Detecting meaningful change using the North Star Ambulatory Assessment in Duchenne muscular dystrophy

ANNA G MAYHEW[1] | STEFAN J CANO[2] | ELAINE SCOTT[3,4] | MICHELLE EAGLE[1] | KATE BUSHBY[1] | ADNAN MANZUR[5] | FRANCESCO MUNTONI[5] | ON BEHALF OF THE NORTH STAR CLINICAL NETWORK FOR NEUROMUSCULAR DISEASE

1 Institute of Genetic Medicine, Newcastle University, Newcastle; **2** Peninsula Schools of Medicine and Dentistry, Plymouth University, Plymouth; **3** School of Health and Related Research, University of Sheffield; Sheffield; **4** Muscular Dystrophy Campaign, London; **5** Dubowitz Neuromuscular Centre, University College London, London, UK.

Correspondence to Dr Anna G Mayhew, The MRC Centre for Neuromuscular Diseases, International Centre for Life, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK.
E-mail: anna.mayhew@newcastle.ac.uk

PUBLICATION DATA
Accepted for publication 30th April 2013.
Published online 5th August 2013

ABBREVIATIONS

DMD	Duchenne muscular dystrophy
MID	minimal important difference
NSAA	North Star Ambulatory Assessment

AIM Clinician-reported outcome instruments such as the North Star Ambulatory Assessment (NSAA) need to be able to detect clinically important change to be suitable for clinical trials. However, in Duchenne muscular dystrophy (DMD), identifying changes in function is not straightforward. In this study, we use Rasch-transformed data to examine the responsiveness and minimal important difference (MID) of the NSAA in males with DMD receiving different corticosteroid regimes.

METHOD NSAA data were examined from 198 males (mean age at assessment was 8y 6mo [SD 2y 6mo] range 4y–18y; 805 assessments). Responsiveness was assessed using mean score changes (using Rasch-transformed data) between adjacent pairs of age groups, pairwise squared *t*-values from paired samples *t*-tests, and an effect size calculation. The MID was assessed using the effect size calculation and 0.5 standard deviation (SD) of mean score differences.

RESULTS Our findings revealed a difference in change scores over time between the two corticosteroid regimes. Mean NSAA person estimates were higher in the daily prednisolone group. The mean MID (0.5 SD) was 8.8 and 6.9 for the daily group and intermittent group respectively.

INTERPRETATION This study, based on Rasch-transformed NSAA data, provides an initial basis for the interpretation of clinical change in DMD over time and between corticosteroid regimes. Our proposed MIDs can be mapped back to differences in specific item content across the range of the NSAA.

Duchenne muscular dystrophy (DMD) is a severe, genetic X-linked disease that affects 1 in 3600 to 6000 live male births.[1,2] The predominant feature of the disease is progressive muscle weakness that in the early years manifests itself as delayed motor milestones and an inability to run and jump. Weakness progresses so that, in untreated individuals, the ability to walk is lost, on average, at the age of 9 years 6 months. Loss of ambulation can be delayed with corticosteroid therapy, which is now an accepted part of standard care, although the benefit of one steroid regime over another has not been established.[3,4] However, recent clinical data collected in the UK, in which the effect of daily versus intermittent (10d on and 10d off) corticosteroid therapy was evaluated, indicate that the current mean age at loss of ambulation is 12 years in children on the intermittent regimen and 14 years 6 months in children on daily corticosteroids.[5]

In recent years, our understanding of DMD disease progression has been improved, in part through the use of clinical rating scales.[6] The North Star Ambulatory Assessment (NSAA) is a promising new clinician-reported outcome instrument consisting of 17 items designed to measure ambulatory function in males with DMD.[7] To date, it has undergone detailed psychometric evaluations based on traditional (reliability and validity) and modern (Rasch analysis) methods.[7–9] However, despite its inclusion in international clinical trials[10] and the availability of longitudinal 12-month data,[11] the extent to which the NSAA captures clinically meaningful change has yet to be formalized. This property of a scale is known as responsiveness.[12]

