

Supplemental Online Content

Mah JK, Clemens PR, Guglieri M, et al; NorthStar UK Network and CINRG DNHS Investigators. Efficacy and safety of vamorolone in Duchenne muscular dystrophy: a 30-month nonrandomized controlled open-label extension trial. *JAMA Netw Open*. 2022;5(1):e2144178. doi:10.1001/jamanetworkopen.2021.44178

eTable 1. Comparison of Coarsened Exact Matched Participants in Duchenne Natural History Study (DNHS) and Higher-Dose VBP15 Long-term Extension (LTE) Study

eTable 2. Pediatric Outcomes Data Collection Instrument (PODCI) Scores

eTable 3. Quantitative Muscle Testing (QMT) Scores

eTable 4. Adverse Events by Dose at Time of Event From 30 mo of Vamorolone Treatment

eFigure 1. Patient-Level Dose Regimen Schematic

eFigure 2. Timed Motor Function Comparisons of Vamorolone Long-term Extension vs Duchenne Natural History Study With Individual Trajectories

eFigure 3. Body Mass Index z Score and Height Percentile Comparisons of Vamorolone Long-term Extension vs Duchenne Natural History Study With Individual Trajectories

eAppendix 1. Duchenne Natural History Study Cohort External Comparison Workflow Plan

eAppendix 2. NorthStar United Kingdom Network Cohort External Comparison Workflow Plan

eAppendix 3. Supplemental Text on Statistical Analysis

This supplemental material has been provided by the authors to give readers additional information about their work.

eTable 1. Comparison of Coarsened Exact Matched Participants in Duchenne Natural History Study (DNHS) and Higher-Dose VBP15 Long-term Extension (LTE) Study

Baseline time to stand (seconds)^a	CEM-matched DNHS participants (n=29)	CEM-matched High-dose VBP15-LTE participants (n=20)	CEM-matched DNHS participants with a 24-month follow-up visit (n=10)	CEM-matched High-dose VBP15-LTE participants with a 24-month follow-up visit (n=18)
Minimum	2.50	2.50	2.62	2.74
1 st quartile	3.59	3.70	3.02	4.10
Median	4.07	4.31	3.66	4.41
Mean	4.03	4.20	3.62	4.36
3 rd quartile	4.50	4.63	4.22	4.68
Maximum	5.89	5.50	4.56	5.50

^a CEM-matched DNHS and LTE participants were well balanced at baseline. However, there was substantial attrition for the matched DNHS subjects (only 10/29=34.5% subjects had a 24-month follow-up visit), which would affect longitudinal modeling. Furthermore, between CEM-matched DNHS participants with long-term follow-up (n=10) and CEM-matched higher-dose VBP15-LTE participants with long-term follow-up (18/20=90% of enrolled), VBP15-LTE participants are consistently slower at baseline, and hence, baseline characteristics are not well balanced for longitudinal comparisons.

eTable 2. Pediatric Outcomes Data Collection Instrument (PODCI) Scores

Patient-reported outcomes and health-related quality of life	VBP15-LTE baseline^a (SD)	24 months from VBP15-LTE baseline (SD)	Change from baseline (SD), Paired t-test p-value; (95% Two-sided CI)
PODCI upper extremity and physical function	75.34 (15.09), n=18	82.32 (10.91), n=18	+ 7.95 (12.55), p=0.028; (1.00, 14.90), n=15
PODCI transfer and basic mobility	86.55 (9.21), n=19	81.44 (17.54), n=18	- 4.38 (22.68), p=0.452; (-16.47, 7.70), n=16

^a Based on participants assigned to higher-dose vamorolone (2.0 and 6.0 mg/kg/day) at start of study

eTable 3. Quantitative Muscle Testing (QMT) Scores

Quantitative Muscle Testing (pounds)	VBP15-002^a baseline (SD)	30 months from VBP15-002 baseline (SD)
QMT Knee extension	11.92 (5.29), n=23	15.37 (12.93), n=5
QMT Knee flexion	7.55 (4.36), n=23	13.43 (6.63), n=5
QMT Elbow extension	5.44 (2.00), n=23	6.13 (3.10), n=5
QMT Elbow flexion	6.36 (2.27), n=23	7.49 (3.61), n=5

^a Based on participants assigned to higher-dose vamorolone (2.0 and 6.0 mg/kg/day) at start of study

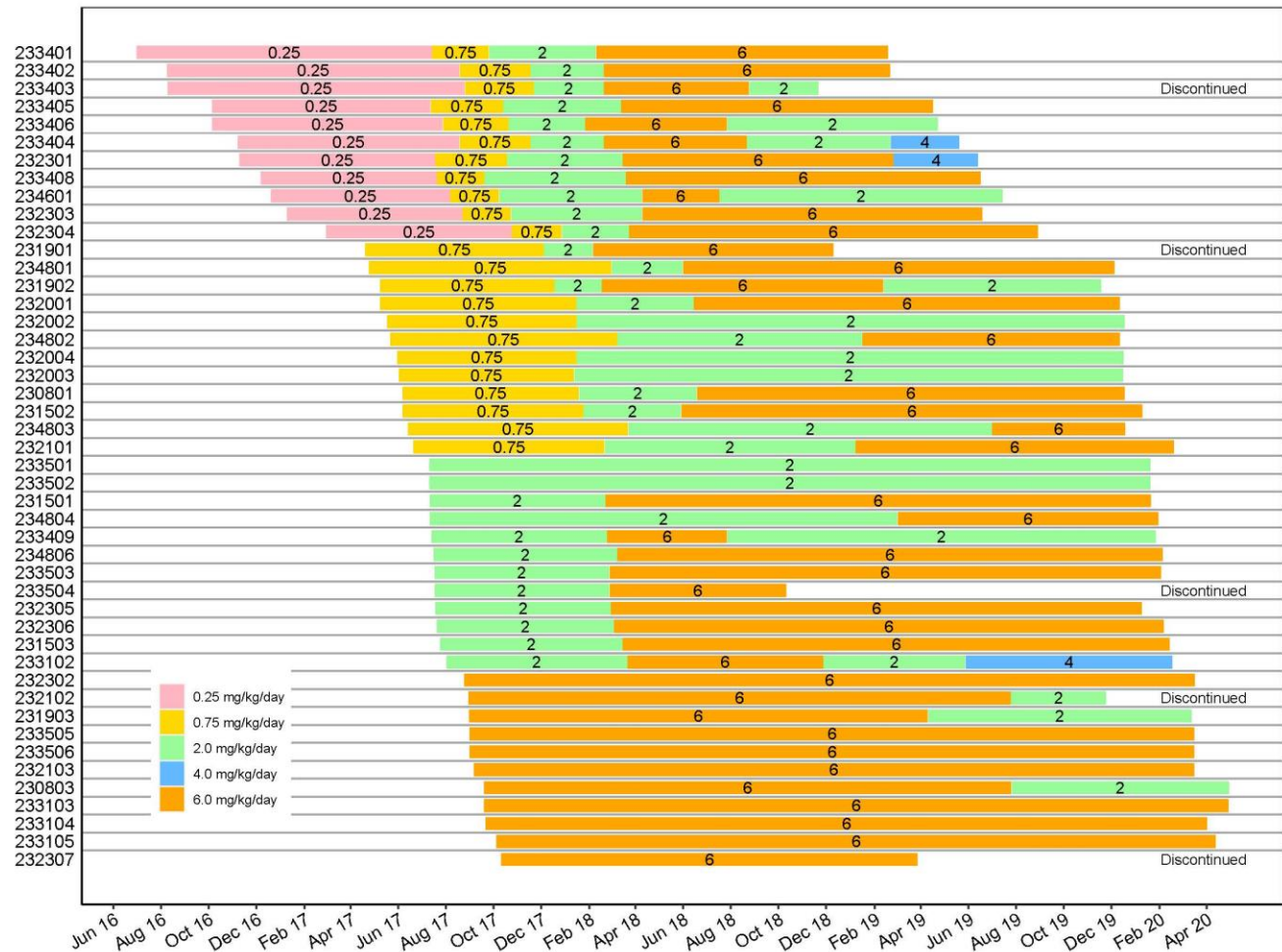
eTable 4. Adverse Events by Dose at Time of Event From 30 mo of Vamorolone Treatment

