STATISTICAL ANALYSIS PLAN

DUCHENNE MUSCULAR DYSTROPHY: DOUBLE-BLIND RANDOMIZED TRIAL TO FIND OPTIMUM STEROID REGIMEN

Study Acronym: FOR-DMD

Protocol Version: 3.2

Clinical Phase: Phase IV

IND #: 109849

Principal Investigators, Clinical Center
Robert C. Griggs, M.D.
Michela Guglieri, M.D.

Principal Investigators, Biostatistics and Coordination Center
Michael P. McDermott, Ph.D.
Rabi Tawil, M.D.

Version 1.0: February 10, 2012
Version 2.0: February 16, 2021

Supported by Grants U01 NS061799 (University of Rochester) and U01 NS061795 (University of Rochester) from the National Institute of Neurological Disorders and Stroke (NINDS)

SIGNATURES

Michael P. McDermott, Ph.D.
Co-Principal Investigator
Director, MSG Biostatistics Center

2/16/21
Date

Rabi Tawil, M.D.
Co-Principal Investigator
Director, MSG Coordination Center

2/16/2021
Date

Robert C. Griggs, M.D.
Co-Principal Investigator
Study Co-Chair

Date

Michela Guglieri, M.D.
Co-Principal Investigator
Study Co-Chair and Chief Medical Director

18 Feb 2021
Date

TABLE OF CONTENTS

1. STUDY OBJECTIVES

1.1. Primary Objective

The primary objective of this study is to compare the three most commonly used corticosteroid regimens (0.75mg/kg/day prednisone; 0.75mg/kg/day prednisone 10 days on/10 days off; 0.9mg/kg/day deflazacort) with regard to functional outcome and participant/parent satisfaction with treatment in boys with Duchenne muscular dystrophy (DMD) previously untreated with corticosteroid therapy. The trial will address the pragmatic hypothesis that both daily prednisone and daily deflazacort will be of greater benefit in terms of function and participant/parent satisfaction with treatment than intermittent prednisone given over a 36-month period.

1.2. Secondary Objective

The secondary objectives are to compare the three corticosteroid regimens over a period of 36 months with regard to tolerability, adverse events, secondary functional outcomes and disease milestones, quality of life, and cardiac function.

2. STUDY DESIGN

FOR-DMD is an international multicenter, randomized, double-blind, controlled trial in which approximately 225 boys with proven DMD between the ages of 4 and 7 will be randomized at 32 sites in 5 countries to receive either daily prednisone (0.75 mg/kg), daily deflazacort (0.90 mg/kg), or intermittent (10 days on, 10 days off) prednisone (0.75 mg/kg) and followed for a minimum of 36 months. Eligible boys will be those with confirmed DMD, age at least 4 years and under 8 years; ability to rise independently from the floor; and ability to maintain reproducible forced vital capacity (FVC) measurements. Detailed eligibility criteria are provided in Sections 4.1 and 4.2 of the Study Protocol.

Following provision of informed consent and screening and baseline visits to determine eligibility, randomization to the study interventions will occur. The boys will be followed prospectively and systematically for 36-60 months of double-blind observation. As described in the Study Protocol (Section 5.3), there is careful monitoring for participant safety and the design includes standardized management protocols for prophylaxis and treatment of adverse events, particularly those that are anticipated to occur with chronic corticosteroid treatment over the course of the trial.

The Schedule of Activities is described in detail in Sections 6.1-6.3 of the Study Protocol. Assessments of the primary and secondary outcomes will be conducted at up to 12 in-person visits (baseline, Months 3, 6, 12, 18, 24, 30, 36, 42, 48, 54, and 60). Telephone contacts half-way between each visit will be made to inquire about adverse events and medication compliance and to aid in study retention. The primary outcome variable will be based on data obtained through Month 36, but all participants will be followed for up

to 60 months until the last participant has completed the 36-month visit. The core evaluations during each in-person visit are described in detail in the Study Protocol. The dosage of study medication will be adjusted as appropriate according to weight and, if necessary, adverse events. Provisions for switching to daily prednisone in case of temporary interruption of study medication supply (e.g., due to expiration or loss) are provided in Section 5.1.2 of the Study Protocol. Participants who withdraw from study medication will continue to be followed in the trial, if willing, even if they are participating in a clinical trial of an investigational treatment and the protocol for that trial allows concurrent participation in the FOR-DMD trial.

3. RANDOMIZATION AND BLINDING

The randomization process is described in Section 4.3.4 of the Study Protocol. A programmer in the Muscle Study Group (MSG) Biostatistics Center will generate the randomization plan using a SAS program, in accordance with the MSG Biostatistics Center SOPs. The randomization will be stratified by country and will include blocking to ensure balance among the treatment groups after a certain number of participants have been enrolled in each country.

Assignment lists for each country will be generated for use in the enrollment module and a separate fill list will be generated for the clinical trials supply company prior to the start of the trial. Randomization will be initiated by the site investigator/designee via a secure web-based system with username and password access. After appropriate information verifying the participant's eligibility and weight, the participant ID number, randomization ID number, and starting dosage will be generated and sent to the clinical trials supply company. Randomization procedures should be completed at least 10 days prior to the baseline visit in order to allow the study medication to be shipped to the site in time to be provided to the participant by the site investigator at baseline visit.

As described in Section 4.3.5 of the Study Protocol, the double-blind is facilitated by the preparation and packaging of study medication. All study personnel, with the exception of a designated unblinded programmer and unblinded statistician in the MSG Biostatistics Center, will have access to the treatment assignments. The chief roles of the unblinded programmer and statistician will involve preparing unblinded reports for closed sessions of meetings of the Data and Safety Monitoring Board (DSMB); they will not communicate with any other involved staff regarding study-related matters.

The unblinded programmer in the MSG Biostatistics Center will prepare sealed emergency drug disclosure envelopes with individual treatment assignments for distribution to the sites as participants are enrolled. Potential reasons for unblinding the treatment assignment for an individual participant include the following:

- Occurrence of a medical emergency such that knowledge of treatment assignment is necessary for providing appropriate medical management.

- Desire of the participant to enroll in a clinical trial of an investigational treatment for which knowledge of the participant's treatment assignment is required for eligibility (allowed only for participants who have completed the Month 36 visit).
- Judgment on the part of the site investigator that revelation of the participant's treatment assignment is clinically warranted (only allowed for participants who have completed the Month 36 visit).
- Temporary interruption of study medication supply that persists beyond 21 days, in which case knowledge of whether the participant has been assigned to a daily vs. intermittent corticosteroid regimen is required for participant safety (see Section 5.1.2 of the Study Protocol).

In each case, unblinding will be restricted to those who have a need to know the identity of the treatment assignment, as outlined in Section 4.3.5 of the Study Protocol.

4. OUTCOME VARIABLES

4.1. Primary Outcome Variable

The primary outcome variable is a three-dimensional (multivariate) outcome consisting of the following three components (each averaged over all post-baseline follow-up visits through Month 36): (1) rise from the floor velocity; (2) forced vital capacity; and (3) participant/parent global satisfaction with treatment, as measured by the Treatment Satisfaction Questionnaire for Medication (TSQM). These components represent different but important aspects of the expected benefits of corticosteroid treatment.

Two components of the primary outcome variable (rise from the floor velocity and FVC) were selected because of obvious clinical relevance. They also relate to longer term outcomes (loss of ambulation and respiratory failure) that are impossible to capture in a study with 36-60 months of participant follow-up. The third component of the primary outcome variable, global satisfaction with treatment, is essential for measuring the benefit/side effect balance that is crucial to consider for a non-curative treatment in a chronic disease. Importantly, all three components of the primary outcome variable are relatively straightforward and easy to measure.

