from baseline at any time post-dose
Weight	Change: > 5% increase from BL
	Change: > 5% decrease from BL

Abbreviations: bpm = beats per minute; mm Hg = millimeters mercury; PDLC = pre-defined limit of change

9.3.2. Physical Examination

Complete physical exams are performed at screening and the Week 52 visit. Brief physical exams are performed at Day 1, Weeks 2, 4, 6, 12, 26, and 38. Baseline physical examination results will be presented for each body system/category examined. The number and percentage of subjects judged to be normal, abnormal, or not performed will be summarized.

Further, the results of the physical examinations will be summarized by visit for subjects in the safety population who had an examination post-baseline. Each site/system will be summarized with respect to being normal or abnormal, or not performed.

The supportive data listing will include the information collected on the eCRF (eg, body system/category, result of the observation [eg, normal or abnormal] or not performed, and any investigator comment).

9.3.3. Electrocardiograms

Electrocardiogram data including heart rate [HR] (beats/min), PR interval [PR] (msec), QRS duration (msec), QT interval (msec), QTcB interval (msec), QTcF interval (msec), and R-R interval [RR] (msec) will be summarized at baseline.

The ECG parameters will be summarized by treatment group and visit. Frequency counts will be presented per treatment for both change and absolute values, with changes categorized as per the

bullets below. Individual listings presenting subjects with flags will be created for both change and absolute values. The summary will be tabulated by treatment group.

Additionally, the number and percentage of subjects with the following overall ECG result will be presented:

- Normal
- Abnormal, not clinically significant
- Abnormal, clinically significant

The PDLCs for quantitative ECG data will be summarized by parameter and treatment group. The summary will indicate the number and percentage of subjects with PDLCs at any time during the study and will not be presented by visit. Unscheduled visits will be included in this summary. It is possible for subjects to appear in both categories for any parameters.

The incidence rates will be compared descriptively between treatment groups. There will be no statistical testing.

Table 11 indicates the PDLC criteria for ECG data. All ECG data will be presented in listings.

Table 11: Pre-defined Limits of Change Criteria for ECG Values

ECG Parameter	Criteria
PR interval (msec)	< 120 msec or ≥ 210 msec
QRS duration (msec)	≤ 50 msec or > 120 msec
QTcF interval (msec) and QTcB interval (msec)	Absolute post-baseline QTcB, QTcF, and uncorrected QT interval: ≤ 450 msec or > 450 msec ≤ 480 msec or > 480 msec ≤ 500 msec or > 500 msec
	Change from baseline for QTcB, QTcF, and uncorrected QT: ≤ 30 msec or > 30 msec; ≤ 60 msec or > 60 msec

Abbreviations: bpm = beats per minute; ECG = electrocardiogram; msec = millisecond; PDLC = pre-defined limit of change; QTcB = QT interval for corrected heart rate (based on Bazett's formula: $QT/(RR^{1/2})$); QTcF = QT interval for corrected heart rate (based on Fridericia's formula: $QT/(RR^{1/3})$). If the RR is not collected, it will be derived as follows: RR = (60/HR).

9.4. Duration of Therapy and Missing Medication Start and Stop Dates

9.4.1. Missing and Partial Concomitant and Other Medication Start and Stop Dates

Incomplete concomitant medication start dates will be imputed in the same manner as imputation of AE start dates described in Section 9.1.6.1. No imputations will be applied to the medication stop date. If, when using those rules, the stop date is before the start date, then the start date will be left as missing and it will be assumed that the medication is concomitant for the purpose of the analysis. Imputed dates will be flagged in the subject data listings.

9.4.2. Duration of Therapy and Treatment Compliance

Study drug exposure and compliance will be summarized for ATB200 and AT2221 separately. Dose completion (number of administered doses), treatment duration, and treatment compliance will be tabulated. A frequency distribution for the number of treated subjects, and descriptive statistics for the duration of treatment (months) and compliance will be provided by treatment group and for all treated subjects. A summary of treatment duration will also be provided by ERT status.