	Vamorolone dose at the time of the event (mg/kg/day)				
	0.25 mg/kg, N = 11 (%)	0.75 mg/kg, N = 23 (%)	2.0 mg/kg, N = 38 (%)	4.0 mg/kg, N = 3 (%)	6.0 mg/kg, N = 41 (%)
Overall					
Any TEAE	4 (36.4)	14 (60.9)	29 (76.3)	1 (33.3)	39 (95.1)
Any Treatment-related TEAE ^a	0	0	8 (21.1)	1 (33.3)	23 (56.1)
Any TEAE with CTCAE Grade ≥ 3	0	1 (4.3)	0	0	1 (2.4)
Any TEAE Leading to Discontinuation of Study	0	0	1 (2.6)	0	0
Any SAE ^a	0	1 (4.3)	0	0	1 (2.4)
Any Serious TEAE	0	1 (4.3)	0	0	1 (2.4)
Preferred Term					
Nasopharyngitis	0	1 (4.3)	7 (18.4)	0	12 (29.3)
Cough	1 (9.1)	1 (4.3)	7 (18.4)	0	8 (19.5)
Pyrexia	0	0	9 (23.7)	0	8 (19.5)
Vomiting	1 (9.1)	0	5 (13.2)	1 (33.3)	7 (17.1)
Weight increased	0	0	2 (5.3)	0	10 (24.4)
Constipation	0	0	3 (7.9)	0	8 (19.5)
Pain in extremity	0	3 (13.0)	5 (13.2)	0	6 (14.6)
Fall	0	1 (4.3)	4 (10.5)	0	6 (14.6)
Upper respiratory tract infection	0	2 (8.7)	5 (13.2)	0	6 (14.6)
Headache	0	0	4 (10.5)	0	7 (17.1)
Influenza	0	1 (4.3)	3 (7.9)	0	6 (14.6)

Abbreviations: TEAE, Treatment emergent adverse events; SAE: serious adverse events; CTCAE: Common Terminology Criteria for Adverse Events

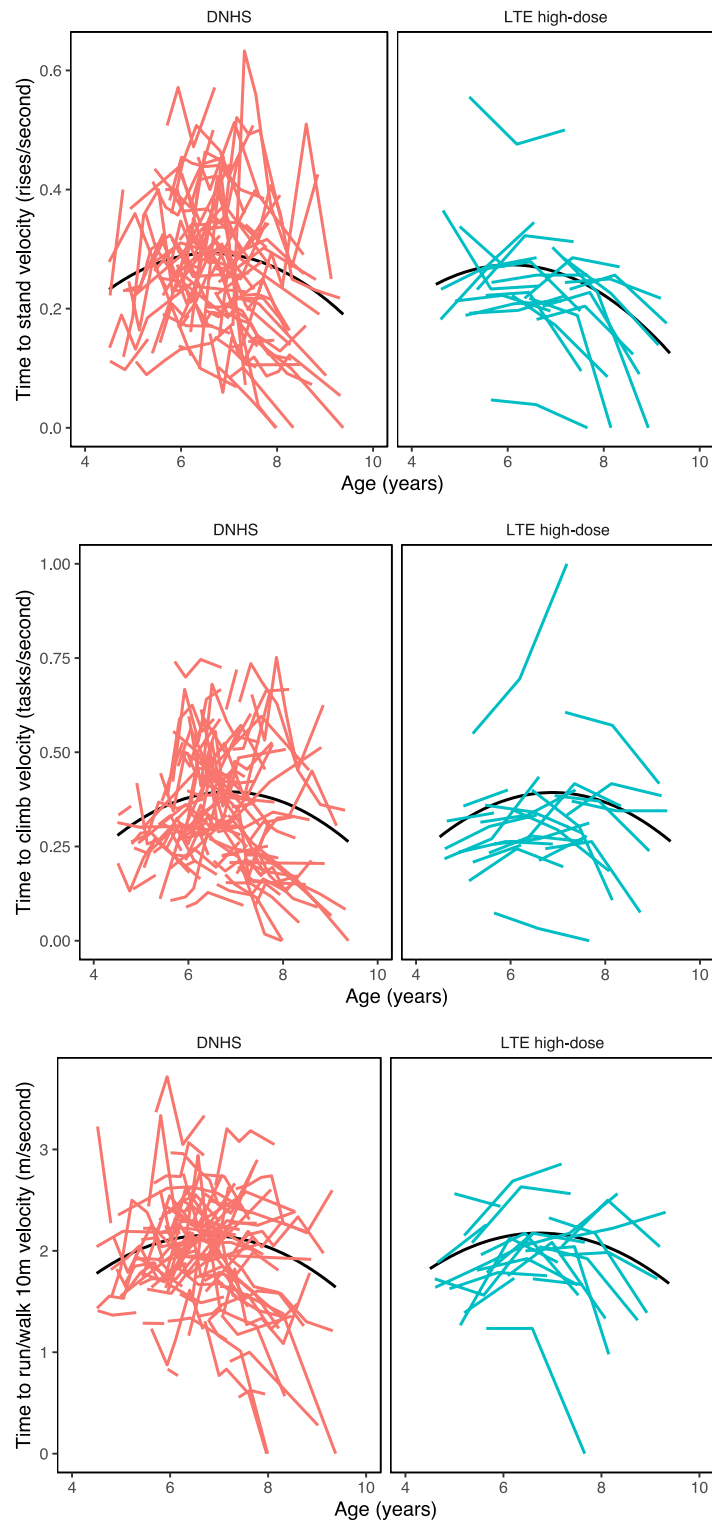
^aNo TEAEs led to discontinuation of study drug. No deaths occurred.