Minimal important difference (MID) can be described as a numerical description of the smallest change relevant to the patient, family, or clinician (positive or negative) which has the potential to alter management.[13] The same amount of change may be interpreted differently depending on the perspective taken. For example, the loss of the ability to stand on one leg to kick a ball may be crucial to a male par-

 DOI: 10.1111/dmcn.12220

ticipating in a football match with his friends, but may appear a small loss to clinicians in the context of the overall clinical assessment. There are two main types of MID. The first type is distribution-based in which scale data are interpreted without reference to an outside measure.[14] As such, an MID can be considered as either (1) corresponding to one-half of the baseline standard deviation in the sample[15] or (2) an effect size, with, according to Cohen, an effect size of 0.2, 0.5, and 0.8 relating to a small, moderate, and large change respectively.[16] The second type is anchor-based, in which rating-scale level data are compared with an anchor (e.g. a global rating related to a specific health construct or change in construct).[17–19] In this study, we focused on the former type, as no global external measure was available. There is no criterion standard for assessing MID values but, by using more than one distribution-based method, a more accurate definition of MID may be achieved.

The aim of this study was to examine the responsiveness and magnitude of clinical differences of the NSAA in ambulant males with DMD in relation to age and corticosteroid regime. Potentially, this could link the identified MID, which would assist our understanding of changed scores, both clinically and within trial settings.

What this paper adds

- The NSAA can detect differences between two corticosteroid treatment regimes used in DMD.
- Estimated MID relates to interpretable clinical change in ambulant children with DMD.
- Rasch analysis improves the process of examining rating scales by linearizing the scale scores.

METHOD

Recruitment

Eligible participants (males with a genetically and/or pathologically confirmed diagnosis of DMD) were recruited through the North Star Database, a secure web-based database which collects de-identified medical and physiotherapy assessment data on ambulant males with DMD from a network of 20 specialist centres in the UK.[7] A request to extract the data from the system was approved by the North Star Governance Committee. This provided prospective collected, longitudinal data.

Data collection

Data relating to the age of the males at assessment, serial assessments on the NSAA, corticosteroid regime, age at which steroids were started, length of time on steroids, and type of steroid used were extracted from the database. Data relating to steroid regime were included only if the regime remained stable over the given time period, that is, not switching from a daily to an intermittent regime or vice versa over any time point for which data were available. Males were excluded from analysis if the regime or type of steroid used altered over time. Males on prednisolone and deflazacort were selected for analysis if they were on a stable regime.

Outcome measure

The NSAA is a multiple-item rating scale (17 items), with three ordered response categories (2, 1, or 0), which are summed to give a total score.[7] Items are scored either 2 ('normal' with no obvious modification of activity), 1 (modified method but achieves goal, independent of physical assistance from another), or 0 (unable to achieve independently). A total 'ambulatory function' score is generated by summing items. A higher score indicates better motor function. Full test details are available from the Muscular Dystrophy Campaign.[20] The data collected represent data from the clinical setting, where the test is performed approximately every 6 months, though this varies from centre to centre (±3mo).

Analysis

There were three stages of analysis: (1) transformation of ordinal-level scores into linearized measurements; (2) responsiveness analysis based on longitudinal data collected between 2006 and 2011; and (3) MID statistics.

Transformation of raw scores into linearized measurements

Before responsiveness analysis, raw (ordinal level) scores on the NSAA were converted to linearized 'person estimates' (or measures). The reason for this transformation is essentially to take account of the fact that fixed changes in ordinal level scores (0–34 in the case of the NSAA) may imply variable changes in interval-level measurements.[21] In other words, a one-point change does not mean the same across the breadth of the scale. Analysing ordinal-level data may hide a true change or amplify it because total scores have a non-linear relationship to the trait they seek to measure, in this case ambulatory function. In contrast, Rasch-derived person estimates are linear measures because they have a linear relationship to the underlying trait they seek to measure. A one-point change means the same across the scale in a Rasch-transformed score.

