Performance of Upper Limb module for Duchenne muscular dystrophy

ANNA G MAYHEW[1] | GIORGIA CORATTI[2] | ELENA STACY MAZZONE[2] | KATRIJN KLINGELS[3,4] | MEREDITH JAMES[1] | MARIKA PANE[2] | VOLKER STRAUB[1] | NATALIE GOEMANS[5] | EUGENIO MERCURI[2] | PUL WORKING GROUP*

1 The John Walton Muscular Dystrophy Research Centre, Institute of Genetic Medicine, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. **2** Child Neurology Unit e Centro Nemo, IRCCS Fondazione Policlinico Gemelli, Università Cattolica del Sacro Cuore, Rome, Italy. **3** Department of Rehabilitation Sciences, KU Leuven – University of Leuven, Heverlee; **4** Faculty of Rehabilitation Sciences, REVAL, Hasselt University, Hasselt; **5** Child Neurology, University Hospitals Leuven, Leuven, Belgium.

Correspondence to Anna G Mayhew at The John Walton Muscular Dystrophy Research Centre, International Centre for Life, Newcastle University and Newcastle Hospitals NHS Foundation Trust, Central Parkway, Newcastle upon Tyne NE1 3BZ, UK. E-mail: anna.mayhew@newcastle.ac.uk

***Members of the PUL Working Group are listed in the Acknowledgements.**

PUBLICATION DATA
Accepted for publication 26th July 2019.
Published online 19th September 2019.

ABBREVIATIONS

DMD	Duchenne muscular dystrophy
PUL	Performance of Upper Limb

AIM To report the differences between Performance of Upper Limb (PUL) versions 1.2 and 2.0, compare the measurement ability of the two versions, and compare their longitudinal performance in Duchenne muscular dystrophy.

METHOD Rasch analysis was performed on the dual data from three centres to confirm whether the two scales measure the same construct. Change scores in natural history for the different domains were compared for the two versions.

RESULTS Rasch analysis demonstrated that both versions measure the same construct and that the PUL 2.0 was a better fit to the construct of motor performance and better able to detect change at 12 months in all levels of ability than the PUL 1.2. This was also true when change scores were reviewed over 2 years.

INTERPRETATION Our results confirm that the PUL 1.2 and 2.0 versions detect change in all domains over 2 years. They also demonstrate that simplifying the original scoring of the PUL 1.2 for the revised PUL 2.0 maintains the validity of the construct and enhances the scale measurement qualities.

The Performance of Upper Limb (PUL) scale was designed specifically to measure upper limb motor performance across the spectrum of severity of Duchenne muscular dystrophy (DMD). This includes weaker ambulant male children where upper limb weakness becomes more apparent and non-ambulant patients with increasing levels of muscle weakness. The proximal to distal progression of muscle weakness typically observed in DMD is tested in the PUL through three domains – shoulder, mid, and distal – with each including items relating to activities of daily living that patients and clinicians identified as relevant. The PUL was developed by an international clinical outcomes group consisting of clinicians, scientists, patient advocacy groups, and industries who had identified a need for a scale to measure motor performance of the upper limb.[1,2] The development of the PUL followed several steps: systematic review and a preliminary study exploring the suitability of the existing measures to the application of a pilot version in a multicentric setting, with Rasch analysis of the preliminary results leading to a version known as the PUL 1.2.[3] This version has been shown to be reliable[4] and related to the 6-Minute Walk Test in ambulant patients.[5] In non-ambulant male children with DMD, it has also been found to be sensitive to the differences in the steroid regime,[6] and a related patient-reported outcome measure has been developed.[7] Since the original version was published, subsequent collection of longitudinal data in a larger cohort, further clinical testing, and Rasch analysis have led to the modification of the scale, with the revised version now known as the PUL 2.0. This has been shown to detect significant change over a 2-year period in the total score as well as within dimensions.[8]

This paper aims to describe in more detail the changes that resulted in the PUL 2.0, report on the differences between PUL versions 1.2 and 2.0, confirm whether the two versions of the scale measure the same construct of upper limb motor performance, analyse how well they measure it, and finally compare the longitudinal scores on these two scales by domain (high-level shoulder, mid-level elbow, distal wrist and hand; see Appendix S1, online supporting information).