eFigure 1. Patient-Level Dose Regimen Schematic



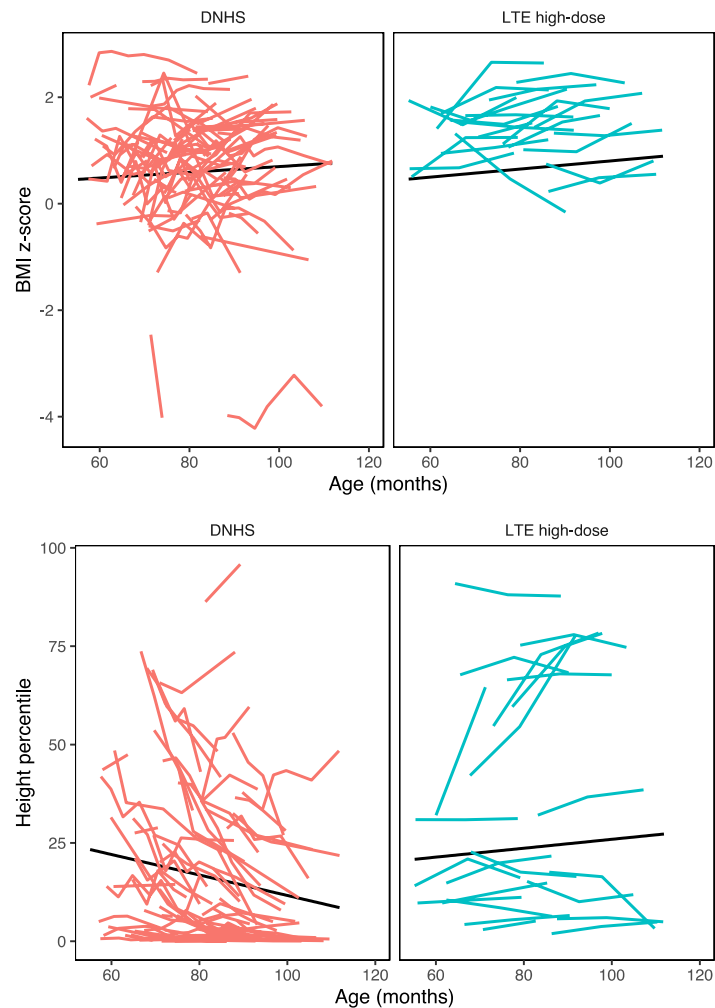
Each horizontal line indicates a participant in VBP15-LTE. The x-axis is date (month year format). The participants started out in a dose-ranging study, with each group (y-axis) of 12 participants started at a specific dose (doses indicated by colors and legend; 0.25 mg/kg/day, 0.75 mg/kg/day, 2.0 mg/kg/day, 6.0 mg/kg/day). VBP15-LTE was a 2-year long-term extension study. Dose escalations and dose de-escalations were permitted in VBP15-LTE at the discretion of the treating physician and the participant's family; dose changes are indicated for each participant by the change in color. Three subjects also were on 4.0 mg/kg/day; the duration of this is indicated by the blue boxes. Participants who discontinued early are annotated as such accordingly.

eFigure 2. Timed Motor Function Comparisons of Vamorolone Long-term Extension vs Duchenne Natural History Study With Individual Trajectories^a



^a Predicted trajectories (in black) are plotted for a subject with the same baseline velocity and age. None of the between cohort comparisons are significant at $\alpha=0.05$.

eFigure 3. Body Mass Index z Score and Height Percentile Comparisons of Vamorolone Long-term Extension vs Duchenne Natural History Study With Individual Trajectories^a



^a Predicted trajectories (in black) are plotted for a subject with the same baseline values. For BMI z-score, the between cohort comparisons is not significant at $\alpha=0.05$ while for height, the comparison was significant ($p=8.94 \times 10^{-07}$).

eAppendix 1. Duchenne Natural History Study Cohort External Comparison Workflow Plan

Version date: Jan 22, 2021

Operating Procedure for Data Extraction, Matching, and Analysis

Workflow for data extraction programming:

1. Two statisticians independent of the sponsor, Dr. Heather Gordish-Dressman and Dr. Utkarsh Dang, who have access to the entire CINRG DNHS dataset from TRINDS, will perform data extraction independently of each other, based on the criteria listed below.
2. Dr. Gordish-Dressman will use *Stata* while Dr. Dang will use *R*.
3. After independently extracting data, the two datasets will be compared using code. If there are any inconsistencies, the issue will be identified and the extraction run again.
4. Number 3 step will be repeated until both independent data extractions lead to the same subset of data, consistent with the criteria mentioned below. This is referred to herein as the criteria-matched data.

Specifics for data extraction of DMD subjects from CINRG DNHS database at TRINDS:

Note that for our comparisons, we will primarily rely on the subgroup of VBP15-LTE subjects who had 6 months of high-dose (2.0/6.0 mg/kg/day combined) exposure at baseline. Then, DMD subjects fulfilling the following general criteria will be extracted from CINRG DNHS data, on par with enrollment into VBP15-LTE (6 months vamorolone exposure, 4.5 to 7.5 years at baseline).

1. Subjects between age 4.5 and <7.5 years old at baseline visit (baseline for this data extraction).
2. Subjects who initiated corticosteroids (prednisone, deflazacort; any dose) approximately 6 months prior to baseline visit and were maintained on corticosteroids for subsequent visits (any steroid regimen is OK but no steroid-naïve visit data). Subjects should have at least two visits including baseline visit (which corresponds to approximately 6 months of steroid exposure) and a maximum follow-up of 24 months included in this comparator data.

Specifics for programming: The lifetime steroid exposure (as well as the current steroid exposure), defined as continuous steroid exposure, at baseline should be 5-12 months. Follow-up with steroid exposure after the baseline visit (~6 months of steroid exposure) should be at least 10 months (approximately 1 year). For programming purposes, there is a three-month additional window around the maximum of 24 months exposure, i.e., maximum of 27 months maximum follow-up.

- a. Data obtained beyond the 2 years of follow-up interesting for comparisons will be used for overall trajectory modeling and visualization.
3. Subjects from all regions.
 4. Subjects not co-enrolled in other exon-skipping clinical trials.

Specifics for data extraction of steroid-naïve DMD subjects from CINRG DNHS database at TRINDS:

Steroid-naïve subjects who at baseline would have between 4 and 7.5 years of age who have at least two visits of data. All visits with continued steroid-naïve status will be obtained.

Specifics for data extraction of healthy controls from CINRG DNHS database at TRINDS:

Dr. Gordish-Dressman expects some healthy peer controls between 4.0 and 7.5 years and have a follow-up visit at 12 months (only 2 visits possible: baseline and 12 months). Timed function tests and anthropometrics are available.

Specifics for statistical matching: Dr. Dang will work on statistical matching.