Duration of treatment (extent of exposure) in months is defined as the number of months on treatment. This will be calculated as: (date of last dose – date of first dose + 1) / 30.4.

The number and percentages of subjects will be tabulated by the extent of exposure categorized into months as: ≤ 3, 3 to ≤ 6, 6 to ≤ 9, 9 to ≤ 12, and > 12.

Note: Due to the potential for delayed visits and catch-up infusions, it is possible that some subjects may stay in the study longer than 12 months.

Compliance for each subject taking ATB200 will be calculated based on the number of infusions as well as the actual infusion dose administered. Compliance based on infusion dose is calculated as:

100 * (total infusion dose administered [mg] / total infusion dose planned or intended [mg]).

Compliance based on number of infusions is calculated as:

100 * (number of infusions administered / number of infusions planned or intended),

where the number of infusions planned is obtained as:

(subject's last date in study while on treatment – date of first infusion + 14) / 14.

Note: Missed infusions due to COVID-19 related policies must be subtracted to ensure accurate compliance.

Compliance for each subject taking AT2221 will be calculated as:

100 * (total dose administered [mg] / scheduled or planned dose [mg]).

Dose compliance rate will be classified as under-compliance: < 80%, within compliance: (80%, 120%), and over-compliance: > 120%.

10. REFERENCES

Agresti, A. (2002). Categorical Data Analysis. Second Edition. New York: Wiley.

Bartlett, J. (2014). "Robustness of linear mixed models." https://thestatsgeek.com/2014/08/17/robustness-of-linear-mixed-models/.

Baschung Pfister, P., E. D. de Bruin, I. Sterkele, B. Maurer, R. A. de Bie and R. H. Knols (2018). "Manual muscle testing and hand-held dynamometry in people with inflammatory myopathy: An intra- and interrater reliability and validity study." PLoS One **13**(3): e0194531.

Bohannon, R. W. and R. Crouch (2017). "Minimal clinically important difference for change in 6-minute walk test distance of adults with pathology: a systematic review." J Eval Clin Pract **23**(2): 377-381.

du Bois, R. M., D. Weycker, C. Albera, W. Z. Bradford, U. Costabel, A. Kartashov et. al. (2011). "Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference." Am J Respir Crit Care Med **184**(12): 1382-1389.

Enright, P. L. and D. L. Sherrill (1998). "Reference equations for the six-minute walk in healthy adults." Am J Respir Crit Care Med **158**(5 Pt 1): 1384-1387.

Evans, J. A. and W. A. Whitelaw (2009). "The assessment of maximal respiratory mouth pressures in adults." Respiratory Care **54**(10): 1348-1359.

Haywood K. L., A. M. Garratt, L. J. Schmidt, A. E. Mackintosh and R. Fitzpatrick (2004). Health Status and Quality of Life in Older People: A Structured Review of Patient-reported Health Instruments Report from the Patient-reported Health Instruments Group (formerly the Patient-assessed Health Outcomes Programme) to the Department of Health.

Koch G. G., C. M. Tangen, J. W. Jung and I. A. Amara (1998). "Issues for covariance analysis of dichotomous and ordered categorical data from randomized clinical trials and non-parametric strategies for addressing them." Stat Med **17**(15-16): 1863-92.

LaVange L. M., T. A. Durham and G. G. Koch (2005). "Randomization-based nonparametric methods for the analysis of multicentre trials." Stat Methods in Med Res **14**(3): 281-301.

nQuery Advisor Sample Size Calculator, Version 8.2.1.0 (2018). Statistical Solutions Ltd., One International Place, 100 Oliver Street, Suite 1400, Boston, MA 02110.

Rpatel. "Combining Z Scores by Weighted Average. Sanity Check Please?" URL (version: 2019-04-27): https://stats.stackexchange.com/q/348605.