4.2. Secondary Outcome Variables for Efficacy

The secondary outcome variables for efficacy include the following (each averaged over all post-baseline follow-up visits):

- Ten-meter walk/run velocity
- Distance walked in 6 minutes (6MWT)
- North Star Ambulatory Assessment (NSAA) total score
- TSQM subscale scores (Effectiveness, Side Effects, Convenience)
- Lean mass (raw value and percentage of total mass) (DXA)
- Range of motion at ankle dorsiflexion (left and right sides)

- Quality of life (Generic PedsQL: Physical Functioning, Emotional Functioning, Social Functioning, School Functioning, Psychosocial Health, and total scores; Neuromuscular disease-specific module: About My Neuromuscular Disease, Communication, About Our Family Resources, and total scores)

Other secondary outcome variables, based on the NSAA, will include times from the baseline evaluation to loss of independent ambulation and time to loss of the ability to rise from the floor. The time at loss of ambulation will be defined in two ways: (1) the post-baseline time when the participant was first classified as "Unable" on NSAA Item 2 (Walk 10 Meters) and (2) the first post-baseline time at which the participant was unable to perform the timed task of Walk/Run 10 Meters. The time at loss of the ability to rise from the floor will be defined as the first post-baseline time at which the participant was unable to perform the timed task of Rise from the Floor.

Transitions over time in other individual NSAA item scores will also be analyzed.

4.3. Outcome Variables for Safety and Tolerability

Measures of safety include the following:

- Corticosteroid dosage
- Adverse events (AEs)
- Serious adverse events (SAEs)
- Height, weight, body mass index (BMI)
- Vital signs (systolic/diastolic blood pressure, pulse)
- DXA bone outcomes (bone area, bone mineral content [BMC], bone mineral density [BMD], and Z-score in the lumbar spine; bone area, BMC, BMD, and Z-score in total body)
- DXA fat mass (raw value and percentage of total mass) and total mass
- Wrist radiography outcomes (bone age, difference between chronological age and bone age)
- Echocardiography and ECG outcomes (ejection fraction, fractional shortening, PR interval, heart rate)
- Iowa Conners ASQ–Parent scale scores (Inattentive/Overactive [I/O], Oppositional/Defiant [O/D], and total scores, I/O score ≥ 10, O/D score ≥ 9)
- PARS-III scores (Peer Relations, Dependency, Hostility, Productivity, Anxiety/Depression, Withdrawal, and total scores, total score < 72)
- Strengths and Difficulties Questionnaire (SDQ) scores (Emotional Problems, Conduct Problems, Hyperactivity, Peer Problems, Prosocial, Externalizing, Internalizing, Total Difficulties, and Impact scores)
- Revised Rutter scale total score
- Laboratory test results (urinalysis, serum chemistry, and hematology)

Measures of tolerability include the following:

- Ability to complete the trial on the originally assigned dosage (for weight) of study medication
- Ability to complete the trial without switching medication regimens (i.e., discontinuing study medication)
- Ability to complete the trial

4.4. Scoring Considerations

4.4.1. Timed Functional Tests

For the time to rise from the floor and time to walk/run 10 meters measures, the value becomes infinite when the participant is not able to perform the task. In the database, the values for these outcomes will be entered as "999". These outcomes will be transformed using reciprocals for purposes of analysis, with those unable to perform the task being assigned a transformed value of zero. This strategy, the use of a velocity-type measure rather than a time, has the following advantages: (1) the transformed values are approximately normally distributed; and (2) like the other two components of the primary outcome variable, higher values of the outcome variable represent better performance. It is becoming standard in the analysis of data from DMD clinical trials.

4.4.2. Forced Vital Capacity (FVC)

The summary measure of FVC will be taken as the maximum value (in liters) over the three trials performed.

4.4.3. Height and Weight

Besides direct measures of height and weight, Z-scores and percentiles will be computed using CDC growth charts (https://www.cdc.gov/growthcharts/percentile_data_files.htm). Heights that were estimated using ulna length will not be included in the analyses.

4.4.4. Treatment Satisfaction Questionnaire for Medication (TSQM)

- Effectiveness score = (Sum of Items 1-3 – 3) × 100 / 18. If one item is missing, then this is computed as (Sum of completed items – 2) × 100 / 12. If more than one item is missing, then the score is missing.
- Side Effects score = 100 if Item 4 = No; otherwise, Side Effects score = (Sum of Items 5-8 – 4) × 100 / 16. If one item is missing, then this is computed as (Sum of completed items – 3) × 100 / 12. If more than one item is missing, then the score is missing.
- Convenience score = (Sum of Items 9-11 – 3) × 100 / 18. If one item is missing, then this is computed as (Sum of completed items – 2) × 100 / 12. If more than one item is missing, then the score is missing.
- Global Satisfaction score = (Sum of Items 12-14 – 3) × 100 / 14. If either Item 12 or Item 13 is missing and Item 14 is non-missing, then this is computed as (Sum of completed items – 2) × 100 / 10. If Item 14 is missing and Items 12 and 13 are

non-missing, then this is computed as (Sum of completed items – 2) × 100 / 8. If more than one item is missing, then the score is missing.

4.4.5. PedsQL Generic Core

Prior to computing scores, all items are reverse scored and linearly transformed to a 0-100 scale: 0 = 100; 1 = 75, 2 = 50, 3 = 25, 4 = 0.

- Physical Functioning score = Mean of Items 1-8. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Emotional Functioning score = Mean of Items 1-5. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Social Functioning score = Mean of Items 1-5. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- School Functioning score = Mean of Items 1-5. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Psychosocial Health score = Mean of all items included in the Emotional Functioning, Social Functioning, and School Functioning subscales. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Total score = Mean of all items included in the instrument. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.

4.4.6. PedsQL Neuromuscular Module

Prior to computing scores, all items are reverse scored and linearly transformed to a 0-100 scale: 0 = 100; 1 = 75, 2 = 50, 3 = 25, 4 = 0.

- About My Neuromuscular Disease score = Mean of Items 1-17. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Communication score = Mean of Items 1-3. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- About Our Family Resources score = Mean of Items 1-5. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.
- Total score = Mean of all items included in the instrument. If 50% or fewer of the items are missing, then the mean is taken over the completed items. If > 50% of the items are missing, then the score is missing.

4.4.7. Iowa Conners ASQ-Parent Scale

- Inattentive/Overactive (I/O) score = Sum of Items 1-5
- Oppositional/Defiant (O/D) score = Sum of Items 6-10
- Total score = Sum of Items 1-10

4.4.8. PARS-III Scale

For Items 1-4 and 15-18, scoring should be as follows:

Never or Rarely = 1, Sometimes = 2, Often = 3, Always = 4

For Items 5-14 and 19-28, scoring should be as follows:

Never or Rarely = 4, Sometimes = 3, Often = 2, Always = 1

- Peer Relations score = Sum of Items 1-4
- Dependency score = Sum of Items 5-8
- Hostility score = Sum of Items 9-14
- Productivity score = Sum of Items 15-18
- Anxiety/Depression score = Sum of Items 19-24
- Withdrawal score = Sum of Items 25-28
- Total score = Sum of Items 1-28

If a 4-item subscale is missing more than 1 item or if a 6-item subscale is missing more than 2 items, then the score is missing. Otherwise, the score is computed as the mean of the completed items × the total number of items on the scale (4 or 6).

4.4.9. Strength and Difficulties Questionnaire (SDQ)

Most items are scored as Not True = 0, Somewhat True = 1, and Certainly True = 2. The exceptions are for Items 7, 11, 14, 21, and 25 which are scored as Not True = 2, Somewhat True = 1, and Certainly True = 0.