METHOD

Ethics

Italian data

The study was approved by the Ethical Committees of all the participating centres (Catholic University of the Sacred Heart; University of Messina; Bambino Gesù Children's Hospital; Second University of Naples; IRCCS Stella Maris; Istituto Giannina Gaslini; University of Torino; University of Padua; National Neurological Institute Casimiro Mondino Foundation; IRCCS Institute of Neurological Sciences, Bellaria Hospital; IRCCS Eugenio Medea; Neurological Institute Carlo Besta; Centro Clinico Nemo Milan).

Newcastle data

The North Star Clinical Network project follows Caldicott Guardian regulations and the information is entered in the database after parents' written informed consent is obtained.

Leuven data

This study was approved by the Ethics Committee of the University Hospitals Leuven.

Development of the PUL 2.0

Alterations to the PUL 1.2 scale have involved deleting some redundant items, simplifying the number of scoring options for most items to 0, 1, or 2, for two items to 0 or 1, adding one completely new item, and capturing some of the entry item details in two new items. These changes are summarized in Table 1.

These changes were suggested both by input from clinicians and by the outcome of a Rasch analysis, which showed that for some items the scoring options were too abundant and potentially not focusing on capturing progression of disease, but rather identified different strategies to compensate for disease progression. For example, patients who cannot turn an object over using full supination of the wrist (item Q) may use a strategy that completes the movement with a 'trick' (score 3), or do not complete the action (score 2). In the revised scale (item 19) this is now captured as one level and is described as 'uses compensation', to remove any possible ambiguity (score 1). It also sought to address any floor and ceiling effects and improve upon the targeting seen in the original 1.2 scale.

The PUL 2.0 (Table S1, online supporting information) includes an entry item to define the broad starting functional level and 22 items subdivided into shoulder level (six items), mid-level (nine items), and distal level (seven items). The entry item is based on a revised version of the Brooke score and ranges from 0 (no useful hand function) to 6 (full shoulder abduction with no weakness). This score is not included in the total PUL score as it is a summary of functional arm ability and contains scoring options that are measured in more detail by individual items. For weaker patients, a low score on the entry item means high-level items do not need to be performed, as they would not be achievable.

What this paper adds

- The original and revised Performance of Upper Limb (PUL) scales measure the same construct.
- Both scales detected change in all domains over 2 years.
- The PUL 2.0 enhances the measurement qualities of the scale.

Each dimension (shoulder, mid, distal) can be scored separately. In the PUL 2.0 there is a maximum score of 12 for the shoulder level, 17 for the mid-level, and 13 for the distal level. A total score of 42 can be achieved by adding the three level scores (compared with the PUL 1.2 with maximum scores of 16 for shoulder, 34 for mid-level, and 24 for distal, and a total of 74). Details of the training sessions and reliability studies have already been reported for the PUL 1.2.[3,4] New training sessions were performed for the new scale with similar levels of agreement.[8]

Rasch analysis

Dual data (PUL 1.2 and 2.0 conducted for each individual) were collected in a multicentre setting (Rome, Newcastle, and Leuven between September 2012 and December 2014) and Rasch analysis was conducted on 224 patients across three international sites in Europe. All patients or their caregivers gave informed consent to participate. The age of the patients ranged between 7 and 24 years (mean 13y, SD 4y 4mo, normally distributed). At baseline, 114 patients were ambulant and 110 non-ambulant. In total, 588 natural history assessments were included in the analysis with 224 conducted at baseline, 187 at year 1, and 177 at year 2.

Data were entered into Rasch Unidimensional Measurement Model RUMM2030 (standard version; RUMM Laboratory Pty Ltd, Perth, Australia) and, to determine the extent to which the two scales measured the same domain, the data were 'racked' (item ratings for each person were replicated horizontally) and analysed as if they were a single rating scale.[9] Analyses assessed: (1) item fit to the underlying construct, in this case motor performance of upper limb (all items should lie within a fit residual range of SD 2.5); (2) whether the scoring within each item reflected disease progression (ordering of thresholds); (3) reliability, as indicated by a Person Separation Index of more than 0.8,[10] which is equivalent to Cronbach's alpha;[11] and (4) targeting of the scales by plotting item location for them separately and comparing individual person–item threshold distribution maps for any significant floor or ceiling effects or gaps in the continuum of measurements in the two scales.[12,13]

Comparison of the PUL 1.2 and 2.0

To compare longitudinal changes collected with both the PUL 1.2 and 2.0, an Italian subset of the above data set, consisting of longitudinal data over 2 years (September 2012–February 2014) on 187 male children and young male adults, was reviewed by domain for change scores.