Transformation can take place only if the data satisfy the requirements of the mathematical model, in this case the Rasch model.[22] This model articulates a theory of how rating scales ought to perform if the values they generate are to be considered scientific measurements. Thus, when the data examined here fit the requirements of the Rasch model (using RUMM 2030 software[23]), there was evidence that the NSAA was a measurement instrument: all items fit the original construct of ambulation (except 'lifts head' which was removed from the total raw score), targeting was sufficient, all response categories were ordered, the person separation index was high (0.91), there was minimal effect of item dependency on reliability (person separation index 0.89), and there was suitable stability over regime and age. This confirmed the findings in our previous NSAA Rasch analysis article[9] (further information available from authors). Under these circumstances, we were able to construct linearized estimates for people from the scale data (Table I). These 'person location estimates' are in

Table I: NSAA scale in the EINSTEIN dataset (linearized data): ordinal level score to linearized person estimates transformation raw score, logit and logit transformed 0–100

Raw score	Logit	Logit (transformed 0–100)
0	−5.11	0[a]
1	−4.25	11
2	−3.63	17
3	−3.19	21
4	−2.83	24
5	−2.52	27
6	−2.24	30
7	−1.99	32
8	−1.74	34
9	−1.51	36
10	−1.29	38
11	−1.07	40
12	−0.85	42
13	−0.64	44
14	−0.44	46
15	−0.23	48
16	−0.03	50
17	0.17	52
18	0.37	53
19	0.57	55
20	0.77	57
21	0.98	59
22	1.19	61
23	1.42	63
24	1.65	65
25	1.91	67
26	2.19	70
27	2.50	73
28	2.87	76
29	3.29	80
30	3.82	85
31	4.54	91
32	5.50	100

[a]Value extrapolated owing to the asymmetry of the data (more information available from authors). NSAA, North Star Ambulatory Assessment.

log-odds units (logits). For each individual's location, the analysis also generates a bespoke standard error.

Responsiveness analyses

The responsiveness of the NSAA was examined at the group level in three ways: (1) examinations of mean score changes between adjacent pairs of age groups (e.g. mean change score from 5 to 6y of age); (2) pairwise squared *t* values from paired-samples *t*-tests (including a reported *p* value); and (3) an effect size calculation, first proposed by Cohen.[16] This reflects the difference between the mean scores of two groups (e.g. aged 10y compared with aged 11y) divided by the SD of the baseline scores, in this instance SD in the 10 to 11 years age group. Interpretation of this change is guided by proposed benchmarking, where an effect size of 0.20 or less represents a change of approximately one-fifth of the baseline SD and is considered small, whereas an effect size of 0.50 is considered moderate and an effect size of 0.8 is viewed as large.[16]

Minimal important difference statistics

In this study, two main sets of standard MID statistics were generated: (1) a distribution-based indicator was generated by calculating a 0.5 SD for the mean scores at all time points from NSAA data generated from the daily versus intermittent regime of prednisolone[24]; and (2) an effect size statistic was calculated as outlined above, which relates to the responsiveness analysis described before.[16]

RESULTS

Sample

A total of 198 males from 16 different UK centres were recruited for analysis from the database; 805 serial assessments were included, with 754 having six serial assessments. The mean patient age at assessment was 8 years 6 months (SD 2y 6mo; range 4y–18y). The mean age at which corticosteroids were started was 6 years 4 months (SD 1y 7mo; range 2y 11mo–12y 1mo), mean length of time on corticosteroids was 2 years 7 months (SD 1y 7mo; range 1mo–7y 8mo). The ratio of participants who had not yet started corticosteroids to prednisolone to deflazacort was 97:648:60. Those on a stable deflazacort regime were excluded from responsiveness analysis, as insufficient data were available. The ratio of type of corticosteroid regime – not yet started to daily to intermittent (10d on 10d off) – was 97:393:315.