Schrover, R, K. Evans, R. Giugliani, I. Noble and K. Bhattacharya (2017). "Minimal clinically important difference for the 6-min walk test: literature review and application to Morquio A syndrome." Orphanet Journal of Rare Diseases 12: 78.

Siddiqui, O., J. H. M. Hung, and R. O'Neill (2009). "MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 NDA datasets." Journal of Biopharmaceutical Statistics **19**(2): 227-246.

Uldry, C. and J-W. Fitting (1995). "Maximal values of sniff nasal inspiratory pressure in healthy subjects." Thorax **50**: 371-375.

van der Ploeg, A. T., P. R. Clemens, R. J. Hopkin, K. Kacena, B.-J. Sanson and P. Lafôret (2017). Poster: Long-term Efficacy of Alglucosidase Alfa in Late-onset Pompe Disease. WORLD. San Diego, CA, Sanofi Genzyme.

White, I. R., J. Carpenter and N. J. Horton (2012). "Including all individuals is not enough: lessons for intention-to-treat analysis." Clin Trials **9**(4): 396-407.

Zink, R. C. and G. G. Koch (2012). "A SAS/IML macro for nonparametric randomization-based analysis of covariance." Journal of Statistical Software **50**(3).

APPENDIX 1. SAMPLE SIZE JUSTIFICATION

The data used for the sample size calculations came from 2 sources:

a. Data for alglucosidase alfa came from the LOTS extension (NCT00455195) which is a long-term extension study in which subjects were treated with the standard of care ERT (alglucosidase alfa) for up to 9 years; and
b. Data for ATB200/AT2221 came from Amicus' ongoing single-arm study, ATB200-02, which enrolled 4 cohorts of subjects. Of these, 2 cohorts had relevant data for use in the sample size calculations: Cohort 1 – ambulatory subjects who had been on ERT treatment for 2 to < 7 years; and Cohort 3 – ambulatory subjects who had not been previously treated with ERT (ie, ERT-naïve).

Firstly, effect sizes were estimated using % predicted 6MWD data from ATB200-02 Cohort 1 subjects (n = 10) treated with alglucosidase alfa for 2 to 6 years before switching to ATB200/AT2221, ATB200-02 Cohort 3 subjects (n = 5) who were ERT-naïve before starting ATB200/AT2221, and from a LOTS extension study (n = 22) where subjects were treated with alglucosidase alfa for up to 9 years (van der Ploeg, Clemens et al. 2017). Longitudinal percent predicted 6MWD data were extracted from Figure 5 of van der Ploeg, Clemens et al. 2017 and subjects were 1:1 matched to ATB200-02 subjects on the basis of treatment duration and baseline % predicted 6MWD. A treatment duration window was specified for each ATB200-02 subject extending from the time of switching from alglucosidase alfa to ATB200/AT2221 (eg, 4.4 years) to the end of the 9 or 12-month treatment period with ATB200/AT2221 (eg, 5.4 years). In order to be eligible for matching to a given ATB200-02 subject, a LOTS extension study subject had to have at least one 6MWD value within the corresponding treatment duration window and one 6MWD value before and after the window. The LOTS extension subjects were then sequentially matched to ATB200-02 subjects based on having the smallest difference in baseline % predicted 6MWD at the corresponding treatment duration starting point. Once the LOTS extension subject with the closest baseline 6MWD was matched to a given ATB200-02 subject, that subject was removed from the pool and the remaining subjects were matched. Calculation of the slope of change in % predicted 6MWD for each matched LOTS extension subject was then based upon cubic spline values extracted every 3 months across the corresponding 9 or 12-month treatment duration window ± 6 months on each side.