- Emotional Problems score = Sum of Items 3, 8, 13, 16, and 24
- Conduct Problems score = Sum of Items 5, 7, 12, 18, and 22
- Hyperactivity score = Sum of Items 2, 10, 15, 21, and 25
- Peer Problems score = Sum of Items 6, 11, 14, 19, and 23
- Prosocial score = Sum of Items 1, 4, 9, 17, and 20

If a subscale is missing more than 2 items, then the score is missing. Otherwise, the score is computed as the mean of the completed items × 5.

- Externalizing score = Sum of Conduct Problems and Hyperactivity scores
- Internalizing score = Sum of Emotional Problems and Peer Problems scores

- Total Difficulties score = Sum of Emotional Problems, Conduct Problems, Hyperactivity, and Peer Problems scores
- Impact score = 0 if the child reports no difficulties. Otherwise, this is computed as the sum of the Upset/Distress, Home Life, Friendships, Classroom Learning, and Leisure Activities items, each scored as Not at All = 0, Only a Little = 0, Quite a Lot = 1, and A Great Deal = 2.

4.4.10. Revised Rutter Scale

The total score is the sum of Items 1-8.

5. SAMPLE SIZE DETERMINATION

As originally planned, the primary outcome variable for the trial was to be three-dimensional consisting of time to rise from the floor (log-transformed) averaged over all post-baseline visits during the 36-month follow-up period, forced vital capacity (FVC) averaged over all post-baseline visits, and (3) participant/parent global satisfaction with treatment, as measured by the TSQM, averaged over all post-baseline visits. The primary statistical analysis was planned as a global test of the null hypothesis that the corticosteroid regimens do not differ with regard to any of the three outcomes against the alternative that they differ (in the same direction) with regard to at least one of the outcomes. The sample size was determined assuming that O'Brien's ordinary least-squares (OLS) statistic [O'Brien, 1984] would be used to carry out this test. The analyses will involve three separate pair-wise comparisons among the three treatment regimens; each will be performed using a Bonferroni-corrected two-tailed significance level of 0.017.

5.1. Supporting Data and Justification for the Chosen Effect Size

Existing data from the trials comparing 0.75 mg/kg/day prednisone vs. placebo suggest that the effects of prednisone on the time to rise from the floor (log-transformed) and FVC outcomes are of a magnitude of one standard deviation unit (difference between the treatment group means divided by the standard deviation of the outcome, adjusted for baseline values) [Mendell et al., 1989; Griggs et al., 1991; Manzur et al., 2004]. A difference in mean response between two corticosteroid regimens of 0.50 standard deviation units (i.e., half of the effect of 0.75 mg/kg/day of prednisone vs. placebo) was considered to be the group difference that is of minimal clinical significance on these two outcomes.

For global satisfaction with treatment on the TSQM, data on 378 subjects (mean age 50.5 years, range 18 to 88 years) using oral medications for a variety of illnesses (migraine, arthritis, high blood pressure, high cholesterol, and depression) suggest that the group difference of minimal clinical significance may be slightly less than 0.50 standard deviation units [Atkinson et al., 2004]. Subjects who described the seriousness of their illness as "Moderate" and those who described it as "Severe" differed by approximately

0.40 standard deviation units on this scale. Also, subjects who appraised their health as "Excellent" or "Very Good" and those who appraised their health as "Good" or "Fair" differed by approximately 0.50 standard deviation units on this scale. These data are limited in that they were not obtained in children or parents of children with Duchenne muscular dystrophy, but they do provide some idea of what group differences in mean response might be considered to be clinically important.

We performed a pilot study of the TSQM in 37 boys with Duchenne muscular dystrophy (DMD) between the ages of 4 and 12 years who were taking corticosteroids, 6 at the University of Rochester, 7 at the University of Kansas, 6 at Children's Hospital of Western Ontario (Canada), and 18 at the University of Newcastle (UK) [Herr et al., 2008]. We collected data on the age and ambulatory status of the participants, allowing us to examine the relationships between the TSQM subscale scores and these characteristics (Table 1). Although the small sample sizes limit our ability to detect significant differences among the groups determined by age or ambulatory status, the expected patterns of responses are apparent: lower mean Effectiveness and Global Satisfaction scores among those > 7 years old ($p < 0.05$) and among those unable to walk outside the home; lower Side Effects scores among those > 10 years old and among those unable to walk outside the home.

Table 1. Results of the TSQM Pilot Study in 37 Boys with DMD

	Age Group			Ambulatory Status	
Scale	**4-6 (n = 9)**	**7-9 (n = 11)**	**10-12 (n = 17)**	**Able (n = 29)**	**Home/Unable (n = 8)**
Effectiveness	79.0 (16.1)	58.6 (22.7)	61.4 (26.2)	68.8 (20.4)	50.7 (31.5)
Side Effects	85.9 (18.2)	83.5 (17.1)	71.2 (29.3)	82.6 (19.4)	68.8 (31.9)
Convenience	88.9 (15.7)	90.6 (11.4)	88.0 (16.6)	89.6 (13.9)	87.3 (17.5)
Global Satisfaction	82.5 (18.2)	67.1 (18.5)	63.9 (26.7)	73.2 (18.7)	54.1 (35.2)

Data from this study indicate that the mean global satisfaction with treatment differed by 0.65-0.80 standard deviation units between boys age 4-6 and those age 7-12. In this context, an effect size of 0.40 standard deviation units does not appear to be unacceptably large.

The power of the test based on O'Brien's OLS statistic also depends on the correlations among the three outcomes. Data on 51 participants between the ages of 5 and 7 from the Clinical Investigation of Duchenne Dystrophy (CIDD) natural history cohort suggest that the correlation between the average time to rise from the floor (log-transformed) and the average FVC over a 36-month follow-up period is very small and possibly even slightly negative. Correlations between these outcomes and the TSQM global satisfaction rating, while expected to be mildly-to-moderately positive, were of unknown magnitude at the onset of the trial. For this reason, an interim sample size re-estimation was performed in July 2015 using all available longitudinal data, blind to treatment group assignment. The

results revealed that the three components of the primary outcome variable (log-transformed time to rise from the floor, forced vital capacity, and participant/parent global satisfaction with treatment) were at best weakly correlated ($0 < r < 0.25$).

5.2. Sample Size Determination for the Primary Outcome Variable

Table 2 provides the power (estimated using Monte-Carlo simulation, with each estimate based on 50,000 replications) required to detect the specified treatment effects (differences between two groups, expressed in standard deviation units) under various assumptions regarding the correlations among the three outcomes, using O'Brien's OLS test and a two-tailed significance level of 0.017. A sample size of 67 participants per group is assumed. This table indicates that a total sample size of 201 participants (67 per group) will provide at least 80% and often greater than 90% power to detect treatment effects of at least 0.45 standard deviation units on at least two of the three components and as little as 0.25 standard deviation units on the third component of the primary outcome variable. The power becomes unacceptably low only in the case where there is no group difference in treatment satisfaction. To account for an anticipated 10% participant withdrawal rate, 225 participants (75 per group) were to be enrolled in the trial.