Table 1: Summary of changes from the Performance of Upper Limb (PUL) 1.2 to 2.0

PUL 1.2: items listed as letters	Score range of each item	PUL 2.0: items listed as numbers	Score range of each item	Summary of changes
Item A: Entry item – not included in total score	0–6	Entry item: A	0–6	Remains unchanged
Item B: Shoulder abduction – shoulder height	0–4			Removed as repetitive and ability level over measured
Item C: Shoulder abduction – above shoulder height	0–4			As above
		Item 1: Shoulder abduction both arms above head	0–2	Scoring based on entry item: 2, score 6; 1, score 5; 0, score 4 or less
		Item 2: Shoulder abduction to shoulder height	0–2	Score similar to level 4 on entry item: 2, able; 1 with compensation; 0, unable
Item D: Shoulder flexion to shoulder height	0–4	Items 3 (no weights) and 4 (500g): Shoulder flexion to shoulder height	0–2	Captured over two items. Simplified scoring with more distinctive difference between weights and clarified scoring: 2, able; 1, with compensation; 0, unable
Item E: Shoulder flexion above shoulder height	0–4	Items 5 (500g) and 6 (1kg): Shoulder flexion above shoulder height	0–2	Simplified with more distinctive difference between weights and clarified scoring: 2, able; 1, with compensation; 0, unable
Item F: Hands to mouth	0–3	Item 7: Hands to mouth	0–2	Same, simplified scoring: 2, able; 1, with compensation; 0, unable
Item G: Hands lap to table	0–3	Item 8: Hands lap to table	0–2	Same, simplified scoring: 2, able; 1, with compensation; 0, unable
Item H: Move weight	0–5	Items 9, 10, and 11: Move weight on table (100g, 500g, 1kg)	0–2	Same as H but scored over three items: 2, able; 1, with compensation; 0, unable
Item I: Move light cans – timed	0–5			Removed as source material difficult and similar to heavy cans
Item J: Move heavy cans – timed	0–5	Item 12: Lift heavy can diagonally	0–2	New item to capture lift and reach across body. Single can, no time
Item K: Stack light cans – timed	0–4			Removed as source material difficult and similar to heavy cans
Item L: Stack heavy cans – timed	0–4	Item 13 (three cans) and 14 (five cans). No time	0–2	Same test but scoring reorganized over two items
Item M: Remove lid		Item 15: Remove lid	0–1	Same: 1, able; 0, unable
Item N: Tearing paper	0–4	Item 16: Tearing paper	0–2	Same, simplified scoring: 2, folded into four; 1, folded in half; 0, unable
Item O: Tracing path	0–4	Item 17: Tracing path	0–2	Same, simplified scoring: 2, able; 1, with compensation; 0, unable
Item P: Push on light	0–3	Item 18: Push on light	0–2	Same, simplified scoring: 2, able permanently; 1, momentarily; 0, unable
Item Q: Supination	0–4	Item 19: Supination	0–2	Same, simplified scoring relates to supination ability when holding a small light
Item R: Pick up coins	0–3	Item 20: Pick up coins	0–2	Same, simplified scoring: 2, six coins; 1, one coin; 0, unable
Item S: Number diagram	0–3	Item 21: Number diagram	0–2	Same, simplified scoring: 2, able; 1, partly; 0, unable
Item T: Finger pinch	0–2	Item 22: Pick up 10g finger pinch	0–1	Same, simplified scoring: 1, able; 0, unable

These data for the PUL 2.0 have recently been reported.[8] In this paper the differences between baseline and 12 months and additionally 24 months were calculated for both scales and compared using a paired *t*-test (Stata version 15; StataCorp, College Station, TX, USA). The level of significance was set at $p<0.01$, to account for multiple testing. In addition, as the two PUL scales are measured on different numeric scales, we calculated 1- and 2-year standardized mean change scores for total PUL scores for both the PUL 1.2 and 2.0.