The CINRG DNHS data extracted and the VBP15-LTE data (initial dose groups: 2.0 or 6.0 mg/kg/day) extracted will be matched using coarsened exact matching (Iacus et al. 2011). This matching will yield subsets of the comparator cohorts with reduced imbalance between the baseline variables. Note that this is data pre-processing, not parameter estimation. The process works as follows on the baseline (prior to relevant outcomes) data (Iacus et al. 2020):

a) coarsen each baseline variable of interest *temporarily* (no information is lost by the end of the procedure),

b) sort all data points into strata with the same values of coarsened variables,

c) prune observations that have no close matches on coarsened baseline variables (exact matching on coarsened data) in the DNHS steroid treated and vamorolone treated cohorts, i.e., a stratum needs at least one DNHS steroid-treated and one vamorolone-treated observation and

d) retain all remaining matched subject IDs and use longitudinal data available on these subjects in analysis.

1. The baseline variables on the basis of which matching will be conducted are the following: time to stand in seconds and time to run/walk 10 meters in seconds as well as age.
2. No ratio of retained vamorolone-treated to DNHS steroid-treated subjects is pre-specified with the goal of retaining as many matched observations as possible.
3. Baseline time to stand and time to run/walk here is defined as the average of 003 final timed function test assessment and the LTE baseline timed function test assessment. This is consistent with six months of vamorolone exposure at baseline.
4. The degree of coarsening is variable-specific:
 - a. Age at baseline: Less than 5.5 years of age vs. greater than or equal to 5.5 years of age.
 - b. Baseline time to stand in seconds: bins of 0.5 seconds, i.e., strata will be defined on the basis of for example, 3.5 to 4, 4 to 4.5, 4.5 to 5.0 seconds, and so on.
 - c. Baseline time to run/walk 10 meters in seconds: bins of 0.75 seconds.
 - d. The above pre-specified coarsening procedure will be compared to an algorithmic automatic coarsening procedure on data in seconds as well as based on velocity and the subset of matched observations from either of the analysis will be picked according to the best tradeoff of imbalance and observations retained.
5. Statistical tests will be conducted on the baseline variables to show reduction of imbalance as compared to the criteria-matched data.
6. The retained matched subjects will be carried forward for analysis, referred to as the CEM-matched subjects. For subsequent analysis, a sensitivity analysis will be conducted with the criteria-matched data as well.

Plan for comparison to steroid-treated subjects:

For statistical considerations, similar criteria will be adopted as in the main SAP (NSAA score if all components available, velocity calculations, zero imputation, etc.). Both efficacy and safety

comparative comparisons will be conducted. Missing data will not be imputed. Any techniques that can account for some missing data (e.g., MMRM with missing at random assumption or survival analysis techniques with censoring) will be used as-is. Because of the specification of the data in the LTE and DNHS data, missing data for timed outcome measures will be used as loss of ability at a particular time point due to disease progression only if a zero velocity has been specified in the dataset already provided to Dr. Dang. All assessments following a zero velocity visit will also be specified to have a zero velocity for the appropriate outcome. No multiple testing correction will be conducted given the number of subjects in high-dose group.

Specifics for comparative analysis:

1. High-dose treated subjects are those with 6 months of exposure (at baseline) to vamorolone at either 2.0 mg/kg/day or 6.0 mg/kg/day (combined together).
2. A table with binning by age of cross-sectional comparisons throughout for steroid-treated, steroid-naïve, and healthy subjects. Mean and SD, median, and interquartile range provided for these bins as long as cell size is greater than or equal to 5 (i.e., 5 subjects at least in each cell).
 - a. Use data from one visit per patient in any given cell. If two visits from a patient are available in a time period, average them.

Age (years)	Clinical Outcome	
[4.5,5)	Mean (SD) n=	Median (IQR)
[5,6)	Mean (SD) n=	Median (IQR)
[6,7)	Mean (SD) n=	Median (IQR)
[7,8)	Mean (SD) n=	Median (IQR)
[8,9)	Mean (SD) n=	Median (IQR)
[9,9.5]	Mean (SD) n=	Median (IQR)

3. Longitudinal visual comparisons will be done of steroid-treated, steroid-naïve, and healthy subjects using all data points available. All visits data will be used to visually compare overall trajectories of DNHS and LTE subjects over time.
4. Time-to-event analysis:
 - a. Nonparametric maximum likelihood estimation (NPMLE) of the survival curve for the first time to stand outcome of ≥ 10 seconds event (no stratification by steroid type) for CEM-matched subjects-based data from CINRG DNHS steroid-treated vs. LTE studies. This is a generalization of Kaplan-Meier analysis to interval-censored data.
 - i. Subjects: CEM-matched subjects-based data
 - ii. Exclude subjects whose time to stand from supine value at baseline is greater than or equal to 10 seconds.
 - iii. Event is defined as the first time a participant reaches a time to stand of more than or equal to 10 seconds.
 - iv. Time variable: Age in years.

- v. Data is interval censored when time to stand of 10 seconds occurred between two scheduled visits (e.g., subject at 7.5 years had a ttstand of 9 seconds and subject at next visit had a ttstand > 10 seconds). In interval censoring parlance, the event happens during the interval $(L_i, R_i]$ where L_i is the largest observation time before the event for subject i and R_i is the smallest observation time at or after the event happened (Fay and Shaw, 2010). If the event time is known exactly, L_i and R_i are the same. Data is considered right censored when at last clinical evaluation of subject (approximately 24 months following baseline visit), event had not yet occurred, i.e., timed test value was less than 10 seconds. The data fits into the above interval censoring format if R_i is specified to be infinity.
 - vi. Median survival times (i.e., median age to event of time to stand outcome of ≥ 10 seconds) along with 95% CIs (based on a modified bootstrap procedure) will be obtained for the two cohorts. The NPMLE of the survival curves will also be graphed with confidence intervals.
- b. A logrank two-sample hypothesis test (permutation based) for interval censored data comparing time to event experiences of the two cohorts based on time to stand outcome of ≥ 10 seconds and interval censoring.
- 5. Longitudinal comparison of motor outcomes:
 - a. TTSTAND, TTCLIMB, TTRW, PODCI (upper extremity and physical function; transfer and basic mobility): mixed models repeated measures (aka mixed-effect models) using all timepoints available adjusting for baseline performance, age, and comparing between cohorts. This will provide us with a comparison of the rate of change of motor outcomes.
 - i. This approach will rely on fixed timepoints as relevant to the trial: baseline, 12 month follow-up, 24 month follow-up, specified as a categorical variable with a cohort-by-time interaction.
 - ii. An age as time variable approach (random slope MMRM rather than based on clinical visits) will also be implemented relying on modeling visits based on age (continuous). In this approach, subjects initially assigned to low doses can be included as well (although only assessment data at LTE month 12 and month 24 is available for these subjects for this analysis). An additional lifetime exposure to high-dose corticosteroids at baseline variable will also be used here.
 - iii. This is stratified (by age) analysis with summaries provided for the following two strata: a) 5.5 years and below at baseline (corresponding to approximately 6 months of steroid treatment and b) Above 5.5 years at baseline.
- 6. Longitudinal comparison of BMI z-scores and height percentiles comparing between the two cohorts (DNHS steroid-treated and LTE) using MMRM.
- 7. Cross-sectional comparison of outcomes:
 - a. TTSTAND, TTCLIMB, TTRW, PODCI (upper extremity and physical function; transfer and basic mobility): An independent samples hypothesis test based on an independent samples t-tests [velocities; distributional assumptions met] or a Wilcoxon sum rank test [velocities or seconds] for the final measurements of timed function tests will be conducted. For a confirmed measurement of 0 velocity, an appropriately highest value for time in seconds will be used only for the purposes of the nonparametric Wilcoxon sum rank test; the use of this approach will be tested before interpretation to make sure that confidence intervals are unaffected by the highest value used. These measurements

correspond to 30-month exposure to steroidal products in both cohorts. This will give us an idea of whether performance between the two cohorts is in the same ballpark or not. Confidence intervals for comparisons of means or pseudo-medians between the two cohorts will be obtained.