Table 2. Power Calculations for the Primary Outcome Variable

Effect Sizes			Correlations			
Δ_T	Δ_F	Δ_S	ρ_{TF}	ρ_{TS}	ρ_{FS}	Power
0.45	0.45	0	0	0	0	72%
0.45	0.45	0	0	0.10	0.10	66%
0.45	0.45	0	0	0.15	0.15	63%
0.45	0.45	0	0	0.25	0.25	58%
0.45	0.45	0.25	0	0	0	92%
0.45	0.45	0.25	0	0.10	0.10	88%
0.45	0.45	0.25	0	0.15	0.15	86%
0.45	0.45	0.25	0	0.25	0.25	82%
0.45	0.45	0.45	0	0	0	98%
0.45	0.45	0.45	0	0.10	0.10	96%
0.45	0.45	0.45	0	0.15	0.15	96%
0.45	0.45	0.45	0	0.25	0.25	93%
0.50	0.50	0	0	0	0	83%
0.50	0.50	0	0	0.10	0.10	77%
0.50	0.50	0	0	0.15	0.15	74%
0.50	0.50	0	0	0.25	0.25	69%
0.50	0.50	0.25	0	0	0	96%
0.50	0.50	0.25	0	0.10	0.10	93%

0.50	0.50	0.25	0	0.15	0.15	92%
0.50	0.50	0.25	0	0.25	0.25	88%

Δ_T = Effect size for log(time to rise from the floor); Δ_F = Effect size for FVC; Δ_S = Effect size for global satisfaction with treatment. All values are expressed in standard deviation units.

ρ_{TF} = Correlation between log(time to rise from the floor) and FVC; ρ_{TS} = Correlation between log(time to rise from the floor) and global satisfaction with treatment; ρ_{TF} = Correlation between FVC and global satisfaction with treatment.

5.3. Sample Size Considerations for Safety Outcomes

A secondary hypothesis concerns potential differences between daily deflazacort and daily prednisone with respect to adverse events, with intolerability (dosage reduction) and weight gain being the events of primary interest. Data from the long-term trials of prednisone [Fenichel et al., 1991] suggest that nearly 50% of participants may require a lower dosage of daily prednisone than 0.65 mg/kg during the 36-month follow-up period. The data from this study also indicate that participants in the daily prednisone group will have an average weight gain of approximately 25% per year (standard deviation = 12%). The sample size of 225 participants (75 per group) will provide 80% power to detect a group difference of 22% (50% in the daily prednisone group vs. 28% in the daily deflazacort group) regarding the percentage of participants taking a reduced dosage at the end of the 36-month follow-up period, using a chi-square test and a 5% significance level. It will also provide > 80% power to detect a group difference of 6% (25% in the daily prednisone group vs. 19% in the daily deflazacort group) in mean annual weight gain during the 36-month follow-up period, using a t-test and a 5% significance level (two-tailed).

Due to difficulties in recruitment, the FOR-DMD trial ultimately randomized 196 participants.

6. GENERAL DATA HANDLING AND STATISTICAL CONSIDERATIONS

6.1. Analysis Cohorts

Given the highly pragmatic nature of the FOR-DMD trial, the primary statistical analyses will be performed according to the treatment to which the participants were randomly assigned and with minimal exclusion of data collected during the trial. All analyses will be based on data collected up to and including the Month 36 visit.

Primary Analysis Cohort: The primary analyses of all efficacy and safety data (other than adverse event data and tolerability outcomes) will be performed using all available data from all randomized participants who contribute at least one post-baseline value of the outcome of interest, with the exception of data obtained after enrollment of the participant in another trial of an investigational treatment. In particular, the following data will be included:

- Data from participants obtained after they discontinued study medication but remained in follow-up on a corticosteroid regimen of their family's choice
- Data from participants obtained after unblinding of the identity of the participant's assigned corticosteroid regimen
- Data from participants obtained during temporary interruption of study medication causing them to switch to open-label prednisone

The analysis of adverse event data and tolerability outcomes will be performed using data from all randomized participants, i.e., the denominators in the calculation of percentages of participants experiencing adverse events or intolerability will be the numbers of randomized participants in the treatment groups.

Sensitivity Analysis Cohort: Analyses of the efficacy and safety data will be repeated after excluding data from the three categories listed above (after discontinuation of study medication, after unblinding, during temporary interruption of study medication).

6.2. General Considerations

All analyses described in this document were pre-specified and defined prior to breaking the study blind. All other analyses performed subsequently will be considered *post hoc* and exploratory.

Unless otherwise specified, all significance testing will be performed using a two-tailed 5% level. Likewise, confidence intervals will be computed using a 95% confidence coefficient.

For purposes of statistical analysis, baseline values will be considered to be those that are the last measurements obtained at or before the Month 0 visit and prior to the first dose of study medication. If a value was scheduled to be obtained at the Month 0 visit but is missing, the value at the screening visit may be used instead.

Due to a variety of factors, most notably scheduling difficulties, several visits may be performed out-of-window as defined by the Study Protocol. The statistical analyses of longitudinal data will use the visit label assigned to each assessment regardless of whether the visit was technically performed out-of-window. This convention will minimize the amount of missing data. Data from unscheduled visits (including early withdrawal visits) will be assigned to the closest scheduled post-baseline visit that includes the assessment(s) of interest. For efficacy assessments, if there is more than one value recorded at a particular visit, the value from the unscheduled visit will not be used. For safety assessments, if there is more than one value recorded at a particular visit, the worst value will be used for analysis.

Considerations for dealing with other specific data anomalies are provided in the Appendix.

6.3. Interim Analyses

Interim analyses of safety data will be performed periodically (approximately every 6 months) throughout the trial. A Data and Safety Monitoring Board (DSMB) appointed by the trial sponsor (NINDS) will review reports of participant safety, data quality, and other indicators of study performance, including accrual. The adverse events associated with daily corticosteroid use in DMD have been reasonably well characterized in previous studies and include behavioral changes, fractures, cataracts, Cushingoid appearance, GI symptoms, glycosuria, hypertension, immunosuppression (e.g., infections), slow growth (height restriction), skin changes (e.g., excessive hair growth, acne, atrophy, easy bruising), and abnormal weight gain. Detailed procedures are in place to clinically manage these anticipated adverse events (see Sections 5.3 and 7.4 of the Study Protocol).

Due to concerns regarding slow enrollment in the trial, a blinded interim analysis was performed after 90 participants had completed 12 months of follow-up (July 2015) in order to determine whether the sample size could be reduced from the originally planned 300 to a number that was feasible given the study timeline. This determination was made based on the correlations among the three components of the primary outcome variable (averaged over available time points), which were completely unknown for those involving the TSQM Global Satisfaction score. The results revealed that the three components of the primary outcome variable were at best weakly correlated ($0 < r < 0.25$), allowing the target sample size to be reduced to 225 (see Sections 5.1 and 5.2 above). The DSMB agreed with this proposal.

A single interim analysis for efficacy was performed after 95 participants had completed (or were scheduled to have completed) 36 months of follow-up (June 2017). The analysis of the primary outcome variable was performed using an overall significance level of 0.001 (instead of 0.05), a conservative Haybittle-Peto-type strategy. Upon reviewing the results of this analysis, the DSMB recommended continuation of the trial. No interim analyses were performed for the purpose of early identification of futility.

7. STATISTICAL ANALYSIS

7.1. Baseline Characteristics

Baseline characteristics of participants, including demographic information, family history, medical history, and efficacy and safety outcomes will be presented overall and by treatment group for all randomized participants. Quantitative variables will be summarized with sample size, mean, standard deviation, median, 1st and 3rd quartiles, minimum, and maximum. Categorical variables will be summarized using frequencies and percentages.

7.2. Participant Disposition

A complete accounting of participant disposition will be summarized overall and by treatment group, including a tabulation of participants in the following categories: completion of follow-up on study medication, completion of follow-up after permanent discontinuation of study medication, withdrawal while on study medication, and withdrawal after permanent discontinuation of study medication.

Other tabulations will include those of temporary interruption of study medication causing a switch to open-label prednisone, unblinding of the identity of the participant's assigned corticosteroid regimen, and enrollment in another trial of an investigational treatment.

Listings of the reasons for withdrawal, permanent discontinuation of study medication, and unblinding of the identity of the participant's assigned corticosteroid regimen will be produced, along with the timing of these events.