Transformation of raw scores into linearized measurements

Table I illustrates the transformation of the raw score to the linearized measurement on a scale of 0 to 100 via a logit transformation. For example, a male who scored 12 out of 32 on the NSAA (excluding his score on item 12 – lifts head) can be calculated to have a linearized measurement of 42 out of 100.

This table demonstrates that raw score differences across the range of each scale have different meanings in terms of person estimates of ambulation. Thus, a raw score change of one point equates to between a 1- and an 11-point difference, depending where on the measurement continuum patients are located. This implies up to an 11-fold difference across the range of the scale, and this especially holds true for the floor and ceiling of the scale, where scores are less precise. Figure 1 illustrates the calibration carried out in order to calculate MID scores. This calibration equalizes the thresholds across items so that the MID can be calculated appropriately. All thresholds are ordered as shown in earlier data sets.[9]

Responsiveness analyses

Table II (daily vs intermittent prednisolone regime) shows the descriptive statistics for each age group for the NSAA. In general, across each of the age groups, mean person estimates (mean change scores) were higher in the daily prednisolone group (see also Fig. 2 for a graphic representation of the data). The positive mean changes in the 3 or 4 to 5 years age group and the 5 to 6 years age group indicate an improvement (range 3.1–7.5), whereas negative values indicate a loss of ambulatory ability. This negative value varied for both groups, with larger values more evident as males grew older (range −0.2 to −13.8).

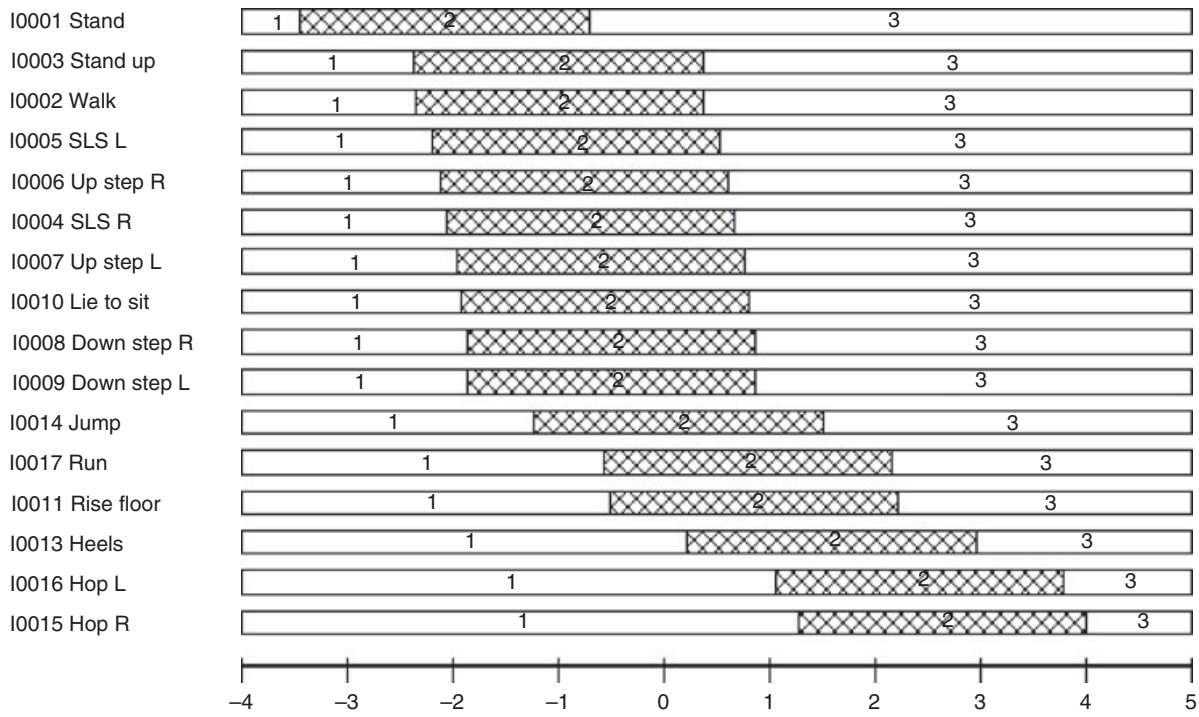

Figure 1: North Star Ambulatory Assessment (NSAA) scale (linearized data). The figure shows the threshold maps for all items in the NSAA. In each instance the *x*-axis represents the construct (ambulatory ability), with ability increasing left to right. The *y*-axis shows each of the items response category: unable labelled as 1; with modification labelled as 2; able labelled as 3. The map shows the rating scale mode generated thresholds, in which the thresholds are equalized across items. This was done to enable appropriate values for minimal important difference calculation.