8. Sensitivity analyses will be conducted with criteria-matched subjects.

References:

Iacus, Stefano M., Gary King, and Giuseppe Porro. "Multivariate matching methods that are monotonic imbalance bounding." *Journal of the American Statistical Association* 106, no. 493 (2011): 345-361.

Iacus, Stefano M., King, Gary, Porro and Giuseppe (2020). cem: Coarsened Exact Matching. R package version 1.1.20. <https://CRAN.R-project.org/package=cem>

Iacus, Stefano M., King, Gary, Porro and Giuseppe (2020). Vignette: CEM: Software for Coarsened Exact Matching. Nov 2, 2020. Accessed from <https://cran.r-project.org/web/packages/cem/vignettes/cem.pdf>

Appendix

Longitudinal Study of the Natural History of Duchenne Muscular Dystrophy (DMD) -

<https://clinicaltrials.gov/ct2/show/NCT00468832>

☐ Patient Information

☒ Participant ID

☒ Visit Information

☒ Visit Number, age at visit, country

☒ Demographics

☒ Age at baseline
☒ Race and ethnicity

☒ Anthropometric

☒ standing and calculated height (cm)
☐ ulnar length (cm)
☒ weight (kg)
☐ dominant side

☐ Vital Signs

☐ BP (sys / dias)
☐ Temperature
☐ Pulse respiratory rate

☒ Genetics

☒ Indication of muscle biopsy or immunoblot or DNA testing performed
☒ mutation start and end exons
☒ presence of a large deletion/duplication or small mutation
☐ eligibility for exons 45, 51, or 53 skipping

☐ Mortality

☐ Age of death, cause of death

☒ Corticosteroid Use

☒ current steroid use at visit (yes/no)
☒ lifetime steroid use up to visit
☒ drug taken

☒ QMT (quantitative muscle testing)

☐ Pinch and key pinch
☐ Grip
☒ Elbow extensor and flexor

☒ Knee extensor and flexor

☐ Extremity Tests

☐ Brooke scale
☐ Vignos scale

☐ Pulmonary function tests

☐ FVC (absolute and % predicted)
☐ FEV1 (absolute and % predicted)
☐ PEFr (absolute and % predicted)
☐ MIP (absolute and % predicted)
☐ MEP (absolute and % predicted)
☐ Peak flow cough (absolute)

☐ Manual muscle testing for:

☐ total
☐ ankle dorsiflexion, eversion, inversion, plantarflexion
☐ elbow extension and flexion
☐ hip abduction, extension, flexion
☐ knee extension, flexion
☐ shoulder abduction, external rotation
☐ thumb abduction
☐ wrist extension, flexion
☐ neck extension, flexion

☐ Range of motion – goniometry –

Muscles tested:

☐ Wrist, elbow, and knee extension
☐ Ankle dorsiflexion

☒ Time Function Tests

☒ Time to run/walk 10 meters
☒ Time to climb 4 stairs
☒ Time to stand from supine
☒ 6-minute walk test
☐ 9-hole peg test

☒ NSAA Scale - Scores for total and each component

☐ EK Scale - Scores for total and each component

☒ Ambulation

☒ ambulatory status
☐ use of assistive devices
☒ age at loss of ambulation
☐ frequency of falls

☐ Lower extremity surgery - Occurrence and age at 1st and 2nd surgery

☐ Other surgeries

☐ Cardiovascular
☐ gastrointestinal/abdominal
☐ genitourinary system
☐ otorhinolaryngology
☐ head/neck
☐ ophthalmologic
☐ dental

☐ Spine assessment and stabilization surgeries

☐ prior and current Cobb angle
☐ type of fusion
☐ lordosis
☐ kyphosis
☐ scoliosis

☐ Non-surgical hospitalizations

☐ Pneumonia
☐ respiratory failure
☐ congestive heart failure
☐ gastrointestinal tract issues
☐ genitourinary issues
☐ otorhinolaryngologic issues
☐ ophthalmologic issues
☐ central nervous system issues

- ☐neurological behavioral issues
- ☐Non-steroid medication history (cardiovascular)**
 - ☐angiotensin converting enzyme inhibitors or angiotensin-receptor blockers
 - ☐diuretics
 - ☐inotropic agents
 - ☐anti-arrhythmics
 - ☐beta-blockers
 - ☐CoQ10
 - ☐Idebenone
- ☐Non-steroid medication history (pulmonary)**
 - ☐bronchodilators
 - ☐anticholinergics
 - ☐mucolytic agents
 - ☐antibiotics
- ☐Non-steroid medication history (GI)**
 - ☐constipation
 - ☐gastro-esophageal reflux or gastritis
- ☐Non-steroid medication (bone health)**
 - ☐bisphosphonates
 - ☐vitamin D
 - ☐calcium
- ☐Non-steroid medication (strength)**
 - ☐Idebenone
 - ☐creatine
 - ☐vitamin E
 - ☐oral albuterol
 - ☐L-arginine
 - ☐anabolic agents
 - ☐glutamine
 - ☐multivitamins
 - ☐minerals
 - ☐pentoxifylline
- ☐Non-steroid medication history (mental health)**
 - ☐behavior/attention deficit/hyperactivity
 - ☐depression
 - ☐anxiety
- ☐Non-steroid medication history (pain)**
 - ☐acetaminophen
 - ☐NSAIDs or Cox-2 inhibitors
 - ☐neuropathic pain meds
 - ☐narcotics/opioids
- ☐Respiratory history (Includes age)**
 - ☐Breathing exercises performed:
 - ☐inspiratory muscle training
 - ☐incentive spirometry
 - ☐PEP
 - ☐glossopharyngeal breathing
 - ☐Cough-assistance measures used including:
 - ☐Vortran percussive nebulizer
 - ☐IPV/IBP
 - ☐in-exsufflator
 - ☐chest percussion
 - ☐TheraVest
 - ☐manual cough assistance
 - ☐Ventilatory assistance used