7.3. Exposure to study medication

Each participant's exposure to study medication will be summarized separately for each treatment group by the total number of days on study medication and the average dosage (expressed in mg/kg of body weight) over all days of exposure. Periods of time when the participant had a temporary interruption of study medication causing a switch to open-label prednisone will not be counted in these summaries. These summaries will be calculated based on the dosage of study medication prescribed by the investigator (i.e., not taking compliance into account).

A plot of average dosage prescribed by the investigator over time (averaged for each 3-month or 6-month visit interval beginning with the day of the first dose) will be provided for each treatment group. For each individual participant, the average dosage between two time intervals will be computed as the cumulative dosage during that interval divided by the number of days in the interval.

Deviations from the planned study medication dosage based on weight band will be listed and summarized by treatment group, including reasons for the deviation.

Descriptive statistics for the time that participants were exposed to open-label prednisone due to a temporary interruption of study medication will be provided by treatment group.

Compliance with study medication will be summarized by treatment group. Compliance will be quantified as the percentage of doses of study medication that were apparently taken (based on the number of pills dispensed compared to the number returned at each study visit) out of the number of doses expected to have been taken. Relevant circumstances such as dosage modifications will be taken into account in the calculation of the expected number of doses.

7.4. Analysis of the Primary Outcome Variable

7.4.1. General Approach

This trial is designed primarily to compare three corticosteroid regimens with regard to efficacy, as measured by a three-dimensional (multivariate) outcome consisting of the following three components (each averaged over all post-baseline follow-up visits through Month 36): (1) rise from the floor velocity; (2) forced vital capacity; and (3) participant/parent global satisfaction with treatment, as measured by the Treatment Satisfaction Questionnaire for Medication (TSQM). The primary statistical analysis will consist of global tests of the null hypothesis that the regimens do not differ with regard to any of the three outcomes against the alternative that they differ (in the same direction) for all three outcome variables, performed separately for each of the three pair-wise comparisons among the three corticosteroid regimens using a Bonferroni-corrected two-tailed significance level of 0.017.

For each pairwise comparison among the treatment groups, the overall strategy for the analysis of the multivariate outcome will be to analyze each of the outcomes separately, and then combine the p-values associated with the three outcomes using a two-sided version of the inverse normal combination statistic [Stouffer et al., 1949]. The null distribution of the statistic has a simple form when the p-values are independent, but in this case independence among the three outcomes (and, hence, the three p-values) cannot be assumed. Therefore, a re-randomization test will be used to compute the overall p-value [Rosenberger and Lachin, 2016]. This strategy will properly account for the dependence among the three p-values.

The p-values for the treatment group comparisons for two of the individual components of the primary outcome variable (rise from the floor velocity and FVC) will be derived from separate repeated measures analysis of covariance models (i.e., the mixed model repeated measures, or MMRM, analysis strategy) [Mallinckrodt, 2008], with terms for treatment group (three levels), country, baseline weight band (A vs. B/C), the baseline value of the outcome variable, time (treated as a categorical variable), and interaction terms for the baseline value of the outcome variable and time, and for treatment group and time. The p-values for the treatment group comparisons for the third component of the primary outcome variable, global satisfaction with treatment, will be derived from a similar model, but since this outcome variable is not measured at baseline, the model will instead include baseline values for rise from the floor velocity and FVC (with no interactions between these baseline variables and time). The covariance matrix for the within-participant observations will be modeled using an unstructured pattern. If the unstructured covariance matrix results in a lack of convergence, the following covariance structures will be used in sequence: heterogeneous Toeplitz, heterogeneous autoregressive, and heterogeneous compound symmetry. The Kenward-Roger approximation [Kenward and Roger, 2009] will be used to estimate the denominator degrees of freedom. The pair-wise treatment group differences will be estimated using appropriate contrasts that quantify the comparisons regarding the average value of the outcome variable over all post-baseline visits.

The overall test statistic will be the inverse normal combination statistic:

$$\psi_N = \sum_{i=1}^{3} \Phi^{-1}(1-p_i),$$

where p_1, p_2, and p_3 are the upper one-tailed p-values for the three outcome variables and $\Phi^{-1}(\square)$ is the inverse normal cumulative distribution function. This statistic is very similar to O'Brien's OLS statistic, i.e., the unweighted sum of the individual standardized test statistics for the three outcome variables. To see this, note that in the context of multivariate normal outcomes, asymptotically, O'Brien's OLS statistic is the sum of the Z-statistics for the three outcomes, rejecting the overall null hypothesis for either large positive values or large negative values of this statistic. Let

$$p_i = P_{H_0}\left(Z_i \geq z_i^*\right) = 1 - \Phi\left(z_i^*\right),\ i = 1,\ 2,\ 3,$$

where z_i^* is the observed value of Z_i. Then

$$\sum_{i=1}^{3} z_i^* = \sum_{i=1}^{3} \Phi^{-1}(1-p_i)$$

and the null hypothesis would be rejected for either large positive values or large negative values of this statistic.

The overall two-tailed p-value for comparing two treatment groups using the inverse normal combination statistic will be determined based on its re-randomization distribution. This will be estimated using a random sample of 15,000 repetitions of the randomization, each performed in exactly the same way as the original randomization was performed (permuted blocks, stratified by country).

It should be emphasized that interpretation of these tests will be in terms of the global three-dimensional outcome and that these tests do not identify the component on which the treatment groups differ. Follow-up analyses of the individual component outcomes, using a closed testing procedure [Marcus et al., 1976], will also be performed in order to aid in the interpretation of the results of the global test for the multivariate outcome. This procedure will involve performing treatment group comparisons with regard to the three possible bivariate outcomes (rise from the floor velocity and FVC, FVC and TSQM, rise from the floor velocity and TSQM) as well as the individual outcomes. A treatment group comparison for an individual outcome will be considered to be statistically significant (at the 0.017 level) only if it is also significant for the trivariate outcome as well as for the bivariate outcomes of which the individual outcome is a component.

Pair-wise treatment group differences in adjusted mean response, averaged over all post-baseline visits, will be summarized for the rise from the floor velocity, FVC, and global

satisfaction with treatment outcomes using 98.3% confidence intervals. The confidence coefficient of 98.3% incorporates a Bonferroni adjustment for the three pair-wise comparisons among the three treatment groups.

7.4.2. Adjustment for Baseline Characteristics

If clinically important differences are found among the three groups at baseline, particularly with regard to important variables such as age, the primary analyses may be repeated after statistically adjusting for these differences. These analyses, however, will be considered to be secondary.

7.4.3. Verification of Model Assumptions

The underlying assumptions of the repeated measures analysis of covariance models for the individual outcomes (normality, linearity) will be thoroughly checked. Secondary analyses that incorporate remedial measures (e.g., variable transformation) may be performed if warranted. The relatively large sample size (196 participants) should alleviate potential concerns about the normality assumption.

7.4.4. Primary Strategy for the Treatment of Missing Data

The primary analyses for the rise from the floor velocity, FVC, and global satisfaction with treatment outcomes will be performed using MMRM; hence, missing data will be appropriately accommodated under the missing at random assumption using direct likelihood [Mallinckrodt, 2008].

7.5. Analysis of the Secondary Outcome Variables for Efficacy

7.5.1. Analysis of Continuous Outcome Variables Obtained at Each Follow-up Visit

The secondary outcome variables for efficacy that were obtained at each follow-up visit and thought to be approximately normally distributed include 10-meter walk/run velocity, distance walked in 6 minutes (6MWT), NSAA total score, TSQM subscales (effectiveness, side effects, convenience), and range of motion at ankle dorsiflexion (left and right sides).