Table II: NSAA (linearized data): daily prednisolone regime compared with intermittent prednisolone regime responsiveness compared across pairs of age groups

Age (y)	Daily prednisolone regime						Intermittent prednisolone regime					
	n	Mean change score[a]	SD	*t*	*p*	Effect size	*n*	Mean change score[a]	SD	*t*	*p*	Effect size
3/4–5	9	3.1	5.2	1.77	0.12	0.57	20	7.2	8.5	3.76	0.00	0.90
5–6	26	7.5	12.2	3.13	0.00	0.63	18	3.9	8.9	1.85	0.08	0.39
6–7	37	−2.1	11.9	−1.09	0.28	−0.15	21	−0.2	6.8	−0.16	0.87	−0.02
7–8	29	−3.7	11.3	−1.79	0.08	−0.25	21	−4.6	12.5	−1.68	0.11	−0.45
8–9	19	−7.0	14.8	−2.07	0.05	−0.48	25	−5.7	10.8	−2.67	0.01	−0.53
9–10	25	−4.3	12.2	−1.75	0.09	−0.23	24	−10.5	16.3	−3.17	0.00	−0.95
10–11	17	−10.3	14.9	−2.84	0.01	−0.48	11	−13.8	11.9	−3.83	0.00	−1.01
11–12	12	−13.8	17.6	−2.71	0.02	−0.67	9	−6.1	15.6	−1.18	0.27	−0.52

Males may be represented more than once in one age group or may be represented in more than one age group. [a]A mean change score value that is positive represents increasing ambulatory ability whereas a negative value represents a loss of ambulatory ability. NSAA, North Star Ambulatory Assessment.

Table III (daily vs intermittent prednisolone regime) shows the mean transformed NSAA score, and the associated SD (responsiveness) for different age groups. This relates to the actual scores on the transformed 0 to 100 scale. In general, the responsiveness statistics suggest a peak in mean scores at 6 to 7 years of age for males on a daily regime (mean 72.2–73.6) and a peak at 6 years of age of those on an intermittent regime (mean 68.9). Mean scores fell below 50 at 11 years of age for those on an intermittent regime and, for those on a daily regimen, the mean remained above 50 for those over 12 years of age (mean 52.9).

Minimal important difference statistics

Table III shows the 0.5 SDs for each adjacent age group according to prednisolone regime (daily vs intermittent). The mean MID (0.5 SD) was 8.8 and 6.9 for the daily prednisolone group and intermittent prednisolone group

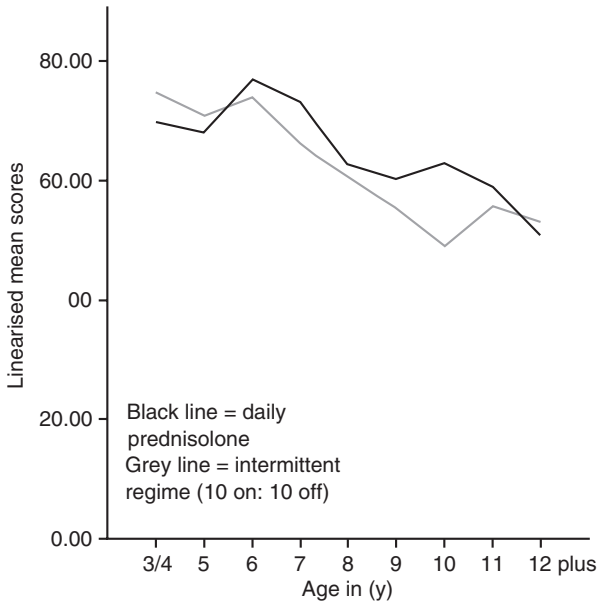

Figure 2: North Star Ambulatory Assessment (NSAA) scalee (linearized data). This figure shows all age groups, with a comparison of mean daily prednisone regime (black) and intermittent prednisolone regime (grey).

