NCT #NCT03648827

STATISTICAL ANALYSIS PLAN

PHASE 2 CLINICAL PHARMACOLOGY STUDY TO ASSESS DYSTROPHIN LEVELS IN SUBJECTS WITH NMDMD BEFORE AND AFTER TREATMENT WITH ATALUREN

PTC124-GD-045-DMD

11 JANUARY 2021

VERSION 1.0

PTC THERAPEUTICS, INC.
100 CORPORATE COURT
SOUTH PLAINFIELD, NJ 07080 USA

PTC THERAPEUTICS APPROVAL SIGNATURES

[redacted]

Date

PTC Therapeutics, Inc.

[redacted]

Date

PTC Therapeutics, Inc.

[redacted]

Date

PTC Therapeutics, Inc.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND DEFINITIONS OF TERMS

Abbreviation or Specialized Term	Explanation
AE	Adverse event
CK	Creatine kinase
COVID-19	Coronavirus disease of 2019
CRF	Case report form
CV	Coefficients of variability
ECL	Electrochemiluminescence
eCRF	Electronic case report form
IHC	Immunohistochemistry
ITT	Intent-to-treat
LLOQ	Lower limit of quantification
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed effect model repeated measurement
N	Number of subjects
nmDMD	Nonsense mutation Duchenne muscular dystrophy
PK	Pharmacokinetics
PP	Per protocol
% normal	Percentage of the predicted levels
PT	Preferred term
rNSAA	Revised North Star Ambulatory Assessment
SAE	Serious adverse event
SAP	Statistical Analysis Plan
SOC	System organ class
TFT	Timed function test
TEAE	Treatment-emergent adverse event
[redacted]	[redacted]

1. INTRODUCTION

The purpose of this Statistical Analysis Plan (SAP) is to describe the procedures and the statistical methods that will be used in analyzing and reporting results for study PTC124-GD-045-DMD (hereafter referred to as Study 045). This SAP is prepared based on the study protocol V3.0 (dated 06May2019). The reader is referred to the study protocol, the electronic case report form (eCRF), general eCRF completion guidelines, and various data collection instruments employed in the study for details of study design, conduct and data collection.

2. STUDY DESIGN

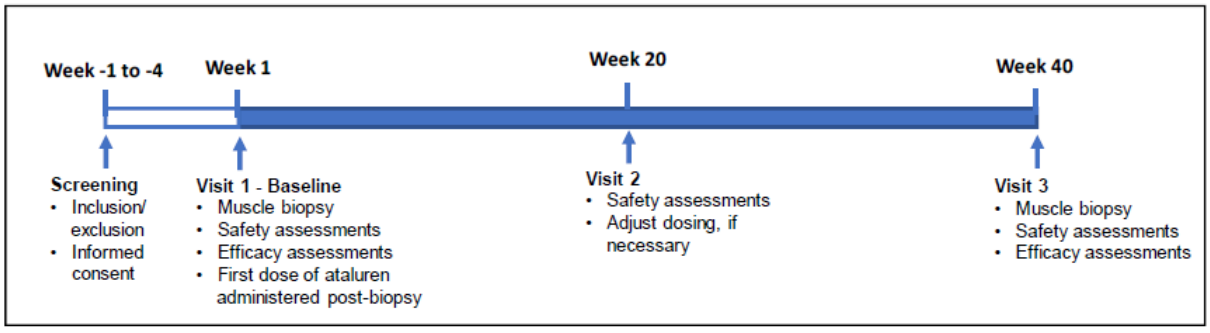
This is an open-label, single-arm, Phase 2 study designed to evaluate the ability of ataluren treatment to increase dystrophin protein levels in muscle cells of patients with nonsense mutation Duchenne muscular dystrophy (nmDMD). The study will evaluate the levels of dystrophin before and after 40 weeks of ataluren therapy using muscle biopsies and two validated assay methods.

Approximately, 15 to 20 ataluren-naïve male subjects, aged >=2 and <8 years of age will be administered ataluren.

Ataluren will be dosed daily 10 mg/kg in the morning, 10 mg/kg at midday, and 20 mg/kg in the evening (10, 10, 20 mg/kg).