Based on this analysis, ATB200-02 Cohort 1 subjects switching from alglucosidase alfa to ATB200/AT2221 had a mean 12-month slope of change of +5.87% predicted 6MWD compared to a mean 12-month slope of -2.15% for matched subjects in the LOTS extension remaining on alglucosidase alfa. The mean difference between groups was +8.01%, favoring ATB2200/AT2221 with a standard deviation of +8.49%. For the ATB200-02 Cohort 3 subjects who were ERT-naïve before starting on ATB200/AT2221, the mean 12-month slope of change was +12.61% predicted 6MWD compared to a mean 12-month slope of -0.65% for matched subjects in the LOTS extension remaining on alglucosidase alfa. The mean difference between groups was +13.26%, favoring ATB2200/AT2221 with a standard deviation of +2.85%. Thus, the effect size of ATB200/AT2221 in ERT-naïve subjects was larger than the effect size in ERT-switch subjects.

Secondly, when the subjects in ATB200-02 Cohort 1 (ERT-switch subjects) were combined with Cohort 3 (ERT-naïve subjects) into a single LOPD sample (n = 15), the overall mean 12-month slope of change for subjects receiving ATB200/AT2221 in ATB200-02 was +8.11% predicted

6MWD compared to a mean 12-month slope of -1.65% for the matched subjects in the LOTS extension on alglucosidase alfa. The mean difference between groups was +9.76% favoring ATB2200/AT2221 with a standard deviation of +7.43%. The resulting estimated standardized effect size (mean difference divided by standard deviation of difference) is then 1.31 (95% CI: 0.60, 2.00) for subjects switching from alglucosidase alfa to ATB200/AT2221.

Thirdly, the sample size calculation used a conservative standardized effect size of 0.7 for the inter-group analysis in the combined ERT-switch and ERT-naïve population. This reduction from the estimated 1.31 effect size was intended to help account for uncertainty of the effect size estimate due to the relatively small sample size, potential selection bias in ATB200-02 versus LOTS study extension data, other potential biases in period effects in combining ERT-switch and ERT-naïve subjects, as well as imbalance in known and unknown risk factors, outliers, informative missing data, and dropout patterns. The standard effect size of 0.7, which is close to the lower confidence limit, constitutes about one-half of the originally calculated standardized effect size of 1.31, thus making it quite conservative. Of note, the standard effect size of 0.7 corresponds to a % predicted 6MWD of 5.0. In the range of expected subject age, weight, and height (using Enright and Sherrill 1998 for % predicted 6MWD calculation), a % predicted value of 5.0 equates to a measured 6MWD value of between 20 to 40 meters. Such a 6MWD range is within the minimal clinically important difference in 6MWD for adults with pathology, and it may also be clinically important across multiple subject groups (Bohannon and Crouch 2017). In addition, 20 to 40 meters is within the range of group-level 6MWD differences (investigational treatment versus comparator) that served as the basis for regulatory approval.

APPENDIX 2. DESCRIPTIONS OF EFFICACY ENDPOINTS

1. 6MWD: The 6-minute walk distance, measured in meters (m), is the distance walked on the 6MWT.
2. Percent predicted 6MWD is calculated as:

   % predicted 6MWD = (actual 6MWD / predicted 6MWD) * 100,

   (rounded up or down based on standard rounding rules), where the predicted values are derived for males and females using the following Enright and Sherill 1998 equations:

   Males: 6MWD = (7.57 x Height in cm) - (5.02 x Age in years) - (1.76 x Weight in kg) - 309 m;

   Females: 6MWD = (2.11 x Height in cm) - (5.78 x Age in years) - (2.29 x Weight in kg) + 667 m.
3. Pulmonary function tests (PFTs):
   - Sitting forced vital capacity (FVC): FVC (L) and percent predicted FVC (%)
   - Supine forced vital capacity: FVC (L) and percent predicted FVC (%)
   - Maximum inspiratory pressure (MIP): MIP (cm of water [cm H_2O]) and % predicted MIP (%)
   - Maximum expiratory pressure (MEP): MEP (cm of water [cm H_2O]) and % predicted MEP (%)
   - SNIP (cm of water [cm H_2O]) and % predicted SNIP (%)

   The percent predicted values of MIP, MEP, and SNIP will be calculated as:

   % predicted = (actual result / predicted result) * 100,

   where the predicted results are obtained using the reference equations in Table 12. The equations for MIP and MEP are from Uldry and Fitting 1995; the equation for SNIP is from Evans and Whitelaw 2009.