- including:
 - ☐Bi-PAP/Mask
 - ☐Bi-PAP/Nasal pillows
 - ☐C-PAP
 - ☐mouthpiece IPPB/IPPV
 - ☐cuirass
 - ☐negative pressure
 - ☐invasive (with tracheotomy)
 - ☐Supplemental oxygen use
 - ☐Suctioning
- ☐Ophthalmologic exam**
- ☐Blood CO2 assessment**
 - ☐end-tidal CO2
 - ☐arterial CO2
 - ☐transcutaneous CO2
- ☐Nutrition history**
 - ☐diet counseling
 - ☐modified diet
 - ☐nutritional supplements
 - ☐formula/calorie supplements
 - ☐modified food texture
 - ☐soft PG tube
 - ☐PEG tube
- ☐Pressure sores (Stage II & greater)**
 - ☐occiput
 - ☐back
 - ☐sacrum
 - ☐ischium
 - ☐trochanter
 - ☐knee
 - ☐foot/ankle
- ☒Fracture history**
 - ☐cervical
 - ☐thoracic
 - ☐lumbar
 - ☐sacral
 - ☐humerus
 - ☐femur
 - ☐tibia
 - ☐fibula
 - ☐radius
 - ☐ulna
 - ☐hand
 - ☐hip/pelvic
 - ☐foot
 - ☐cranial/facial
 - ☐elbow
 - ☐ankle
 - ☐clavicle
- ☐Echocardiogram results including LVEF & SF**
- ☐Skin issues**
 - ☐skin rash or hives
 - ☐skin fragility or easy bruising
 - ☐severe acne
 - ☐open pressure sores
- ☐Respiratory & cardiac issues**

- ☐chronic cough
- ☐bronchitis
- ☐pneumonia
- ☐asthma
- ☐weak cough
- ☐hay fever/allergies
- ☐shortness of breath
- ☐night sweats
- ☐morning headaches
- ☐nightmares
- ☐difficulty speaking
- ☐snoring
- ☐difficulty sleeping
- ☐breathing while supine
- ☐breathing while exercising
- ☐excessive sleepiness
- ☐difficulty concentrating
- ☐chest pains
- ☐racing heartbeat
- ☐leg or ankle swelling
- ☐congestive heart failure
- ☐Nutritional, gastrointestinal & urinary issues**
 - ☐feeding himself
 - ☐poor appetite
 - ☐difficulty chewing
 - ☐difficulty swallowing
 - ☐choking
 - ☐early satiety
 - ☐nausea
 - ☐indigestion
- ☐Nutritional, gastrointestinal & urinary issues**
 - ☐gastric reflux
 - ☐ulcers
 - ☐bloating
 - ☐constipation
 - ☐dehydration
 - ☐bowel obstruction
 - ☐diarrhea
 - ☐bowel accidents
 - ☐rectal bleeding
 - ☐dark stools
 - ☐blood in stool
 - ☐weight loss
 - ☐weight gain
 - ☐poor growth
 - ☐excessive appetite
 - ☐excessive thirst
 - ☐diabetes
 - ☐kidney stones
 - ☐bed wetting
 - ☐incontinence
 - ☐frequent urination
 - ☐fluid retention
 - ☐UTI
 - ☐night urination
- ☐Neuromuscular & neurological issues**
 - ☐abnormal bleeding/blood clotting problems
 - ☐unexpected infections
 - ☐joint/muscle pain or swelling
 - ☐muscle cramping

- ☐dizziness
- ☐vision changes
- ☐excessive fatigue
- ☐hearing problems
- ☐persistent headaches

☐Neurodevelopmental issues

- ☐significant behavioral problems
- ☐depression
- ☐autism
- ☐speech delay
- ☐language impairment
- ☐learning disability
- ☐sensory integration disorder
- ☐cognitive impairment
- ☐ADD/ADHD
- ☐mental retardation
- ☐mild development delay
- ☐pervasive developmental delay

☐Therapeutics attended/assistive devices used

- ☐an aquatic program
- ☐recreation therapy program
- ☐home passive stretching program
- ☐night splints for ankle positioning
- ☐body jacket/brace for the spine
- ☐standing board
- ☐Hoyer lift
- ☐hand splints
- ☐support group
- ☐individual social or behavioral therapy
- ☐group social or behavioral therapy

☐School

- ☐highest grade/years of school

- completed
- ☐have an IEP or 504 plan
- ☐received special services in the classroom
- ☐received special education outside classroom

☐Doctors (has seen/currently sees):

- ☐primary care physician or pediatrician
- ☐physiatrist
- ☐orthopedist
- ☐neurologist
- ☐cardiologist
- ☐pulmonologist
- ☐gastroenterologist
- ☐psychiatrist
- ☐ophthalmologist
- ☐ENT specialist
- ☐urologist

☐Allied health professionals seen

- ☐physical therapist
- ☐occupational therapist
- ☐speech therapist
- ☐respiratory therapist
- ☐nutritionist
- ☐psychologist/mental health worker
- ☐social worker

☐Life Satisfaction Index scores

- ☐general well-being
- ☐interpersonal relationship
- ☐personal development
- ☐personal fulfillment
- ☐leisure and recreation

☐NeuroQoL Disorders scores

- ☐anxiety
- ☐depression
- ☐fatigue
- ☐upper extremity
- ☐lower extremity
- ☐mobility, wheelchair

☐PedsQoL Inventory scores

- ☐physical functioning
- ☐emotional functioning
- ☐social functioning
- ☐school functioning

☒POSNA Pediatric Musculoskeletal Functional Health Questionnaire scores

- ☒global function
- ☒happiness
- ☒pain/comfort
- ☒sports & physical functioning
- ☒transfer & basic mobility
- ☒upper extremity
- ☒physical function

☐WHOQoL scores

- ☐physical
- ☐psychological
- ☐social
- ☐environment

☐PedsQL Neuromuscular Module scores

- ☐about my neuromuscular disease
- ☐communication
- ☐about our family

☐Questionnaires

- ☐responses to individual questions for each questionnaire administered

eAppendix 2. NorthStar United Kingdom Network Cohort External Comparison Workflow Plan

Version Date: Jan 22, 2021

Operating Procedure for Data Extraction and Analysis

Specifics for data extraction of DMD subjects:

1. Subjects 4.5 - 7.5 years old at baseline visit (baseline for this data extraction).
2. Subjects able to perform the time to stand test unsupported (NSAA score 1 or 2) at baseline.
3. Steroid treated for approximately 6 months at baseline and maintained on corticosteroids (prednisone, prednisolone, deflazacort) for subsequent visits (daily throughout or intermittent throughout or switch from daily to intermittent or vice versa) regimen. Subjects should have at least two visits including baseline visit (which corresponds to approximately 6 months of steroid exposure) and a maximum follow-up of 24 months included in this comparator data.
Specifics for programming: Current steroid exposure at baseline should be between 5 and 12 months (this is the window around approximately 6 months of steroid exposure). Follow-up with steroid exposure after the baseline visit (~6 months of steroid exposure) should be at least 10 months (approximately 1 year). For programming purposes, there is a three-month additional window around the maximum of 24 months exposure, i.e., maximum of 27 months maximum follow-up.
4. Subjects not co-enrolled in other exon-skipping clinical trials. This applies to both during the 6 months of pre-steroid exposure as well as the follow-up time of 24 months.
5. Subjects not on ataluren.
6. Baseline measures available: time to stand from supine (time in seconds) and NorthStar Ambulatory Assessment total score (composed of **all** 17 subscores).

