Performance of the Upper Limb Module—A Reliable and Valid Evaluation for Chinese Patients with Duchenne Muscular Dystrophy

Yuen Yee Alice Chiu
The Duchess of Kent Children's Hospital
Chun Wai Lo
The University of Hong Kong
Chi Kuk Connie Hui
THe Duchess of Kent Children's Hospital
Wai Chong Susanna Choi
The Duchess of Kent Children's Hospital
So Lun Lee
The University of Hong Kong; Queen Mark Hospital
Hoi Shan Sophelia Chan (sophehs@hku.hk)
University of Hong Kong Li Ka Shing Faculty of Medicine https://orcid.org/0000-0002-2990-0163

Research article

Keywords: Hammersmith motor scale, Performance of Upper Limb Module for Duchenne Muscular Dystrophy, pulmonary function, reliability, validity

Posted Date: February 4th, 2020

DOI: https://doi.org/10.21203/rs.2.22633/v1

Abstract

Background

Duchenne muscular dystrophy is a genetic disease leading to progressive muscle weakness and degeneration. Effective assessment tool is needed to allow monitoring of progress to guide the management. This study assessed the reliability and validity of the Performance of Upper Limb (PUL) Module when used in patients with Duchenne Muscular Dystrophy (DMD).

Methods

Total thirty-three Chinese DMD patients were included. Twenty-five video-recorded PUL Module version 1.3 assessments were performed for the recruited patients with three raters evaluated the same recorded video for inter-rater reliability and evaluated the same performance one month later for intra-rater reliability. Construct validity was assessed correlating the PUL Module scores with the patients' age, their forced vital capacity (N=25) and their Hammersmith motor scale scores (N=25) performed on the same day.

Results

The intra-rater and inter-rater reliability (ICC 0.92 - 0.99), internal consistency (Cronbach's alpha 0.97 - 0.99) and known groups validity (AUC 0.97) of PUL module were excellent. PUL was negatively correlated with age (r = -0.912), and positively correlated with the forced vital capacity (r = 0.87) and the Hammersmith motor scale (r = 0.84). The findings confirm the high reliability and validity of PUL module, and its high clinical relevancy in monitoring the deteriorating upper limb motor performance that strongly correlated with the lung function and generalized motor performance as age increased in DMD.

Conclusion

This first study of PUL module in Chinese patients with DMD confirmed that it is a reliable valid tool to monitor clinical progress and outcome for DMD.

Introduction

Duchenne muscular dystrophy (DMD) is a genetic disease leading to progressive muscle weakness and degeneration affecting skeletal, heart and respiratory muscles, which currently has no known cure [1]. DMD is the most common inherited neuromuscular diseases and the prevalence is 1 in 10,000 males in our locality [2]. DMD patients start showing symptoms between the age of 2 and 5, followed by progressive muscle weakness and eventually becoming wheelchair bound by the age of 12 [3]. Heart and respiratory muscle problems are also observed in the teen years [3]. Due to the advancement of cardiorespiratory interventions in recent years, the lifespan of DMD patients has increased to 30 – 40 [4, 5]. Previous researches showed that corticosteroid therapy can improve strength, prolong ambulation, improve cardiopulmonary function and increase survival and quality of life of DMD patients [6, 7].

The recent rapid development of approved therapies and many ongoing clinical trials highlighted the need to have reliable clinical outcome measures. The development of Performance of the Upper Limb module (PUL) for Duchenne muscular dystrophy[8, 9] was first established in 2013. There is also an updated version of PUL 2.0 [10].

The objective of this study was to investigate the psychometric properties the internal consistency, and validityf PUL when used in Chinese for DMD. DMD involves the deterioration of cardiopulmonary functions and motor performance over time [11-13], thus monitoring is crucial to assess the disease progression over time as well as the treatment response [11-15]. Forced vital capacity usually expressed as a percentage of the predicted value (FVC%) declines almost linearly in DMD patients from ages 5 to 24 [16, 17]; while Hammersmith motor ability score decreases proportionally to age in DMD patients from 5 to 12 years old [18, 19]. We also aimed to study the construct validity by comparing the PUL scores with the FVC% and the Hammersmith motor ability scores to assess the correlation.

Materials And Methods

Study Design

This is an observational study using a cross-sectional design with repeated measurement undertaken.