The study will have a screening visit and 3 additional visits (Figure 1). At the screening visit, which will be performed at a patient's local site, inclusion/exclusion criteria, demographics and medical history will be assessed. In addition, blood will be drawn for genotyping to confirm that a subject carries a nonsense mutation in the dystrophin gene. For all subjects, Visit 1 and Visit 3 will occur at the same site, that of [redacted]. During Visit 1 and Visit 3 biopsies will be performed and used for assessing dystrophin protein levels. At Visit 1, the first dose of ataluren will be administered following the biopsy at [redacted] and subjects, parent/caregivers or legal guardian will be dispensed a supply of ataluren which will be administered at home. Visit 2 will be in the clinic at the subjects' local site and weight will be measured to determine if the dose of ataluren requires adjustment. Adverse events (AE), serious adverse events (SAE), concomitant medications, and compliance will also be captured, and an additional supply of ataluren will be dispensed.

Figure 1: Study Schematic



No interim analysis will be conducted for this study.

3. STUDY OBJECTIVES AND ENDPOINTS

3.1. Study Objectives

3.1.1. Primary Objective

To assess the change in levels of dystrophin in ambulatory nmDMD subjects after treatment with ataluren for 40 weeks using quantitative assay, such as electrochemiluminescence (ECL).

3.1.2. Secondary Objective

To assess dystrophin levels/intensity and protein localization by immunohistochemistry (IHC).

3.1.3. Exploratory Objectives

The exploratory objectives are as follows:

- To assess the change from baseline in revised North Star Ambulatory Assessment (rNSAA) and timed function tests (TFTs) after 40 weeks of ataluren treatment
- To assess changes from baseline in creatine kinase (CK) levels after 40 weeks for ataluren treatment
- To evaluate the steady state pharmacokinetics (PK) of ataluren after 40 weeks of treatment

3.2. Study Endpoints

3.2.1. Primary Endpoint

The percent change from baseline in dystrophin levels (final concentration) following 40 weeks of ataluren therapy, as measured by an ECL-based immunoassay that has been validated and will be run under Good Laboratory Practice conditions.

3.2.2. Secondary Endpoint

The percent change in dystrophin levels (mean membrane stain density) from baseline after 40 weeks of ataluren therapy as determined by a validated IHC assay.

3.2.3. Exploratory Endpoints

The exploratory endpoints are as follows:

- Change from baseline in rNSAA after 40 weeks of treatment
- Change from baseline in TFTs after 40 weeks of treatment
- Change from baseline in CK levels after 40 weeks of treatment
- PK assessments at Week 1 and Week 40

3.3. Sample Size Determination

Approximately, 15 to 20 ataluren-naïve male subjects, aged >=2 and <8 years of age is planned to be enrolled in this study. Limited information is available about the magnitude and intra-subject variability of the change in dystrophin levels, as measured by an immunoassay such as ECL following 40 weeks ataluren therapy in this patient population. Assuming a baseline dystrophin level of 0.2% of control levels (as per previous dystrophin replacement therapy [Center for Drug Evaluation and research Application number: 206488Orig1s000 Eteplirsen Summary Review]), Table 1 shows the number of subjects needed to achieve an 85% power to detect an increase from baseline in dystrophin of 0.04%, 0.06% and 0.08%, for intra-subject coefficients of variability ranging from 20% to 35%, using a one-sided test at alpha 5%. As an example, if the increase from baseline is 0.08% and the intra-subject coefficient is 30%, 13 evaluable subjects would be required to have a power of 85% to detect this increase.

Table 1: Summary of number of subjects required to achieve 85% power to detect increase in dystrophin levels from baseline

Increase of dystrophin from baseline	Intra Subject CV			
	20%	25%	30%	35%
0.04%	19	29	41	55
0.06%	10	15	21	28
0.08%	7	10	13	18

Abbreviations: CV, coefficients of variability

4. STUDY POPULATIONS

4.1. Screened Population

Screened population is defined as all patients who have signed informed consent.

4.2. Enrolled Population

Enrolled population is defined as the subset of screen population who are not screen failures.

4.3. Intent-to-Treat Population

This analysis set will include all enrolled subjects with treatment assignments. In addition, subjects in this analysis set must have a valid assessment of dystrophin levels at baseline, as measured by ECL. This analysis set will be used for the primary endpoint analyses and summaries of other efficacy endpoints.