These outcome variables will be analyzed using repeated measures analysis of covariance models (i.e., the MMRM analysis strategy) with terms for treatment group (three levels), country, baseline weight band (A vs. B/C), the baseline value of the outcome variable, time (treated as a categorical variable), and interaction terms for the baseline value of the outcome variable and time, and for treatment group and time. Again, since the TSQM was not administered at baseline, the models for the TSQM subscales will include baseline values for rise from the floor velocity and FVC (with no interactions between these baseline variables and time) instead of the baseline value of the outcome variable. The covariance matrix for the within-participant observations will be modeled using an unstructured pattern. The pair-wise treatment group differences will be estimated using

appropriate contrasts that quantify the comparisons regarding the average value of the outcome variable over all post-baseline visits. Confidence intervals for pair-wise group differences in adjusted mean response at each visit will also be obtained using this model, but these will be of secondary interest.

7.5.2. Analysis of Continuous Outcome Variables Obtained at Annual Visits

The secondary outcome variables for efficacy that were measured at annual visits include lean mass (DXA) and quality of life scores (generic PedsQL and its neuromuscular disease-specific module). These will be analyzed using an analysis of covariance model with terms for treatment group (three levels), country, baseline weight band (A vs. B/C) (for the quality of life scores only), and the baseline value of the outcome variable. Due to several participants having no follow-up data for these outcome variables, the models will include further adjustment for baseline age and baseline NSAA total score. Separate analyses will be performed for outcomes at each annual visit. For each quality of life score, analysis of the average score across the annual visits will also be performed. Confidence intervals for pair-wise group differences in adjusted mean response will be obtained using these model.

Due to the relatively large percentage of participants with missing baseline data for these outcome variables, multiple imputation will be used to accommodate missing data. First, the monotone data augmentation method using Markov Chain Monte Carlo [Li, 1988; Liu, 1993] will be used to impute the small amount of missing data that are required to make the missing data pattern monotone. Next, for participants with complete data up to a particular visit, a multiple regression model will be fit that includes the outcome at that visit as the dependent variable and outcomes at previous visits, treatment group, country, baseline weight band (for the quality of life scores only), baseline age, and baseline NSAA total score as independent variables. Separate models will be similarly constructed for each visit. Using these regression models, a missing value for a participant at a particular visit will be imputed as a draw from the predictive distribution given the outcomes at previous visits (some possibly imputed), treatment group, and country. This will be done sequentially starting with the Month 12 visit. This process will be repeated 100 times, resulting in 100 complete analysis data sets. The analyses will be performed separately for each of the 100 complete analysis data sets and the results will be combined into one multiple imputation inference (estimated adjusted treatment group differences and p-values) using established methods [Schafer, 1997].

7.5.3. Examination of Treatment-by-Covariate Interactions

Analyses will be performed to explore potential interactions between treatment group and covariates with respect to selected outcome variables (averaged over all post-baseline visits). These outcome variables will include rise from the floor velocity, FVC, TSQM global satisfaction with treatment, 10-meter walk/run velocity, distance walked in 6 minutes (6MWT), and NSAA total score.

The interaction between treatment group and country/region will be investigated by including the appropriate interaction term in the analysis of covariance model and testing for its significance. Interactions between treatment group and the following baseline variables will also be explored: age group (< 6, ≥ 6 years), ethnicity (hispanic, non-hispanic), time to rise from the floor (< 5, 5-8, > 8 seconds), NSAA total score (< 20, 20-25, < 25), and initial weight band (A, B/C) will be examined separately in a similar fashion.

Since the power to detect potentially meaningful interactions will be limited, the magnitudes of mean responses to treatment in the relevant subgroups will be examined. The observation of clinically important subgroup differences in mean treatment response (e.g., age < 6 vs. age ≥ 6) will serve as hypothesis generation for possible future studies designed to specifically address the issue of differential therapeutic response.

7.5.4. Treatment of Missing Data for Secondary Outcome Variables

As noted in Section 7.4.4 above, for the secondary outcomes obtained at each follow-up visit, missing data will be appropriately accommodated under the missing at random assumption using direct likelihood [Mallinckrodt, 2008]. Multiple imputation will be used instead, as noted in Section 7.5.2 above, for the secondary outcome variables obtained at annual visits due to the elevated frequency of missing baseline values.

7.5.5. Analyses of Individual NSAA Items

The individual items on the NSAA are generally scored as 0 = Unable to achieve independently, 1 = Able to achieve independently with a modified method, and 2 = Able to achieve normally. Although DMD is a progressive disease, most of these item scores can transition in either direction (i.e., improvement or decline) in the context of corticosteroid treatment. Therefore, multi-state models for panel data [Kalbfleisch and Lawless, 1985; Jackson, 2011] will be used to compare the treatment groups with respect to the patterns of transition between the different states (i.e., NSAA scores of "0", "1", or "2").

Multi-state models assume that the transitions between the different states occur in continuous time, even though the data are observed only at discrete visits. In these models, therefore, it will be assumed that one can only transition between adjacent states. For example, one cannot transition directly from a state (score) of "2" to a state of "0" without having transitioned through a state of "1". The models also employ a Markov assumption, i.e., that the probability of transitioning to a particular state at any time depends only on the current state and not on the observation process leading up to that state. It will also be assumed that the transition probabilities remain constant over time (time-homogeneous model), though this assumption can be relaxed if needed.

The models will include treatment group, country, age, and initial weight band (A, B/C) as independent variables. The transition intensity matrices for each treatment group (for selected values of country, age, and initial weight band) will be presented, and treatment

group differences regarding the transition between different states will be summarized with hazard ratios and their associated 98.3% confidence intervals under a proportional intensity assumption. The models will also allow estimation of the mean total time in each state (out of 36 months) for each treatment group, with bootstrap confidence intervals, for selected values of country, age, and initial weight band.

NSAA items that are assumed to only progress and not improve include rising from the floor and walking. Times from the baseline visit to loss of the ability to rise from the floor and to loss of independent ambulation will be primarily analyzed using a Cox proportional hazards model that will include treatment group as the factor of interest and age and initial weight band as covariates. Confidence intervals (98.3%) for the adjusted hazard ratios representing pair-wise treatment group comparisons will be computed using this model.

Event times will be censored at the last follow-up time for participants who do not reach the endpoint (milestone). Kaplan-Meier curves will be used to describe the cumulative probabilities of reaching the endpoint over time in the three treatment groups.

7.6. Analysis of Tolerability and Safety Outcomes

7.6.1. Tolerability Outcomes

Tolerability outcomes will include ability to complete the trial on the originally assigned dosage (for weight) of study medication, ability to complete the trial without switching drug regimens (i.e., discontinuing study medication), and ability to complete the trial. The proportions of participants with these outcomes will be compared among the treatment groups using Fisher's exact tests.

7.6.2. Adverse Events

A treatment-emergent adverse event is defined as an adverse event with an onset date on or after the start of dosing of study medication. Analyses of adverse events will only be performed for treatment-emergent adverse events. Adverse events that occurred prior to the start of dosing of study medication will only be included in the listings specified below.

The frequencies of individual adverse events will be tabulated by treatment group, body system, and preferred term. For each preferred term, the number (%) of participants experiencing the event and the total number of events will be presented. For each adverse event, pair-wise comparisons among the treatment groups will be performed regarding the occurrence of at least one event using Fisher's exact tests. The comparisons will be repeated excluding all mild events; if a participant had repeat occurrences of the same event, the event with the highest severity will be considered for these comparisons. Similar analyses will be performed at the level of body system rather than preferred term.

Listings of individual adverse events will be produced, including treatment group, participant ID number, site, preferred term, days after randomization, event duration (days), perceived relationship to study medication, action taken, expected (yes/no), severity, serious (yes/no), and outcome. Similar listings of individual serious adverse events and adverse events leading to a modification in dosage will also be produced.