RESULTS

Rasch analysis of the two scales is summarized in Table 2. Both scales showed good reliability with high Person Separation Index (0.95), and good person fit. Item fit was improved for the PUL 2.0, as was the overall fit of the items to the construct as demonstrated by a higher percentage of items with ordered thresholds and a fit residual inside the recommended range and non-significant χ^2 probability.[14]

Targeting of the scales as illustrated by the relative item locations in Figure 1 and the person–item location distribution in Figure 2 demonstrated that the PUL 2.0 was better able to capture higher functioning individuals with the potential to reduce ceiling effect. It also showed that the range of items had a better spread, with fewer gaps between items; the PUL 2.0 is therefore more likely to capture change across the whole range of abilities.

Comparison of longitudinal data

Comparison of the changes observed between baseline and 12 or 24 months using the two scales showed that statistically significant changes were observed using both scales (they performed in a similar way). Our analysis demonstrated that a difference in the total score was observed between baseline and 12 months in the whole cohort, namely regardless of ambulatory status ($p<0.001$), and in the non-ambulant subgroup ($p<0.001$) with both scales (Table 3 and Table S2, online supporting information).

A difference was also observed between baseline and 24 months in the whole cohort ($p<0.001$) and in both ambulant ($p<0.001$) and non-ambulant subgroups ($p<0.001$) with both scales. Details of the subscores are also reported in Table 2. Furthermore, the standardized total scores for the PUL 1.2 and 2.0 are presented in Table S1, supporting the idea that there is no impactful reduction in sensitivity between the old and new scales.

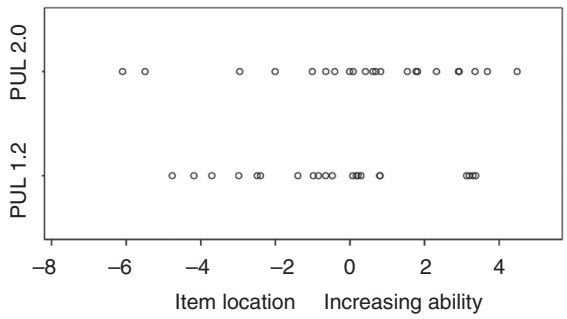

Figure 1: Comparative item location for individual items placed on common metric for the Performance of Upper Limb (PUL) 1.2 and PUL 2.0. PUL 2.0 item location demonstrates a better range and an improved spacing along the horizontal axis, which will mean measurement is more accurate and better captures all levels of ability. The upper line of dots representing the PUL 2.0 items extends beyond the range of PUL 1.2 dots both to the right (more ability) and to the left (less ability).

DISCUSSION

The need for a revision of the PUL 1.2 became apparent after its use in clinical and research settings. We wanted to address some of its limitations related to the redundancy of items and to the presence of a ceiling effect, and we aimed to improve scoring options. The changes from versions 1.2 to 2.0 were agreed on the basis of both Rasch analysis and input by clinical experts. A new cohort of patients was assessed using the revised version to ascertain the effect of the proposed changes and to verify whether the new version had better psychometric properties compared with the PUL 1.2, without losing sensitivity.

Our analysis of data in a large cohort of male children with DMD has confirmed that the PUL 2.0 has the same construct as the PUL 1.2, but as a measurement tool it has improved item fit to the model. Reliability, as measured by the Person Separation Index, has been maintained.

The logical progression of the scoring options has also improved, and from a targeting perspective the PUL 2.0 can measure more able individuals with DMD (Fig. 2). This is important as one of the concerns related to the PUL 1.2 was the risk of a ceiling effect for many patients with DMD, even at the time when they were losing ambulation and upper limb weakness might have been be clinically evident. In the PUL 2.0, the floor effect was negligible despite deleting some of the distal grip items; however, the measurement of the strongest individuals could be improved. This is not surprising, as many of those patients assessed in this study were ambulant with well-preserved arm function (37% had an entry item score of 6, the top score) and in older individuals (who are not limited in scoring by developmental) a top score at this point might negate the need to perform the rest of the scale. The ceiling effect seen in the PUL 1.2 may have been reduced by the inclusion of item 1 in the PUL 2.0, which captures this high level of function in the total score. Future work could examine the relation of the North Star Ambulatory Assessment score to the PUL score and provide additional information on when ambulant individuals with DMD have progressed sufficiently (as assessed by the North Star Ambulatory Assessment) so that the PUL 2.0 could start to identify meaningful clinical decline.