Specifics for analysis:

2. A table with binning by age of cross-sectional comparisons throughout. Mean and SD, median, and interquartile range provided for these bins as long as cell size is greater than or equal to 5 (i.e., 5 subjects at least in each cell).
 - a. Use data from one visit per patient in any given cell. If two visits from a patient are available in a time period, average them.

Age (years)	Total NSAA score		Time to stand velocity	
	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[4.5,5)	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[5,6)	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[6,7)	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[7,8)	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[8,9)	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)
[9,9.5]	Mean (SD) n=	Median (IQR)	Mean (SD) n=	Median (IQR)

3. Nonparametric maximum likelihood estimation (NPMLE) of the survival curve for the first time to stand outcome of ≥ 10 seconds event (no stratification by steroid type). This is a generalization of Kaplan-Meier analysis to interval-censored data.
 - a. Subjects: 4.5 - 7.5 years, steroid exposure ~ 6 months at baseline
 - b. Exclude subjects whose time to stand from supine value at baseline is greater than or equal to 10 seconds.
 - c. Event is defined as the first time to stand assessment of more than or equal to 10 seconds.
 - d. Time variable: Age in years.
 - e. Data is interval censored when time to stand of 10 seconds occurred between two scheduled visits (e.g., subject at 7.5 years had a ttstand of 9 seconds and subject at next visit had a ttstand > 10 seconds). In interval censoring parlance, the event happens during the interval $(L_i, R_i]$ where L_i is the largest observation time before the event for subject i and R_i is the smallest observation time at or after the event happened (Fay and Shaw, 2010). If the event time is known exactly, L_i and R_i are the same. Data is considered right censored when at last clinical evaluation of subject (approximately 24 months following baseline visit), event had not yet occurred, i.e., timed test value was less than 10 seconds. The data fits into the above interval censoring format if R_i is specified to be infinity. For *R* users, the *interval* (Fay and Shaw, 2010) package can be used.
 - f. Provide mean, SD, median, and interquartile range of baseline time to stand from supine values in seconds.
 - g. Provide mean, SD, median, and interquartile range of baseline age in years.
 - h. Provide median age at event in years from NPMLE along with 95% CI, how many subjects at baseline, how many events occurred (≥ 10 seconds). Confidence intervals is based on a modified bootstrap procedure.
 - i. In *R*, the *interval* package allows for this with (default) 200 bootstrap replications.
 - ii. The *interval* package does not explicitly output an estimate for median survival and for the 95% CI. Within intervals, an estimate of median survival as well as the 95% CI can be inferred through similar calculations as in the Supplementary Material of Dugué et al., 2016 where *R* code has been provided.
 - i. Provide plot of survival curve with survival probability (y-axis) and age of DMD subject (x-axis) with 95% confidence interval. Add a horizontal line colored red at y-axis = 0.50.
 - j. Provide survival table with survival curve, confidence limits, and other information provided by default through print (objectname) in *R*.
 - k. Provide commands (code) for the analysis in *R* so that we can replicate those arguments. Use default arguments if appropriate.
4. Change in NorthStar total points (raw, not linearized) over 2 years
 - a. Subjects: 4.5 - 7.5 years, steroid exposure ~ 6 months at baseline
 - b. NorthStar Ambulatory Assessment total score should be composed of ALL 17 subscores; otherwise, it is incomplete for the purposes of this analysis and should not be used.
 - c. For programming: Note that the 24th month visit can be anywhere from 21 months to an upper tolerance (we specify 27 months here).
 - d. Provide mean at baseline (with SD), mean at 2 years in (with SD), mean change over 2 years (with SD), number of subjects with data at both these timepoints,

and 95% CI for the mean change. This is stratified (by age) analysis with summaries provided for the following two strata:

- i. 5.5 years and below at baseline (corresponding to approximately 6 months of steroid treatment).
- ii. Above 5.5 years at baseline.
- e. Provide command (code) for the paired analysis (t-test) in *R*.
5. Shift analysis for NorthStar UK cohort (shift up or shift down or no change)
 - a. Subjects: 4.5 - 7.5 years, steroid exposure ~ 6 months at baseline
 - b. NorthStar Ambulatory Assessment total score should be composed of **ALL** 17 subscores; otherwise, it is incomplete for the purposes of this analysis and should not be used.
 - c. Shift up (gain) over follow-up (minimum 10 months; maximum 24 months): Defined as proportion of subjects who shifted from 0 → (1,2) or 1 → 2 over 24 months. Provide proportion for each of the 17 functions.
 - d. Shift down (loss) over follow-up (minimum 10 months; maximum 24 months): Defined as proportion of subjects who shifted from 2 → (0,1) or 1 → 0 over 24 months. Provide proportion for each of the 17 functions.
 - e. No change over follow-up (minimum 10 months; maximum 24 months): Defined as no change in score over 24 months. Provide proportion for each of the 17 functions.
 - f. Provide a frequency table of subjects whose data corresponded to 0, 1, and 2 scores on the 17 activities at baseline. This is stratified (by age) analysis with tables provided for the following two strata:
 - i. 5.5 years and below at baseline (corresponding to approximately 6 months of steroid treatment).
 - ii. Above 5.5 years at baseline.

Activity	0	1	2	Total number of subjects
Stand				
Walk				
Sit to stand				
Get to sitting				
Stand on one leg (L)				
Stand on one leg (R)				
Climb box step (L)				
Climb box step (R)				
Descend box step (L)				
Descend box step (R)				
Lifts head				
Jump				
Run				
Rise from floor				
Stands on heels				
Hop (L)				
Hop (R)				

- g. Provide a frequency table of subjects whose data corresponded to 0, 1, and 2 scores on the 17 activities at the 24-month visit. This is stratified (by age) analysis with summaries provided for the following two strata:
 - i. 5.5 years and below at baseline (corresponding to approximately 6 months of steroid treatment).

ii. Above 5.5 years at baseline.
Specifics for programming: 24th month visit can be anywhere from 21 months to an upper tolerance (we specify 27 months here).

Bibliography:

Dugué, Audrey Emmanuelle, Marina Pulido, Sylvie Chabaud, Lisa Belin, and Jocelyn Gal. "How to deal with interval-censored data practically while assessing the progression-free survival: a step-by-step guide using SAS and R software." *Clinical Cancer Research* 22, no. 23 (2016): 5629-5635.

Fay, Michael P., and Pamela A. Shaw. "Exact and asymptotic weighted logrank tests for interval censored data: the interval R package." *Journal of Statistical Software* 36, no. 2 (2010).

eAppendix 3. Supplemental Text on Statistical Analysis

Analysis Population: All subjects who received at least one dose of vamorolone medication in the VBP15- LTE extension study were included in the Safety Population. The Safety Population was the primary analysis population for safety assessments. This was also the modified Intention to Treat (mITT) population.