Table III: NSAA (linearized data): daily prednisolone regime compared with intermittent prednisolone regime responsiveness at each time point (minimal important difference)

Transformed location age at assessment (y)	Daily prednisolone regime				Intermittent prednisolone regime			
	n	Mean	SD	0.5 SD[a]	*n*	Mean	SD	0.5 SD[a]
3 and 4	10	61.6	5.9	3.0	23	62.3	8.9	4.4
5	31	65.7	11.5	5.7	30	65.7	10.9	5.4
6	48	73.6	13.6	6.8	35	68.9	12.9	6.4
7	56	72.2	14.4	7.2	39	64.2	13.8	6.9
8	38	67.0	18.2	9.1	39	60.1	15.0	7.5
9	39	63.3	20.4	10.2	41	58.3	14.2	7.1
10	37	66.1	22.8	11.4	30	52.7	19.8	9.9
11	22	57.1	27.5	13.7	16	50.0	13.9	6.9
12	15	52.9	24.4	12.2	11	46.2	14.5	7.2
Mean				8.8				6.9

[a]0.5 SD defined here as minimal important difference. A minimal important difference can be summarized to represent a 7 to 9 point change on the transformed North Star Ambulatory Assessment (NSAA) scale.

respectively. For the group on daily prednisolone, the 0.5 SD increased until the age of 11 years (13.7) and then fell slightly (12.2). For the group on intermittent steroids, the 0.5 SD increased by smaller increments to 9.9 at 10 years of age and then demonstrated a greater fall (7.2). This MID, shown to be about 10 points (between 7 and nearly 14 points), can be translated for males at different levels of ability. Using Table II, a transformed score of 50 may fall by 10 points to 40 (logit fall 0.03 to −1.07). This logit fall can be interpreted by examining Figure 1, which shows that a male would lose the ability to speed up from walking speed (item 2) and would now need the support of furniture to get off the floor or would be unable to get up without assistance (item 11).

Significant effect sizes (moderate and above) were seen at either end of the age groups examined: those on daily prednisolone showed a moderate increase in effect size up to 6 years of age and a moderate decrease from 10 years of age. Males on an intermittent regime showed a moderate increase in effect size up to 5 years of age and a moderate to large fall in effect size from 9 years of age and older. This relates to responsiveness analysis, where an MID is considered to equate to an effect size of 0.5.[25]

DISCUSSION

The aim of this study was to examine the responsiveness and magnitude of clinically meaningful change using the NSAA in ambulant males with DMD, in relation to age and corticosteroid regime. This analysis was based on the link between measures of responsiveness and MID. Our findings show that the NSAA was able to detect differences in the two regimes and that responsiveness could be shown to change with age. It can be seen that ambulation improves in younger males – positive mean change score – but then falls significantly after 10 years of age. This is in accordance with the known clinical course of the disease[3] and demonstrates the need to take into account age and stage when assessing the impact of treatment. The MID, as calculated using 0.5 SD and in relation to an effect size of 0.5, was demonstrated to be around 10 points on the transformed NSAA scale.

Importantly, the proposed MIDs can be equated to significant 'milestones' of loss. In more able males, a fall in score from 90 to 80 means they can no longer hop, and a fall from 50 to 40 fits with an inability to rise independently from the floor. In weaker males, a fall in score from 21 to 11 means they lose the ability to stand still. This fits with our clinical understanding of the hierarchy of difficulty of items which we reported in our previous paper, in which we compared 'expert' physiotherapy opinion on item difficulty with the hierarchy calculated within the Rasch model.[9] This hierarchy of difficulty has held true within this larger analysis. Agreement between the two was high (Spearman's rho=0.8) and also highlighted the difficulty that experts found in fitting 'lifts head' as an item into the model, which continued to misfit in this analysis and was therefore removed from the total score.