Table 12: Predicted Equations for MIP, MEP, and SNIP

PFT	Male	Female
MIP	120 – (0.41 x Age in years)	108 – (0.61 x Age in years)
MEP	174 – (0.83 x Age in years)	131 – (0.86 x Age in years)
SNIP	126.8 – (0.42 x Age in years)	94.9 – (0.22 x Age in years)

4. Manual muscle testing (MMT):
   - Right/left shoulder adduction
   - Right/left shoulder abduction
   - Right/left elbow flexion

- Right/left elbow extension
- Right/left hip flexion
- Right/left hip abduction
- Right/left knee flexion
- Right/left knee extension

Each manual muscle test is evaluated on a scoring scale from 0 to 5, as follows: 0 = no muscle movement; 1 = visible muscle movement, but no movement at the joint; 2 = movement at the joint, but not against gravity; 3 = movement against gravity, but not against added resistance; 4 = movement against resistance, but less than normal; 5 = normal strength.

- MMT Lower Extremity score: The total score for MMT lower extremity strength is obtained by summing the test scores across the following 8 body parts: right/left hip flexion, right/left hip abduction, right/left knee flexion, and right/left knee extension. The total score ranges from 0 to 40, with lower scores indicating lower muscle strength. That is, Lower = Hip scores + Knee scores.
- MMT Upper Extremity score: The total score for MMT upper extremity strength is obtained by summing the test scores across the following 8 body parts: right/left shoulder abduction, right/left shoulder adduction, right/left elbow extension, and right/left elbow flexion. The total score ranges from 0 to 40, with lower scores indicating lower muscle strength. That is, Upper = Shoulder scores + Elbow scores.
- MMT Total score: The MMT total score is the sum of the lower extremity score and the upper extremity score. The total score ranges from 0 to 80, with lower scores indicating lower overall muscle strength. That is, Total = Upper extremity + Lower extremity.
- Proximal body muscle groups include right/left hip flexion, right/left hip abduction, right/left shoulder abduction, and right/left shoulder adduction with the total score ranging from 0 to 40, based on these 8 muscle groups. That is, Proximal = Shoulder scores + Hip scores.

5. Quantitative muscle test (QMT) was measured using the hand-held dynamometer (HHD). Larger values/scores (in kg) indicate greater muscle strength.
    - Right/left shoulder adduction
    - Right/left shoulder abduction
    - Right/left elbow flexion
    - Right/left elbow extension
    - Right/left hip flexion
    - Right/left hip adduction
    - Right/left hip abduction

  - Right/left knee flexion
  - Right/left knee extension

- QMT value for upper extremities (kg) is obtained by summing up the values/scores from the following: right/left shoulder abduction, right/left shoulder adduction, right/left elbow extension, and right/left elbow flexion.
- QMT value for lower extremities (kg) is obtained by summing up the values/scores from the following: right/left hip flexion, right/left hip abduction, right/left hip adduction, right/left knee flexion, and right/left knee extension.
- QMT total value (kg) is obtained by adding up the QMT upper extremity and QMT lower extremity values/scores.
- QMT proximal body value (kg) is obtained by adding up the scores/values from the following: right/left shoulder abduction, right/left shoulder adduction, right/left hip flexion, right/left hip abduction, and right/left hip adduction.