Our results also show that deleting redundant items and modifying the remaining ones helped to improve the hierarchical response options of individual items within the context of all items. This was particularly true for items where the scoring options were excessive and perhaps not

Table 2: Summary Rasch results for the Performance of Upper Limb (PUL) 1.2 and 2.0

	Item fit	Person fit		Reliability PSI with extremes	Item fit	
	Mean (SD)	Mean (SD)	Overall item–trait interaction χ^2 value (df)		Ordered thresholds	Number of items with good fit
PUL 1.2	0.01 (3.58)	−0.17 (0.25)	189	0.95	7/21 (33%)	6/21 (28%) 10[a]/13[b]
PUL 2.0	−0.19 (2.41)	−0.14 (0.16)	198	0.95	12/22 (55%)	15/22 (68%) 17[a]/18[b]

Fit defined as [a]fit residual inside the recommended range (−2.50 to 2.50) and [b]non-significant χ^2 probability ($p<0.001$). PSI, Person separation index, a measure of reliability within the Rumm 2030 programme; df, degrees of freedom.

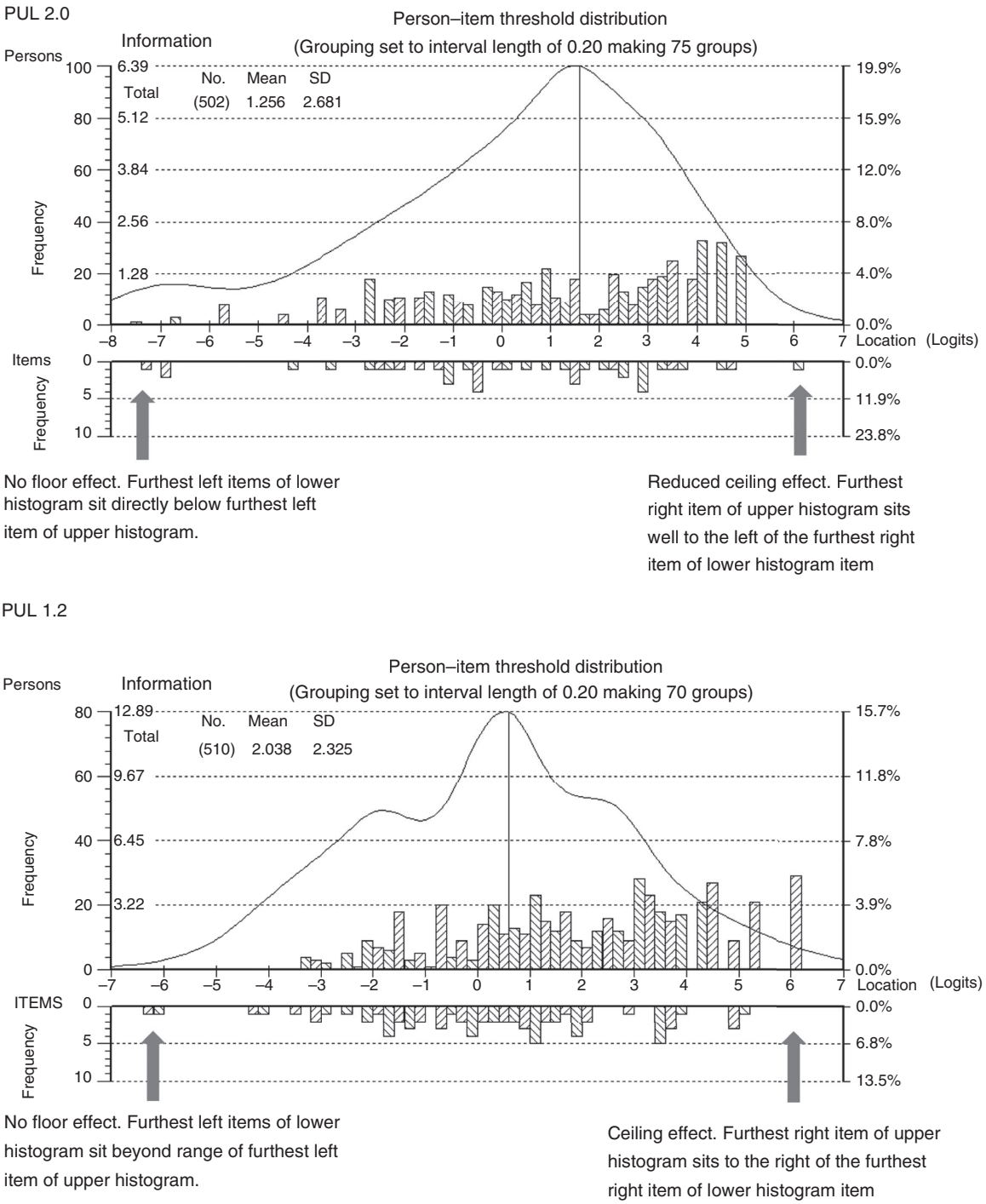

Figure 2: Person–item location distribution for the Performance of Upper Limb (PUL) 1.2 and PUL 2.0. Targeting of the patient sample (upper histograms) to the items (lower histograms). The figure shows the improved targeting between the distribution of person measurements and the distribution of item locations for the PUL 2.0 compared with the PUL 1.2. For the PUL 2.0, the ceiling/floor effects are minimal as the range of the person measurements (upper histogram 'blocks') closely matched the item locations (lower histogram 'blocks').

capturing progression but rather different strategies to compensate for progression. One of the concerns from the health care professionals involved in the process of revising the PUL 1.2 was that reducing the number of scoring options, while appropriate for statistical purposes, could have decreased the possibility of detecting changes and therefore reduced the sensitivity of the scale. On the contrary, our results clearly showed that simplifying scoring options for many individual items still enables change to be captured across the whole range of the scale. Sensitivity