This loss of function also fits with the known clinical progression of the disease, although the age at which this happens varies considerably. Able males either gain the ability to hop, or never gain the ability to hop or, if they do, this is one of the first skills to be lost. The inability to stand still is seen in males just as they are about to lose ambulation altogether.

The calculation of a linearized numerical value to signify clinically meaningful change is key to demonstrating change, regardless of a male's ability. This does not hold

true for ordinal-level total scores, a change in which is difficult to interpret and the meaning of which may change, especially at either end of the scale. Understanding that a one-point change has an equivalent meaning in measurement terms across the range of ambulatory function means that we are better able to compare different levels of ability in the same frame of reference. This methodology, transforming scores before analysis, has not been employed routinely by other groups seeking to establish relative responsiveness of their rating scales, although the benefits of such a technique have been reported.[26]

These findings do not affect the clinical utility of the NSAA. Clinicians can continue using the original 17-point scale, although, within a trial setting, a transformed score (with lifts head removed from the total) would be more appropriate. The NSAA would benefit from further analysis to enhance the precision of the transformed ruler and confirm these analyses in a larger population, particularly in younger males and adolescents who remain ambulant.

Our study has three main limitations. First, the study population, although sufficiently large for Rasch analysis, when broken into age groups for responsiveness analysis consisted of small sample sizes, especially in the case of the younger and older age groups, and it is possible that data from the same individual are represented more than once in the same age group or in more than one age group. However, the database is on-going, and, as already suggested, analysis could be extended. Second, statistics to quantify meaningful change are applied across many disciplines; however, the paired *t*-test statistic depends not only on the size of the observed change, but also on the sample size, which perhaps makes its application in assessing responsiveness limited.[12] Finally, effect size calculations based on group-level analyses (even though data had been transformed using Rasch) can be misleading, especially in the case of scales with poor psychometric properties,[21] although this is not the case for the NSAA. The fact that the calculated MID can be readily related to ambulatory abilities, often quoted as significant areas of loss or gain by individuals, supports the proposed MID.[27] It is also true that the analyses are driven by psychometric evaluation rather than statistical analysis. This could be addressed in the future.

Full reporting of the results of the Rasch analysis, which took place before transformation of the raw scores to a linearized measurement, was outside of the scope of this paper. In a previous analysis, the NSAA was shown to be a scientifically robust measure using modern psychometric methods[9]; however, the results of this expanded analysis in a larger longitudinal study have not been fully reported in this paper. Critical evaluation explored the removal of one item – lifts head, which did not fit the construct – and the issue of dependency between the right- and left-sided items, where minimal impact was seen on inflation of reliability statistics (person separation index). A satisfactory solution was developed which supported the transformation. (Details of this interim analysis can be obtained from the authors).

This study also illustrates the added value of Rasch-transformed data, which enhances subsequent analyses of responsiveness. The generation of interval-level scores gives confidence to the interpretation of change scores and enables the production of values associated with MID, which may assist future DMD research and strengthen the case for the NSAA's inclusion as a clinical endpoint within clinical trials. Regardless of a male's ability, a 10-point change across the scale has the same meaning and the associated change can be related to a specific loss or gain of function.

This preliminary demonstration of the ability of a Rasch transformed NSAA score to detect differences in current treatment regimens could be enhanced by analysis using a larger subset of data, in order to produce a more accurate linearized measurement. The application of Rasch-transformed scores and subsequent responsiveness analysis is a relatively new approach and would benefit from additional applications. This methodology could be applied to data collected as endpoints within clinical trials.

ACKNOWLEDGEMENTS

This study was conducted on behalf of the North Star clinical network for paediatric neuromuscular disease. The support of the Muscular Dystrophy Campaign to the North Star Network and of the MRC Neuromuscular Translational Research grant is gratefully acknowledged. We thank CERTUS Technology Associates Limited, who host and maintain the database. Francesco Muntoni is supported by the Great Ormond Street Children's Charity.