6. PROMIS Short Forms

- PROMIS – Physical Function Short Form 20a (v2.0) consists of 20 questions. The first 14 questions are each scored on a scale from 1 to 5 as follows: 1 = unable to do; 2 = with much difficulty; 3 = with some difficulty; 4 = with a little difficulty; 5 = without any difficulty; the next 6 questions are each scored on a scale from 1 to 5 as follows: 1 = cannot do; 2 = quite a lot; 3 = somewhat; 4 = very little; 5 = not at all.
- PROMIS – Fatigue Short Form 8a consists of 8 questions, each scored on a scale from 1 to 5 as follows: 1 = not at all; 2 = a little bit; 3 = somewhat; 4 = quite a bit; 5 = very much; and 2 questions, each scored on a scale from 1 to 5 as follows: 1 = never; 2 = rarely; 3 = sometimes; 4 = often; 5 = always.
- PROMIS – Upper Extremity Short Form 7a consists of 7 items each scored on a decreasing scale from 1 to 5 as follows: 1 = unable to do; 2 = with much difficulty; 3 = with some difficulty; 4 = with a little difficulty; 5 = without any difficulty.
- PROMIS – Dyspnea Severity Short Form 10a consists of 10 questions each scored on a scale from 0 to 3 as follows: 0 = no shortness of breath; 1 = mildly short of breath; 2 = moderately short of breath; 3 = severely short of breath.

7. Gait, Stairs, Gowers' maneuver, Chair (GSGC total score is the sum of the component scores from the following 4 functional tests:

- 10-meter walk test (10MWT) is the time in seconds (s) it takes for the subject to walk 10 meters. The test is scored as: 1 = normal; 2 = mild waddling, lordosis and/or toe walking; 3 = moderate waddling, lordosis and/or toe walking; 4 = severe waddling, lordosis and/or toe walking; 5 = walks only with assistance (ie, braces, cane, crutches); 6 = stands, but unable to walk; 7 = confined to wheelchair. The score from this test is used as the Gait score in GSGC.

- Stairs score (based on the subject climbing 4 stairs). This is scored as: 1 = climbs four stairs without assistance; 2 = supports one hand on thigh; 3 = supports both hands on thighs; 4 = climb stairs in upright position but with aid of railing; 5 = climbs while clinging to the railing with both hands; 6 = manages to climb only a few steps; 7 = unable to climb steps.
- Gowers' maneuver score (based on the subject lying down on the floor, then rising from the floor to get to a standing position). This is scored as: 1 = normal; 2 = butt first maneuver, one hand on floor; 3 = butt first maneuver, two hands on floor; 4 = unilateral hand support on thigh; 5 = bilateral hand support on thighs; 6 = arises only with aid of an object (table, chair, cane, etc); 7 = unable to rise.
- Chair score (based on the subject arising from a sitting position in a chair to a standing position). This is scored as: 1 = normal; 2 = with wide base and/or difficulty but without support; 3 = with support on one thigh; 4 = with support on both thighs; 5 = with support on arms of chair or on a table; 6 = not possible.

For each of the above motor function tests, the actual time (in seconds [s]) that the subject takes to perform the test is also recorded for analysis.

The total GSGC score ranges from 4 (normal performance) to 27 (worst performance).

8. Timed Up and Go [TUG] test represents the time it takes the subject to stand up from the chair, walk to the line on the floor (about 3 meters), turn around, walk back to the chair, and sit down, all at the regular pace.
9. R-PAct Scale total score: The Rasch-built Pompe-specific Activity (R-PAct) scale consists of 18 questions, each scored on a scale from 0 to 2 as follows: 0 = no; 1 = yes, but with difficulty; 2 = yes, without difficulty.

   The total score is calculated simply by summing up the observed scores across the 18 items and it ranges from 0 to 36, with higher values representing lower level of disease impact on the muscles.
10. EQ-5D-5L: European Quality of Life-5 Dimensions 5 Response Levels health status:

    The EQ-5D-5L consists of the EQ-5D descriptive system and the EQ-VAS (visual analogue scale). Each of the 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) has 5 categorical responses/levels of perceived problems coded as follows:

    Level 1 = indicating no problem; Level 2 = indicating slight problems;
    Level 3 = indicating moderate problems; Level 4 = indicating severe problems;
    Level 5 = indicating extreme problems (for pain and anxiety) or indicating unable to (for mobility, self-care, and activity).