7.6.3. Additional Safety Outcomes

Continuous measures of safety including height, weight, BMI, vital signs, DXA outcomes, bone age, difference between chronological age and bone age, echocardiography outcomes, and behavioral scale scores will be analyzed using the methods described in Section 7.5.1 or Section 7.5.2 above, depending on the completeness of the baseline information for each outcome variable. Dichotomous Iowa Conners outcomes (I/O score ≥ 10, O/D score ≥ 9) and PARS-III total score < 72 will be tabulated by treatment group and visit.

The frequencies of individual laboratory abnormalities (high and low separately) will be tabulated by treatment group. For each laboratory test, the number (%) of participants experiencing the abnormality and the total number of abnormal tests will be presented. For each laboratory test, pair-wise comparisons among the treatment groups will be performed regarding the occurrence of at least one abnormality using Fisher's exact tests. The comparisons will be repeated excluding all abnormalities that are not deemed clinically significant; if a participant had repeat occurrences of the same abnormality, the abnormality with the highest level of clinical significance will be considered for these comparisons.

Results of laboratory tests (actual values and changes from baseline) will be summarized by treatment group and visit using descriptive statistics.

Listings of individual laboratory abnormalities will be produced, including treatment group, participant ID number, site, laboratory test, days after randomization, value, and clinical significance.

Echocardiography and ECG abnormalities will be summarized in a similar manner as the laboratory tests.

7.7. Miscellaneous Analyses and Outcomes

7.7.1. Concomitant Medications

Concomitant medication use will be summarized by treatment group according to the percentages of participants using particular medications (or classes of medications) at or after the start of dosing of study medication.

A listing of concomitant medications will be provided, including participant ID number, site, treatment group, medication, dosage, frequency, form, indication, start date, and stop date.

7.7.2. Blindedness Assessment

Guessed treatment assignment from the Blindedness Assessment will be tabulated by actual treatment assignment separately for the parents/guardians, site investigators, and site clinical evaluators/coordinators. Chi-square tests will be used to determine whether or not the distributions of the guessed treatments of the parents/guardians (or site investigators, or site clinical evaluators/coordinators) differ among the actual treatment groups. The certainty of the guesses and the primary and secondary reasons for the guesses will also be summarized by guessed treatment/actual treatment combination.

7.7.3. Protocol Deviations

A listing of protocol deviations will be provided, including participant ID number, site, treatment group, month, date of the deviation, nature of the deviation, reason for the deviation, and comments.

8. CHANGES FROM THE STUDY PROTOCOL

The following modifications were made to the plans for statistical analysis specified in the protocol (all modifications were made prior to unblinding):

- The analysis plan for the primary outcome variable was modified due to a greater number of participants than expected who lost the ability to rise from the floor during the trial. The protocol states that participants who do not have the ability to rise from the floor at a particular visit will be assigned the largest observed time in the sample for purposes of analysis. Apart from the arbitrary nature of the assignment of the largest observed time in the sample to these incompletely ascertained outcomes, this strategy also renders the multivariate normality assumption somewhat questionable, both for the statistical model and for the multiple imputation strategy that was proposed to accommodate missing data.

  The analysis plan contains two substantial modifications: (1) use of the reciprocal transformation of the time to rise from the floor, expressing the outcome as the rise from the floor velocity, and (2) analyzing each of the components of the three-dimensional outcome variable separately and then combining the resulting p-values rather than using O'Brien's OLS test. The latter modification yields a test based on re-randomization that is similar to O'Brien's OLS test but does not rely on the assumption of multivariate normality (though it does assume normality for the individual outcome variables).

- The analysis cohorts for the primary and secondary outcome variables were specified in greater detail, with the principal changes due to (1) a small number of participants who contributed no post-baseline data for the primary outcome variable and (2) the focus on the trial having pragmatic aims. With respect to the latter point, the "per-protocol" analyses mentioned in the protocol have been eliminated.

- The analysis of the time to walk/run 10 meters was modified to use the reciprocal transformation (10-meter walk/run velocity) to accommodate the loss of the ability of some participants to perform this task.

- The analysis plan was modified for the secondary efficacy outcome variables collected at annual visits due to (1) the number of participants who had missing data at baseline (necessitating the use of multiple imputation) and (2) the number of participants who had no post-baseline data (necessitating the adjustment for additional covariates).

- The analysis plan for the individual NSAA items was modified in recognition of the fact that the scores on these items could either improve or worsen in the context of corticosteroid treatment and disease progression. The analyses of times to disease milestones were restricted to time to loss of independent ambulation and time to loss of the ability to rise from the floor; with very rare exception, these abilities were not regained once lost. Due to the relatively low incidence of these milestone events, the plan to stratify by country in the analyses was eliminated.

9. REFERENCES

Atkinson MJ, Sinha A, Hass SL, Colman SS, Kumar RN, Brod M, Rowland CR. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health Qual Life Outcomes* 2004; 2:12.

Fenichel GM, Florence JM, Pestronk A, Mendell JR, Moxley RT III, Griggs RC, Brooke MH, Miller JP, Robison J, King W, Signore L, Pandya S, Schierbecker J, Wilson B. Long-term benefit from prednisone therapy in Duchenne muscular dystrophy. *Neurology* 1991; 41:1874-1877.

Griggs RC, Moxley RT III, Mendell JR, Fenichel GM, Brooke MH, Pestronk A, Miller JP. Prednisone in Duchenne dystrophy. A randomized, controlled trial defining the time course and dose response. *Arch Neurol* 1991; 48:383-388.

Herr BE, Hart KA, McDermott M, Griggs RC, Herbelin L, Barohn R. Surrogate completion of the Treatment Satisfaction Questionnaire for Medication for a trial of corticosteroids in Duchenne muscular dystrophy. *Neurother* 2008; 5:490.

Jackson CH. Multi-state models for panel data: the msm package for R. *J Statist Softw* 2011; 38:1-28.

Kalbfleisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *J Am Statist Assoc* 1985; 80:863-871.

Kenward MG, Roger JH. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Comput Statist Data Anal* 2009; 53:2583–2595.

Li KH. Imputation using Markov chains. *J Statist Comp Simul* 1988; 30:57-79.

Liu C. Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data. *J Mult Anal* 1993; 46:198-206.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf J* 2008; 42:303-319.

Manzur AY, Kuntzer T, Pike M, Swan A. Glucocorticoid corticosteroids for Duchenne muscular dystrophy. *Cochrane Review* 2004; 2:1-55.

Marcus R, Peritz E, Gabriel KR. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 1976; 63:655-660.

Mendell JR, Moxley RT, Griggs RC, Brooke MH, Fenichel GM, Miller JP, King W, Signore L, Pandya S, Florence J, Schierbecker J, Robison J, Kaiser K, Mandel S, Arfken C, Gilder B. Randomized, double-blind six-month trial of prednisone in Duchenne's muscular dystrophy. *N Engl J Med* 1989; 320:1592-1597.

O'Brien PC. Procedures for comparing samples with multiple endpoints. *Biometrics* 1984; 40:1079-1087.

Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*, Second Edition. Hoboken, NJ: John Wiley and Sons, 2016.

Schafer JL. *Analysis of Incomplete Multivariate Data*. Boca Raton: Chapman and Hall/CRC, 1997.

Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams RM. *The American Soldier, Volume 1: Adjustment during Army Life*. Princeton, NJ: Princeton University Press, 1949.