4.4. Per Protocol Population

The per protocol (PP) analysis set consists of all ITT subjects without major protocol deviation that will impact the efficacy analyses. Any ITT subjects who meet the following criteria will be excluded from the PP population:

- study treatment not received
- Week 40 visit not within 38 to 42 weeks from the Baseline visit
- not compliant to study drug dose regiment; treatment compliance <80% or >120%
- baseline dystrophin levels [redacted] predicted
- do not have a valid assessment of dystrophin levels at either Baseline or Week 40 visits

This analysis set will be used for supportive efficacy analyses.

4.5. Safety Population

This analysis set will include all subjects who received at least one dose of ataluren and will be used for all safety summaries.

5. GENERAL CONSIDERATIONS

5.1. Tables, Figures, and Listings

In general, continuous variables will be summarized using descriptive statistics including n (number of subjects), mean, standard deviation, median, minimum, and maximum. Categorical variables will be summarized using frequency count as number and percentage of subjects. Unless specified, all hypothesis testings are 1-sided at a 5% alpha level.

By subject data listings will be created for selected electronic case report form (CRF) domains sorted by subject and associated dates.

5.2. Visit Window

Study day 1 (baseline) is defined as the date of the first dose of study drug during the treatment period. The nominal visit (ie, study visit captured on the CRF) will be used as the analysis visits in all efficacy analyses. For safety analyses, the unscheduled and Early Termination visits will be reclassified into analysis visits according to Table 2. If multiple visits occurred within the same visit window, the one closest to the target visit date will be used in the analysis. If two visits are equidistant to the target visit date, the latest will be used.

Table 2: Visit Window for Safety Analyses

Timepoint	Target Visit Date	Visit window (in study days)
Visit 1 (baseline)	1	-45 to 1
Visit 2 (week 20)	140	71 to 210
Visit 3 (week 40)	280	>=211

5.3. Missing Data Handling

Unless specified, observed data will be used for all analyses. Missing data will not be imputed.

5.4. Interim Analysis

No interim analysis will be conducted for this study.

5.5. Multicenter, Multiplicity control

This is a multicenter study with all muscle biopsies performed at one center. No multicenter analysis will be conducted for this study.

There is only one primary efficacy endpoint, and one secondary efficacy endpoint for this study. The secondary efficacy endpoint will be analyzed if the primary efficacy endpoint is statistically significant at 0.05 significance level. All other endpoints and analyses are supportive. Therefore, no multiplicity adjustment will be conducted for this study.

5.6. COVID-19 Related Considerations

Due to the Coronavirus Disease of 2019 (COVID-19) pandemic, 8 subjects will have Visit 3 (Week 40 visit) later than planned due to closure of the study site to study procedures. These subjects will receive continued therapy with ataluren until final study biopsies can be obtained. Since many subjects will have their Week 40 biopsy performed much later than after 40 weeks of ataluren treatment therapy, a sensitivity analysis will be conducted to include the Week 40 biopsy study day as a covariate in the analysis model. This sensitivity analysis will be performed for both the primary and secondary efficacy endpoints as detailed in Section 7.1.3 (Sensitivity Analyses) for the primary efficacy endpoint and in Section 7.2.2 (Sensitivity Analyses) for the secondary efficacy endpoint.

6. SUBJECT DATA

6.1. Subject Disposition and Study Populations

The number of subjects in each analysis population will be summarized using frequency count.

The number of subjects who completed or discontinued from the study and the reasons of discontinuation will be summarized based on both the ITT and the safety populations if they are different.

6.2. Treatment Exposure

The extent of exposure (duration of exposure) to study drug will be summarized using descriptive statistics and tabulated with frequency count by categories. Duration of exposure of study drug will be defined as the number of days between the date the subject received the first dose of study drug during treatment period and the date the subject received the last dose of study drug during treatment period, inclusive. Duration of exposure will be categories as <266 days, 266 to 295 days and >295 days. Treatment exposure will be summarized based on both the ITT and safety populations if they are different.

6.3. Study Drug Compliance

Treatment compliance (in %) is defined as follows:

$$\frac{100 \times (\text{total number of sachets dispensed} - \text{total number of sachets returned or lost})}{\text{number of sachets expected to be taken}}$$

Treatment compliance will be summarized using descriptive statistics based on the safety population. Treatment compliance will also be summarized using the categories <80%, ≥80% to ≤100%, >100% to ≤120%, and >120%.