    The EQ-VAS is a quantitative measure of health outcome that reflects the patient's own judgement. The EQ-5D is a 5-digit number (a categorical variable) that describes the subject's health state.

11. SGIC: The Subject's Global Impression of Change consists of the following 8 items (or functional areas):

    - overall physical wellbeing
    - effort of breathing
    - muscle strength
    - muscle function
    - ability to move around
    - activities of daily living
    - energy level
    - level of muscular pain

    Each of the 8 functional items is scored on a 7-point rating scale that ranges from '1 = very much worse' to '7 = very much improved.'

    For each functional item, a tertiary response variable (improving, declining, stable) will be defined as: "Improving" which consists of improved, moderately improved, and very much improved; "Declining" consists of worse, moderately worse, and very much worse; and "Stable" equals no change.

12. PGIC: The Physician's Global Impression of Change is based on a single item that is scored on a 7-point rating scale. The response options/category are similar to the SGIC described above.

APPENDIX 3. SAMPLE SAS PROGRAMS

All analysis programs will be submitted. Three sample SAS codes are provided below to illustrate the MMRM analysis (based on ITT-OBS), ANCOVA (based on ITT-LOCF) and the composite endpoint analysis that combines two separate logistic regression analyses.

Sample Code for the MMRM Analysis

```
/*************************************************************************/
/** Description: Sample program for implementing the MMRM Analysis.     */
/** This sample program applies to the primary analysis of the primary  */
/** endpoint(6MWD), and can be adapted to all other MMRM analyses       */
/** specified in the SAP.                                               */
/*                                                                      */
/*  Response variable(CHG): Change from baseline at each visit timepoint*/
/** Explanatory variables in the model are:                            */
/**       TRTPN AVISIT TRTPN*AVISIT SEX BASE ERTEXP AGEBL HGTBL WGTBL   */
/*                                                                      */
/*  Assume that the 6MWD dataset is called ANADAT                      */
/*                                                                      */
/*  Some Definitions:                                                  */
/**     TRTPN  = Treatment group (1 = new treatment, 2 = control)      */
/**     AVISIT = Analysis Visit                                        */
/**     ERTSTA = ERT status (Experienced or Naïve)                     */
/*************************************************************************/

/*************************************************************************/
/* Perform the analysis. Then obtain treatment means and differences at */
/* each visit, with p-values and CIs.                                   */
/*************************************************************************/

ods output LSMeans=lsm LSMEstimate=lsmest;
proc mixed data = anadat method = reml;
   class TRTPN AVISIT ERTSTA SEX USUBJID;
   model CHG = TRTPN AVISIT TRTPN*AVISIT SEX BASE ERTSTA AGEBL HGTBL WGTBL /
               ddfm=kr;
   repeated AVISIT / type=UN subject=USUBJID;
   lsmeans TRTPN*AVISIT / slice=AVISIT alpha=0.05 OM diff cl s;  *** Obtain
      treatment means and differences at each visit, with CIs.**;

   lsmestimate TRTPN*AVISIT "Treat vs. Control at Week 52" 0 0 0 1 0 0 0 -1
      / upper testvalue=0; *** Get treatment diff and 1-sided p-value  **;

   lsmestimate TRTPN*AVISIT 'Treat vs. Control at Week 52' 0 0 0 1 0 0 0 -1 /
      testvalue=0 alpha=0.05 cl; **Get treatment diff and 2-sided p-value *;
```