Table 3: Comparison of change score for the Performance of Upper Limb (PUL) 1.2 and 2.0 over 12 and 24 months

		Baseline PUL 1.2	12mo changes	24mo changes	Baseline PUL 2.0	12mo changes	24mo changes
Total scores	Mean	56.4	−2.1	−5.3	28.9	−1.3	−3.2
Whole cohort	Range	7–74	−23 to 9	−25 to 10	3–42	−19 to 6	−21 to 7
(*n*=177)	95% CI	53.6–59.2	−2.8 to −1.4	−6.4 to −4.3	27.0–30.7	−1.8 to −0.8	−3.9 to −2.5
	p		<0.001[a]	<0.001[a]		<0.001[a]	<0.001[a]
Total scores	Mean	69.8	−0.8	−3.1	38.4	−0.4	−2.1
Ambulant	Range	52–74	−23 to 7	−25 to 10	24–42	−19 to 6	−21 to 7
(*n*=87)	95% CI	68.8–70.9	−1.7 to 0.1	−4.6 to −1.7	37.5–39.3	−1.1 to 0.2	−3.1 to −1.0
	p		0.09	<0.001[a]		0.20	<0.001[a]
Total scores	Mean	43.5	−3.4	−7.5	19.7	−2.2	−4.4
Non-ambulant	Range	7–74	−21 to 9	−25 to 5	3–42	−18 to 6	−21 to 7
(*n*=90)	95% CI	39.7–47.4	−4.4 to −2.4	−8.8 to −6.1	17.4–22.0	−2.9 to −1.4	−5.3 to −3.4
	p		<0.001[a]	<0.001[a]		<0.001[a]	<0.001[a]

[a]For further information on standardized change score for total score for PUL 1.2 and PUL 2.0, see Table S1 (online supporting information). CI, confidence interval.

has not been lost and utility has been improved, which will also increase reliability of the scale's performance and consistency in its scoring across multiple centres.

In our 24-month longitudinal data using both the PUL 1.2 and 2.0, we were able to detect the progression of upper limb involvement both in ambulant and non-ambulant male children. Even if the magnitude of changes was different, as the PUL 2.0 has a lower maximum possible score (42) compared with the PUL 1.2 (74), the point change remains meaningful and easier for health care professionals to interpret, as any double counting (scoring the same level of ability more than once) was removed. The difference between baseline and 12 or 24 months was always concordant both in the whole cohort and in the ambulant and non-ambulant subgroups (Table S1).