6.4. Demographic and Baseline Characteristics

Unless otherwise specified, baseline measurement is the last observed measurement prior to the first dose of study drug for a given assessment.

Continuous demographic and baseline variables including age, age at onset of phenotypic evidence, age at first observation of difficulty with ambulation, height, weight, and body mass index will be summarized using descriptive statistics. Categorical variables including ethnicity group, race, and type of corticosteroid use will be tabulated using frequency count. Demographic and baseline characteristics will be summarized based on both the ITT and safety populations if they are different.

Genotype information including the location of nonsense point mutation (exon number) and the stop codon type will be tabulated using frequency count and displayed in subject listings.

6.5. Concomitant Medications and Non-Drug Treatments

Medications and non-drug treatments will be coded using the World Health Organization Drug Dictionary, version September 1, 2020. Prior medications and prior non-drug treatments are defined as those taken any time prior to first dose of study drug. Concomitant medications and non-drug treatments are defined as those taken any time on or after first dose of study drug.

The number and percentage of subjects who take prior and concomitant medications will be summarized using the safety population by Anatomical Therapeutic Chemical class, and WHO DD preferred term (PT) for safety populations. If a subject takes the same medications for the same class level or drug name, the subject will be counted only once for that class level or drug name.

6.6. Medical History

Medical history other than the history of dystrophinopathy will be coded using the Medical Dictionary for Regulatory Affairs (MedDRA), version 23.1. The number and percent of subjects with medical history will be summarized by System Organ Class (SOC) and PT based on safety population.

7. EFFICACY EVALUATION

The primary efficacy assessment for this study is the dystrophin levels obtained from muscle biopsies. Muscle biopsies of [redacted] muscles will be performed using an established core biopsy procedure at baseline and after 40 weeks of ataluren therapy. If the [redacted] muscle is considered by the investigator to be too small for a biopsy sample (or the obtained sample is considered not evaluable for analysis), the [redacted] muscle may be used. [redacted] cores of muscles will be taken from each location and multiple samples will be prepared from each muscle cores. Dystrophin levels will be measured from each sample by both ECL and IHC. The dystrophin levels measured by ECL will be considered as the primary efficacy endpoint. The dystrophin levels measured by IHC will be considered as the secondary efficacy endpoint.

7.1. Primary Efficacy Analyses

7.1.1. Primary Analysis

The primary efficacy endpoint is the percent change from baseline in dystrophin levels (final concentration) following 40 weeks of ataluren therapy, as measured by ECL. The final concentration is calculated by dividing a sample's dystrophin concentration by its total protein concentration. In addition, the percentage of the predicted levels (% normal) will also be calculated as dystrophin levels normalized to dystrophin concentration of the control population (Syneos Health Report No. 14597.051619.A):

% normal (in %) = 100 final concentration/33.8*

Within each muscle location, the best, median, and the average of the dystrophin levels and the corresponding % normal values of the muscle cores will be summarized using descriptive statistics for each subject. In case any measured dystrophin concentrations are below the lower limit of quantification (LLOQ), the dystrophin concentrations will be set at 0.5*LLOQ in the final concentration calculation for the statistical analysis.

The null hypothesis of the primary analysis is that the percent change from baseline in dystrophin levels following 40 weeks of ataluren therapy as measured by ECL is zero with the alternative hypothesis being that the change is greater than zero. To evaluate this hypothesis, a one-sided test at the 5% significance level will be employed.

The percent change from baseline at Week 40 in dystrophin levels measured by ECL will be analyzed using the mixed effect model repeated measurement (MMRM) analysis with factors of muscle locations, visits (baseline and Week 40) as fixed effects, and subjects as a random effect, based on ITT population. An unstructured covariance matrix will be used, and if the model fails to converge, then a compound symmetry matrix will be used. Since dystrophin level is known to be non-normally distributed, a log-transformation will be used in the analysis. The best dystrophin levels across all muscles are considered the primary analysis. A sample SAS code of the MMRM model is provided as follows:

```
proc mixed data = dataset;
  class subjid visit mlocation;      /*mlocation is muscle location*/
  model lnaval = visit mlocation; /*lnaval is log-transformed dystrophin level*/
  random subjid;
  repeated visit / type = un;
  estimate "Week 40 change from baseline" visit -1 1;
run;
```

As supportive analyses, the average and median dystrophin levels across all muscles, and the subgroup analyses of the best, median, and mean dystrophin levels within each muscle location will be considered. In addition, the two-sided 90% confidence interval of percent change from baseline in dystrophin levels at Week 40 and the ratio (%) of dystrophin level at Week 40 versus baseline will also be provided.