APPENDIX: DATABASE MODIFICATIONS FOR ANALYSIS

Imputation of Missing Scale Items

Missing total scores on the NSAA, Iowa Conners, and Revised Rutter scales due to a small number of missing items were imputed as follows:

1. Imputing the missing item scores by carrying forward the item scores from the previous visit or, if the item scores were missing at the screening/baseline visit, by carrying backwards the item scores from the subsequent visit. For the NSAA, item imputation was also based on the ambulatory status of the participant in some cases.

2. Recomputing the total score using the observed and imputed item scores.

NSAA Scale

Participant ID	Month	Change
CAN04R01	36	Item 5: Changed to "1"; Item 11 Changed to "0"
GBR01R02	6	Item 17: Changed to "1"
GBR02R09	36	Item 5: Changed to "1"
GBR07R19	24	Item 11: Changed to "0"
GBR09R01	SC	Item 13: Changed to "1"
GBR09R01	0	Items 8 and 9: Changed to "2"
USA01R11	3	Items 11 and 17: Changed to "1"
USA01R12	24	Item 11: Changed to "1"
USA03R02	18	Items 11 and 17: Changed to "0"
USA03R02	24	Items 11 and 17: Changed to "0"
USA03R02	30	Items 11 and 17: Changed to "0"
USA03R02	36	All items except for Item 10: Changed to "0"
USA05R13	30	Items 10 and 14: Changed to "2"
USA05R13	36	Item 3: Changed to "1"
USA07R01	SC	Item 10: Changed to "1"
USA07R05	24	Item 17: Changed to "1"
USA09R03	30	Item 13: Changed to "0"
USA10R10	3	Item 12: Changed to "1"

Iowa Conners Scale

Participant ID	Month	Change
CAN01R03	12	Item 9: Changed to "0"
CAN01R03	36	Item 6: Changed to "0"
DEU04R03	12	Item 4: Changed to "0"
GBR03R18	6	Item 10: Changed to "1"

GBR03R18	24	Item 3: Changed to "1"; Item 10: Changed to "0"
ITA04R06	0	Item 5: Changed to "0"
ITA04R06	36	Item 5: Changed to "0"
ITA06R09	18	Item 7: Changed to "0"
USA01R01	0	Item 2: Changed to "1"
USA01R04	36	Item 3: Changed to "2"
USA01R06	3	Item 8: Changed to "0"
USA01R06	12	Item 1: Changed to "2"; Item 3: Changed to "0"
USA01R06	30	Item 1: Changed to "1"
USA01R10	3	Item 7: Changed to "0"
USA02R07	0	Item 2: Changed to "1"
USA03R02	30	Item 1: Changed to "1"
USA10R01	36	Item 7: Changed to "0"
USA10R06	0	Item 7: Changed to "2"
USA10R06	24	Item 7: Changed to "2"
USA10R07	0	Item 7: Changed to "1"
USA11R03	3	Item 10: Changed to "1"

Revised Rutter Scale

Participant ID	Month	Change
CAN01R01	36	Item 4: Changed to "0"
DEU01R08	6	Item 7: Changed to "1"
DEU03R05	0	Item 1: Changed to "0"
GBR06R09	18	Item 4: Changed to "0"
GBR09R04	24	Item 6: Changed to "1"
USA01R06	30	Item 1: Changed to "0"
USA02R02	0	Item 5: Changed to "1"
USA05R09	6	Item 4: Changed to "1"
USA06R01	36	Item 7: Changed to "0"
USA07R03	6	Item 8: Changed to "1"
USA10R06	12	Item 6: Changed to "0"
USA14R11	0	Item 4: Changed to "0"

Use of Efficacy Outcome Data from Unscheduled or Shifted Visits

Participant ID	Month	Change
DEU06R01	36	Use ECHO/ECG data from Month 30 visit
GBR06R01	24	Use 10-meter walk/run time and six-minute walk test data from unscheduled visit on 12/01/15
GBR06R06	3	Use NSAA, range of motion at ankle dorsiflexion, and six-minute walk test data from unscheduled visit on 11/04/14
GBR06R06	12	Use data from unscheduled visit on 04/17/15

Imputation of Missing NSAA and Six-Minute Walk Test Data Due to Loss of Ability to Perform the Task

For the NSAA timed tasks (rise from the floor and walk/run 10 meters), a time of "999" is recorded in the database if the corresponding NSAA item (11 or 17) is scored as "0" and the time was entered as either missing or "0". There were several instances where there were missing NSAA total scores and times as well as six-minute walk test distances, but it was clear from the comments recorded on the corresponding case report forms that the participant no longer had the ability to perform the tasks.

Participant ID	Months	Change
CAN01R03	24	Six-minute walk test distance set to "0"
CAN01R03	30, 36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
CAN02R05	30	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
CAN02R05	24, 36	Six-minute walk test distance set to "0"
DEU06R06	36	Six-minute walk test distance set to "0"
GBR03R09	36	Six-minute walk test distance set to "0"
GBR07R19	24	Time to rise from the floor set to "999"
ITA03R02	30	Six-minute walk test distance set to "0"
ITA03R02	36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA02R05	18	Six-minute walk test distance set to "0"
USA02R05	24, 30, 36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA03R02	18, 24, 30, 36	NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA06R01	36	Six-minute walk test distance set to "0"
USA09R04	24, 30, 36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA10R07	24, 30	Six-minute walk test distance set to "0"
USA10R07	36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA10R08	36	Six-minute walk test distance set to "0"
USA10R10	36	Six-minute walk test distance set to "0"
USA11R01	36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"
USA15R01	24	Six-minute walk test distance set to "0"
USA15R01	36	All NSAA items set to "0"; NSAA timed tasks set to "999"; six-minute walk test distance set to "0"

Six-minute Walk Test Results Deemed Invalid

The following six-minute walk test distances were set to missing due to the evaluator deeming the results invalid on the case report form.

Participant ID	Months	Participant ID	Months	Participant ID	Months
CAN01R04	6	GBR07R11	30	USA09R05	3
CAN02R06	0	GBR07R15	6	USA10R01	SC, 0, 6, 12, 18, 24, 30, 36
CAN02R11	0	USA01R04	12	USA10R03	6
CAN04R01	0, 6, 18	USA01R07	3	USA10R04	SC, 0, 3, 12, 30
GBR01R10	SC	USA02R09	24	USA10R05	0
GBR02R11	24	USA07R01	3, 6, 12	USA11R02	18
GBR03R03	36	USA07R02	SC	USA14R11	12
GBR03R16	36	USA07R04	0	USA15R09	3
GBR07R01	3	USA07R05	24	USA18R01	0, 6
GBR07R04	SC	USA09R01	SC, 0, 6, 18, 30, 36		

Imputation of Missing Baseline NSAA Timed Tasks

One four year-old participant (USA18R01) did not have the NSAA timed tasks performed at either the Screening visit or the Baseline visit due to issues with cooperation. These values were singly imputed using a regression model for the relationship between the baseline value (Y) and the Month 3 value (X) for participants with data at both time points.

Values Deemed Implausible

The following values were deemed implausible and set to missing.

Participant ID	Month	Change
GBR06R05	36	Six-minute walk test distance set to missing
GBR06R09	18	Ten-meter walk/run time set to missing
USA11R03	30	Forced vital capacity set to missing

Missing Medication Start Dates

Participant ID	Change
GBR06R06	Missing medication start date imputed as one day after the medication ship date
GBR06R09	Missing medication start date imputed as one day after the medication ship date
GBR06R10	Missing medication start date imputed as one day after the medication ship date

Height Values Estimated from Ulna Length

Participant heights at the following visits were estimated from ulna length; therefore, height, height percentile, height Z-score, and body mass index were set to missing.

Participant ID	Months
CAN01R03	30
USA01R08	30, 36
USA02R05	36
USA03R02	18, 24, 30
USA09R03	SC
USA09R04	18, 24, 30, 36
USA10R07	30