REFERENCES

1. Drousiotou A, Ioannou P, Georgiou T, et al. Neonatal screening for Duchenne muscular dystrophy: a novel semiquantitative application of the bioluminescence test for creatine kinase in a pilot national program in Cyprus. *Genet Test* 1998; **2**: 55–60.
2. Emery AE. Population frequencies of neuromuscular diseases. I. Amyotrophic lateral sclerosis (motor neurone disease). *Neuromuscul Disord* 1991; **1**: 323–5.
3. Bushby K, Finkel R, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet Neurol* 2010; **9**: 77–93.
4. Manzur AY, Kuntzer T, Pike M, Swan A. Glucocorticoid corticosteroids for Duchenne muscular dystrophy. *Cochrane Database Syst Rev* 2008: CD003725.
5. Ricotti V, Ridout D, Scott E, et al. Long-term benefits and adverse effects of intermittent versus daily glucocorticoids in boys with Duchenne muscular dystrophy. *J Neurol Neurosurg Psychiatry* 2013; **84**: 698–705.
6. Mazzone E, Martinelli D, Berardinelli A, et al. North Star Ambulatory Assessment, 6-minute walk test and timed items in ambulant boys with Duchenne muscular dystrophy. *Neuromuscul Disord* 2010; **20**: 712–6.
7. Scott E, Eagle M, Mayhew A, et al. Development of a functional assessment scale for ambulatory boys with Duchenne muscular dystrophy. *Physiother Res Int* 2012; **17**: 101–9.
8. Mazzone ES, Messina S, Vasco G, et al. Reliability of the North Star Ambulatory Assessment in a

multicentric setting. *Neuromuscul Disord* 2009; **19**: 458–61.

9. Mayhew A, Cano S, Scott E, et al. Moving towards meaningful measurement: Rasch analysis of the North Star Ambulatory Assessment in Duchenne muscular dystrophy. *Dev Med Child Neurol* 2011; **53**: 535–42.
10. Kinali M, Arechavala-Gomeza V, Feng L, et al. Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *Lancet Neurol* 2009; **8**: 918–28.
11. Mazzone E, Vasco G, Sormani MP, et al. Functional changes in Duchenne muscular dystrophy: a 12-month longitudinal cohort study. *Neurology* 2011; **77**: 250–6.
12. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; **53**: 459–68.
13. Lemieux J, Beaton DE, Hogg-Johnson S, Bordeleau LJ, Goodwin PJ. Three methods for minimally important difference: no relationship was found with the net proportion of patients improving. *J Clin Epidemiol* 2007; **60**: 448–55.
14. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; **56**: 395–407.
15. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; **37**: 469–78.
16. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; **27** (Suppl.): S178–89.
17. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; **10**: 407–15.
18. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific Quality of Life Questionnaire. *J Clin Epidemiol* 1994; **47**: 81–7.
19. Deyo RA, Inui TS. Toward clinical applications of health status measures: sensitivity of scales to clinically important changes. *Health Serv Res* 1984; **19**: 275–89.
20. Muscular Dystrophy Campaign. (2013) Clinical Databases. http://www.muscular-dystrophy.org/how_-we_help_you/for_professionals/clinical_databases (accessed 16 May 2013).
21. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? *J Neurol Neurosurg Psychiatry* 2010; **81**: 1044–8.
22. Andrich D. Rating scales and Rasch measurement. *Expert Rev Pharmacoecon Outcomes Res* 2011; **11**: 571–85.
23. Andrich D, Sheridean B. RUMM2030. Perth, WA: Laboratory Pty Ltd, 1997–2012.
24. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes* 2006; **4**: 70.
25. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care* 2003; **41**: 582–92.
26. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; **13**: iii.
27. Bushby K, Finkel R, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. *Lancet Neurol* 2010; **9**: 177–89.