7.1.2. Subgroup Analyses

The primary endpoint described in Section 7.1.1 will be summarized using descriptive statistics by each subgroup listed below. No hypothesis testing will be performed in the subgroup analyses.

- Muscle location
- Steroid use
- Exon mutation location
- Visit 3 biopsy date (impacted by COVID-19 and not impacted by COVID-19)

7.1.3. Sensitivity Analyses

The following analyses will be considered as sensitivity analyses:

- PP analysis: The same primary efficacy analysis described in Section 7.1.1 (Primary Analysis) will be repeated based on the PP population for the primary efficacy endpoint.
- The average and median dystrophin levels across all muscle locations based on the same method as described in Section 7.1.1 (Primary Analysis).
- Excluding subjects with baseline dystrophin levels [redacted] of the predicted levels: The same primary analysis described in Section 7.1.1 (Primary Analysis) will be repeated excluding subjects with baseline dystrophin levels [redacted] of the predicted levels. The same analysis on the average and median dystrophin levels across all muscle locations will also be repeated excluding such subjects.
- COVID-19 consideration: The Week 40 biopsy study day (defined as Week 40 biopsy date – first study drug date +1) will be added to the primary efficacy analysis model described in Section 7.1.1 (Primary Analysis) as a covariate. This analysis model will

be applied to the best, average, and median dystrophin levels analyses based on the ITT population.

7.2. Secondary Efficacy Analyses

7.2.1. Secondary Efficacy Analysis

The secondary efficacy endpoint is the percent change from baseline in dystrophin levels (mean membrane stain density) after 40 weeks of ataluren therapy as determined by a validated IHC assay. Within each muscle location, the best, median, and the average dystrophin levels of the muscle cores will be summarized using descriptive statistics for each subject. The secondary efficacy endpoint will be analyzed using the same method as described in Section 7.1.1 (Primary Analysis). The best dystrophin levels across all muscles are considered the primary interest. As supportive analyses, the average and median dystrophin levels across all muscles, and subgroup analyses of the best, median, and mean dystrophin levels within each muscle location will be considered. In addition to dystrophin levels (the mean staining intensity), the percentage of positive fibers will also be summarized using descriptive statistics.

7.2.2. Sensitivity Analyses

The following analyses will be considered as sensitivity analyses for the secondary efficacy endpoint:

- PP analysis: The same secondary efficacy analysis described in Section 7.2.1 (Secondary Analysis) will be repeated based on PP population for the secondary efficacy endpoint.
- The average and median dystrophin levels (mean membrane stain density) across all muscle locations will be analyzed using the same analysis method as described in Section 7.2.1 (Secondary Analysis).
- Excluding subjects with baseline dystrophin levels [redacted] of the predicted levels: The same analysis described in Section 7.2.1 (Secondary Analysis) will be repeated excluding subjects with baseline dystrophin levels [redacted] of the predicted levels. The same analysis on the average and median dystrophin levels across all muscle locations will also be repeated excluding such subjects.
- Subgroup analysis: The subgroup analysis by muscle location will be performed on secondary efficacy endpoint as described in Section 7.1.2 (Subgroup Analyses).
- COVID-19 considerations: The Week 40 study day (defined as Week 40 biopsy date – first study drug date +1) will be added to the efficacy analysis model described in Section 7.2.1 (Secondary Analysis) as a covariate. This analysis model will be applied to the best, average, and median dystrophin levels analyses based on the ITT population.

7.3. Exploratory Analyses

No hypothesis testing will be performed on all exploratory endpoints. All exploratory endpoints will be summarized using descriptive statistics based on ITT population.

7.3.1. Revised North Star Ambulatory Assessment

The rNSAA consists of 17 activities, each scored as 0, 1, or 2. The sum of the 16 activities (exclude head lift) will form a total score. If fewer than 13 of the 16 activities are performed, the total score will be considered missing. If from 13 to 16 activities are performed, the total score will be standardized as follows:

$$\textit{Standardized total score} = \frac{\textit{observed total score}}{\textit{number of non-missing activities}} \times 16$$

Transformation of the total score into a linear scale (Mayhew 2013a) will also be performed.