Sample Code for the ANCOVA

```
/*************************************************************************/
/* ANCOVA is based on the ITT-LOCF population. This sample program can  */
/* be adapted to other ANCOVA analyses for the 6MWD, percent predicted  */
/* FVC, and all others specified in the SAP.                            */
/*************************************************************************/

/* Perform the ANCOVA, obtain estimates of the treatment means, treatment */
/* differences, p-values and 95% CIs.                                     */

ods output LSMeans=lsm LSMEstimate=lsmest;
proc mixed data = anadat;
   class TRTPN ERTSTA SEX;
   model CHG = TRTPN SEX BASE ERTSTA AGEBL HGTBL WGTBL /ddfm=kr;
   lsmeans TRTPN / OM s alpha=0.05 cl diff;  ** Obtain lsmeans and lsmean
differences;

   lsmestimate TRTPN "Treat vs. Control at Week 52" 1 -1 / upper
       testvalue=0; **Get treatment diff and 1-sided p-value at W52 *;

   lsmestimate TRTPN 'Treat vs. Control at Week 52' 1 -1 / testvalue=0
       alpha=0.05 cl; **Get treatment diff, 2-sided p-value, 95% CIs at W52*;
```

Sample Code for the Analysis of Composite Endpoint

```
*************************************************************************
* Implement separate logistic analyses for positive responses and      *
* negative responses, and then combine the two results using weights.  *
*                                                                       *
*  Assume input dataset (COMPOS) is structured as below, and has the   *
*  following variables: RESPONSE, TRTPN, ERTSTA, DUR, SEX, AGEBL       *
*                                                                       *
   1 1 N 2.4 F 0
   3 1 N 3.5 M 0
   1 1 E 2.7 M 37
   0 2 N 5.6 F 0
   3 1 E 7.0 M 67
   3 2 N 2.0 F 0
   2 1 E 4.1 M 64
   1 2 N 4.8 M 0
   0 2 E 7.7 F 28
   3 1 E 5.5 F 32
   2 2 E 6.3 F 48
   . . .  .  . .

*   RESPONSE variable = composite response with values 0, 1, 2, 3     *
*   TRTPN = Treatment group (1 = new treatment, 2 = control)          *
*   ERTSTA = ERT status (Experienced or Naïve)                        *
*   DUR = duration of ERT: value for ERT-experienced, and 0 for naïve.*
*************************************************************************;

** Analysis of positive responses vs. not positive responses **;
```

```
data comp1;
  set compos;
  if response=3 then resp1='Y';
  else resp1='N';
run;

** Analysis of stable responses vs. negative responses;
data comp2;
  set compos;
  if response=3 then delete;
  if response=2 then resp2='Y';
  else resp2='N';
run;

** Run two separate logistic regression models **;
ods output ParameterEstimates=test1;
proc genmod data=comp1;
  class TRTPN ERTSTA SEX;
  model resp1(event='Y') = TRTPN ERTSTA DUR SEX AGEBL / dist=binomial
                           link=logit type3;
run;

ods output ParameterEstimates=test2;
proc genmod data=comp2;
  class trt ert;
  model resp2(event='Y') = TRTPN ERTSTA DUR SEX AGEBL / dist=binomial
                           link=logit type3;
run;

data combo;
  merge test1(where=(level1='1') keep=level1 estimate stderr
              rename=(estimate=estimate1 stderr=stderr1))
        test2(where=(level1='1') keep=level1 estimate stderr
              rename=(estimate=estimate2 stderr=stderr2));
run;

** Calculate the weights, combine the test statistics and obtain the **;
** p-value using weight as the inverse of the standard error estimate *;
data combo2;
  set combo;
  w1=1/stderr1**2;
  w2=1/stderr2**2;
  z1=estimate1/stderr1;
  z2=estimate2/stderr2;
  zw=(w1/(w1+w2)*z1 + w2/(w1+w2)*z2)/sqrt((w1/(w1+w2))**2+(w2/(w1+w2))**2);
  pvalue=1-cdf('normal',zw);
run;
```