Real-world Comparator Population: Previous papers have noted down-titration of glucocorticoids compared to standard of care guidelines (for example, average dose for daily prednisone at 75% recommended and for daily deflazacort at 83% recommended, according to Bello et al 2015). Our objective was to compare vamorolone treatment to real-world glucocorticoid (GC) usage, which includes intermittent treatment as well as switching between daily and intermittent regimens. We provide a table below for the two cohorts (NSUK and DNHS) and proportions of unknown (but continuous treatment without breaks) regimen, intermittent, daily, and switching between intermitted and daily regimens (either direction). For DNHS cohort (used in subject-level comparisons with MMRM), the data is very similar, with the majority ($\geq 85\%$) on a daily GC regimen. For NSUK, after accounting for the large proportion of unknown and switched regimens (combined $\geq 30\%$ participants), there appears to be a similar split between daily and intermittent only regimens.

	Daily	Intermittent	Switch (between daily and intermittent)	Unknown
Proportion based on all subjects (n=75; DNHS)	85.3%	6.7%	2.7%	5.3%
Proportion based on subjects with long-term* follow-up (n=30; DNHS)	90%	6.7%	3.3%	0%
Proportion based on all subjects (n=110; NSUK)	33.6%	35.5%	14.5%	16.4%
Proportion based on subjects with long-term* follow-up (n=49; NSUK)	32.7%	30.6%	20.4%	16.3%

* Long-term follow-up defined here as >18 months of follow-up for DNHS, and having a 2-year follow-up visit for NSUK.

Data Analysis Considerations:

Transformations: For TTSTAND, TTRW, and TTCLIMB, results in seconds were converted to velocities (1 / TTSTAND expressed as rises/second [TTSTAND velocity], 10 / TTRW expressed as meters/second [TTRW velocity], and 1 / TTCLIMB expressed as tasks/second [TTCLIMB velocity]).

Missing data: Different subsets of higher-dose (≥ 2 mg/kg/day) participants completed all planned outcomes. In terms of baseline (multiple-ascending-dose

[VBP15-002] study for all but PODCI for which baseline was at start of dose-finding [VBP15-003] study), long-term extension [VBP15-LTE] month 24 visit, and pairwise complete data (participants with both VBP15-002 baseline and VBP15-LTE month 24 final visits), the number of samples per outcome are TTSTAND (23; 21; 21), TTCLIMB (23; 18; 18), TTRW (23; 18; 18), 6MWT (20; 17; 15), NSAA (23; 18; 18), QMT (23; 5; 5), PODCI (18; 18; 15). Reasons for missing data included subject early withdrawal, disease progression, COVID19-related restrictions, and equipment/software issues (for QMT) (see **Table** below).

Time to event analysis: As noted in Appendices 1 and 2, the time to event analysis excluded participants whose time to stand from supine value at baseline was greater than or equal to 10 seconds. Hence, for this analysis, the following number of participants were available for the three cohorts: 108 (NSUK), 74 (DNHS), and 22 (VBP15-LTE).

Withdrawals and terminal visits: If a participant withdrew but their end-of-follow up clinic visit was close to a scheduled visit, their data was considered to occur at that scheduled visit for the purpose of the analysis. For analysis, only the observed data was utilized, i.e., no multiple imputation was done.

Table: Number of participants treated throughout VBP15-003/-LTE with higher-dose vamorolone (≥ 2.0 mg/kg/day) at baseline and at VBP15-LTE month 24 visit

Time to stand test (n=23 at baseline, n=21 at LTE month 24 visit)	Time to climb test (n=23 at baseline, n=18 at LTE month 24 visit)	Time to run/walk 10m test (n=23 at baseline, n=18 at LTE month 24 visit)	Six-minute walk test (n=23 at baseline, n=17 at LTE month 24 visit)	NorthStar Ambulatory Assessment (n=23 at baseline, n=18 at LTE month 24 visit)
Missing velocity: Withdrew early, n=2	Missing velocity: COVID-19, n=3 Withdrew early, n=2	Missing velocity: COVID-19, n=3 Withdrew early, n=2	Missing data: COVID-19, n=3 Disease progression*, n=1 Withdrew early, n= 2	Missing data: COVID-19, n=3 Withdrew early, n=2

*This participant had missing data for six-minute walk test, had a zero velocity imputed for the timed function tests, but the NSAA was measured at LTE Month 24 visit.

More details on Statistical Modeling:

Delayed start analysis:

For the delayed start analyses, participants initially assigned to vamorolone 0.25 mg/kg/day (group A) and 0.75 mg/kg/day (group B) (<2.0 mg/kg/day dosing

group, n=23) were considered “delayed starters”. Outcomes for these participants were compared to those who were initially assigned to, and retained on, higher dose throughout their follow-up (≥ 2.0 mg/kg/day, group C (2 mg/kg/day) and group D (6 mg/kg/day), n=23). For the delayed start analyses, the estimates and standard errors were calculated with a mixed-effect model with repeated measures (MMRM). Data from all relevant post-baseline visits (two visits from dose-finding studies [VBP15-003] and two visits from long-term extension [VBP15-LTE] trials) were used as response variables. The explanatory/independent variables included initial dose levels (high vs low), visit, interaction between dose level and visit, and baseline value as a covariate.

Comparison of vamorolone-treated to CINRG DNHS participants (1st historical control cohort):

Analysis was pre-specified as based on an external comparison workflow plan (Supplement 2). Comparison of higher-dose VBP15-LTE participants with GC-treated historical controls from DNHS was based on longitudinal outcome data using MMRM. Baseline responses and age were included as covariates and every post-baseline comparison visit was used, to a maximum follow-up in DNHS of 27 months. Given that the age range of interest has a period of improvement followed by deterioration for muscle outcomes, a quadratic term was also included to model trajectories. A random-intercept (no random slopes) model was used. Nonparametric maximum likelihood estimation (NPMLE) was used to estimate the observed survival curve for the first TTSTAND outcome of ≥ 10 seconds event. The median time to event estimate as well as confidence intervals (bootstrapped) were obtained for the VBP15-LTE and DNHS cohorts.

Comparison of vamorolone-treated to NorthStar UK participants (2nd historical control cohort):

Analysis was pre-specified as based on an external comparison workflow plan (Supplement 3). Due to NorthStar UK participant-level data sharing restrictions, the NSUK data were analyzed separately by UK Network researchers. Nonparametric maximum likelihood estimation (NPMLE) was also obtained for the observed survival curve (median time to event estimate as well as bootstrapped confidence intervals) for the first TTSTAND outcome of ≥ 10 seconds event. Summaries of NSAA change were compared with the higher-dose VBP-LTE participants using independent t-tests (and not MMRMs) and relative risk analysis (with Wald confidence intervals) of loss of ability.

References:

Bello L, Gordish-Dressman H, Morgenroth LP, et al. Prednisone/prednisolone and deflazacort regimens in the CINRG Duchenne Natural History Study. *Neurology*. 2015;85(12):1048-1055. doi:10.1212/WNL.0000000000001950