The mean and mean change from baseline in the total score and its linear transformation will be summarized at the Baseline and Week 40 visits based on ITT population. In addition, the proportion of subjects who has lost function after 40 weeks of ataluren treatment, defined as a drop of score from 1 or 2 at baseline to a score of 0 at the Week 40 visit, will be tabulated for each activity.

7.3.2. Timed Function Test

The TFTs, the time taken to run/walk 10 meters, climb 4 stairs, descend 4 stairs, and stand from supine, will be assessed at the Baseline and Week 40 visits. Subjects who cannot perform a timed function test within 30 seconds, including those who loss of ambulation or the timed function test is above 30 seconds, will be assigned a value of 30 seconds for the respective test. The mean and mean change from baseline for each timed function test will be summarized at each visit based on the ITT population.

7.3.3. Creatine kinase levels

The CK levels will be assessed at the Screening and Week 40 visits. The mean and mean change from baseline of CK levels will be summarized at each visit based on the ITT population.

8. SAFETY EVALUATION

All safety summaries will be generated based on the safety population.

8.1. Adverse Events

Adverse events will be coded using the MedDRA version 23.0. A treatment-emergent adverse event (TEAE) is defined as an adverse event that occurs or worsens on or after the first dose of study drug up to 6 weeks after the last dose of st\udy drug. Any AE with an end date prior to the first dose of the study drug will not be considered as an TEAE.

The following summaries will be provided for TEAEs:

- TEAE overview
- TEAEs by SOC and PT

If a subject has multiple events under any given SOC and PT, the subject will be counted only once under that SOC and PT. If a subject has the same AE on multiple occasions, the highest severity reported on the CRFs ("life-threatening", "severe", "moderate", "mild", "unknown", and "not reported") will be presented in the summaries.

In addition, subjects with treatment-emergent SAE and TEAE leading to discontinuation will also be summarized.

8.2. Clinical Laboratory

Clinical laboratory assessments for hematology, biochemistry, and urine evaluations are collected at Screening, Baseline and Week 40 visits. Mean and mean change from baseline for all hematology and biochemistry laboratory assessments, and continuous variables for urinalysis will be summarized using descriptive statistics at Baseline and Week 40 visits. Categorial variables for urinalysis will be tabulated using frequency count at Baseline and Week 40 visits. In addition, shift tables for each laboratory parameters from baseline to post-baseline visits will also be provided.

Creatine kinase, summarized as an efficacy endpoint, will be excluded from all laboratory summaries and listings.

8.3. Vital Signs

Vital signs including systolic blood pressure, diastolic blood pressure, pulse rate, and oral temperature are collected at Baseline and Week 40 visits. Mean and mean change from baseline for each vital sign parameters will be summarized using descriptive statistics at each visit.

8.4. Other Safety Assessments

Other safety assessments including physician assessments will be presented in the subject listings only.

9. PHARMACOKINETIC BLOOD SAMPLE

The PK assessments using sparse sampling and population PK modeling will be presented in a separate report. The mean, median, minimum, maximum, %CV and geometric mean of plasma concentration at pre-dose and post-dose at each visit will be summarized based on the ITT population. In case any measured plasma concentrations are below LLOQ, the plasma concentrations will be set as 0 in calculations for the statistical analysis.

10. CHANGES FROM THE PLANNED ANALYSES

The following changes to study protocol V3.0 (dated 06May2019) are implemented in this SAP:

- As requested by US FDA, inference for the primary and secondary efficacy endpoints will be evaluated using a one-sided test at the 0.025% significance level.
- The primary and secondary efficacy analyses will be performed including all subjects. Subjects with baseline dystrophin levels [redacted] predicted will not be removed from the primary and secondary efficacy analyses.
- Sensitivity analyses excluding subjects with baseline dystrophin levels [redacted] predicted will be added to both primary and secondary efficacy endpoints.

11. REFERENCES

Mayhew, AG, Cano, SJ, Scott, E, Eagle, M, Bushby, K, Manzur, A, et al. Detecting meaningful change using the north star ambulatory assessment in duchenne muscular dystrophy. Dev Med Child Neurol 2013a;55(11):1046-1052.

Syneos Health Report No. 14597.051619.A, Analysis of dystrophin in human muscle tissue homogenate samples from control subjects for baseline determination and preparation of assay quality controls (non-regulated).