In conclusion, this study clearly supports the use of the PUL 2.0 as a functional outcome measure, without excluding the ongoing use of the PUL 1.2. Studies in larger cohorts are needed to further validate and explore the statistical properties of the PUL 2.0, including the assessment of the minimal clinical important difference, which was not established in this cohort.

ACKNOWLEDGEMENTS

The members of the PUL Working Group are as follows: Valeria Ricotti, Francesco Muntoni, Deborah Ridout, and Victoria Selby. VR is currently employed by Solid Biosciences. FM is a member of the Rare Disease Scientific Advisory Group for Pfizer and has participated in scientific advisory board meetings for PTC, Sarepta, Summit, and Wave Therapeutics. UCL and Great Ormond Street Hospital are recipients of grants from Pfizer, Sarepta, Italfarmaco, Wave Therapeutics, and Summit regarding clinical trials (Muntoni PI). We acknowledge the contribution of the North Star Network (https://www.northstardmd.com), in particular Newcastle. The financial support of MDUK for the North Star Project (AM and FM) and of the MRC Translational Research Centre to UCL and Newcastle (MR/K501074/1) is gratefully acknowledged. This research was supported by the NIHR GOSH BRC and the NIHR Newcastle BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. The authors' conflicts of interest are available in Appendix S2 (online supporting information).

SUPPORTING INFORMATION

The following additional material may be found online:

Appendix S1: PUL 2.0 worksheet.

Appendix S2: Authors' conflicts of interest.

Table S1: Standardized change scores for total score for PUL 1.2 and PUL 2.0.

Table S2: Comparison of change score for the PUL 1.2 and 2.0 subscales over 12 and 24 months.

REFERENCES

1. Mercuri E, McDonald C, Mayhew A, et al. International workshop on assessment of upper limb function in Duchenne muscular dystrophy. Rome, 15-16 February 2012. *Neuromuscul Disord* 2012; **22**: 1025–8.
2. Mazzone ES, Vasco G, Palermo C, et al. A critical review of functional assessment tools for upper limbs in Duchenne muscular dystrophy. *Dev Med Child Neurol* 2012; **54**: 879–85.
3. Mayhew A, Mazzone ESE, Eagle M, et al. Development of the performance of the upper limb module for Duchenne muscular dystrophy. *Dev Med Child Neurol* 2013; **55**: 1038–45.
4. Pane M, Mazzone ES, Fanelli L, et al. Reliability of the performance of upper limb assessment in Duchenne muscular dystrophy. *Neuromuscul Disord* 2014; **24**: 201–6.
5. Pane M, Mazzone ES, Sivo S, et al. The 6 minute walk test and performance of upper limb in ambulant Duchenne muscular dystrophy boys. *PLoS Curr* 2014; **6**. https://doi.org/10.1371/currents.md.a93d9904d57dcb08936f2ea89bca6fe6
6. Pane M, Fanelli L, Mazzone ES, et al. Benefits of glucocorticoids in non-ambulant boys/men with Duchenne muscular dystrophy: a multicentric longitudinal study using the performance of upper limb test. *Neuromuscul Disord* 2015; **25**: 749–53.
7. Klingels K, Mayhew AG, Mazzone ES, et al. Development of a patient-reported outcome measure for upper limb function in Duchenne muscular dystrophy: DMD upper limb PROM. *Dev Med Child Neurol* 2017; **59**: 224–31.
8. Pane M, Coratti G, Brogna C, et al. Upper limb function in Duchenne muscular dystrophy: 24 month longitudinal data. *PLoS One* 2018; **13**: e0199223.
9. 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, ix–x, 1–177.

10. Andrich D. Distinctions between assumptions and requirements in measurement in the social sciences. In Keats JA, Taft R, Heath RA, Lovibond SH, editors. Proceedings of the XXIVth International Congress of Psychology. North Holland: Elsevier Science Publications BV, 1989: 7–16.76.
11. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; **16**: 297–334.
12. Hobart JC, Cano SJ, Zajicek JP, Thompson AJ. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007; **6**: 1094–105.
13. Boone WJ. Rasch analysis for instrument development: why, when, and how?. *CBE Life Sci Educ* 2016; **15**: rm4.
14. Wright BD, Masters GN. Rating Scale Analysis. Chicago, IL: Mesa Press, 1982.