Participants

All participants have genetically confirmed DMD and undergone regular follow-ups under our Paediatric Neuromuscular Disorder Program and Pulmonary Rehabilitation program in our hospital. The study period was from 2016 to 2019. The inclusion criteria for patient recruitment included Chinese, genetically and clinically compatible DMD, aged 4 or above, and able to understand and follow instructions. Exclusion criteria included recent orthopedic interventions for the upper limbs, poor eyesight despite corrective lens, and recent acute medical conditions affecting their general status. Ethical approval was obtained from the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Written informed consent was obtained from patients aged 18 or above, or from parent or guardian for patients aged under 18.

Sample Size Calculation

All sample size estimations were performed based on an alpha level of 0.05, a power of 0.8. For establishing intra-rater and inter-rater reliability, the participants were evaluated twice. For a high test-retest and inter-rater reliability (ICC= or > 0.8) for the test, a minimum sample size of 25 individual assessment would be required to detect significant changes. Since DMD progresses from the ambulatory to non-ambulatory stage from about age 10 onwards, the ratio of ambulatory vs non-ambulatory participants was adjusted with a ratio of 1:2.

Qualification and Training of Assessors

All raters (rater A, B and C) were registered physiotherapists with over 10 years of clinical experience in paediatric patients. All were trained in the administration of the PUL for DMD through studying the instruction manual. Prior to data collection, all raters were engaged in a pilot trial, in which 2 individuals with DMD were assessed. The ratings were compared, and any discrepancies in ratings were thoroughly discussed until consensus was reached.

Measurements

For the reliability study, twenty-three patients participated the study with two DMD individuals had assessments at both the ambulatory phase and later the non-ambulatory phase. There was a total of 25 individual PUL for DMD assessment. Information on demographic data, physical and neurological examination, and steroid usage, was collected from systemic review of medical records. To study the reliability, each patient underwent the PUL for DMD

assessment conducted by one of the three registered physiotherapists with the entire individual assessment process recorded. The other two raters who did not perform the assessment watched the video of the patient PUL assessment and gave the scoring (inter-rater reliability). One month later, the three raters who were blinded about the participants data gave the scoring individually again (intra-rater reliability).

For the construct validity study, thirty-three DMD patients participated, with different number of DMD patients participated in each of the correlation analysis. The validation process was conducted by comparing the PUL for DMD with the age, the forced vital capacity (FVC % predicted value) and the Hammersmith Functional Motor Scale that were performed on the same day of the PUL assessment. The serial PUL for DMD scores changes with age changes were also analyzed.

The PUL for DMD (Version 1.3) was used. The test consists one entry item, the high-level shoulder dimension with 4 items, the mid-level elbow dimension with 9 items and the distal-level wrist and hand dimension with 8 items. The item score range is different for individual item ranging from 0 to 6. Total score ranges from 0 to 74. Serial measurement of the PUL at different ages for the 23 DMD patients during their subsequent clinical follow-up were also included for analysis.

Forced vital capacity (FVC% predicted value) which measures the total amount of air a person can exhale during a forced breath, assesses the cough effort. As part of the pulmonary function test, FVC% was obtained by spirometry (COSMED, Pony FX, Rome, Italy.) with the subject in sitting position and the data was taken on the same day of the PUL for DMD assessment.

Hammersmith Functional Motor Scale [20]is a validated test to evaluate the child's ability to perform various motor activities. The performance was rated by the registered physiotherapist who had performed the PUL scale for the DMD patient on the same day. The scale consists of 20 items assessing rolling, sitting, standing, stairs, static and dynamic standing balance. Each item score ranges from 0 to 2. Total score ranges from 0 when all activities failed to 40 when all activities completed.

Data Analysis

IBM SPSS for Windows (version 20, IBM Corporation, Armonk, NY, USA) was used for statistical analysis unless indicated otherwise. The level of significance was set at $P \leq 0.05$ to reduce the probability of making a type I error due to the many variables involved. Demographic data was analysed by descriptive statistics (e.g., means and standard deviations).

Floor and Ceiling Effects

The proportion of individuals with the lowest and the highest possible scores for each test was examined. Floor or ceiling effects were considered to be substantial if the proportion was greater than 20%. The 20% cut off is commonly used in to define substantial ceiling effect in previous studies that assessed pyschomtric properties of different measurement tools for various patient populations. The coefficient of skewness for the distribution of balance scores was assessed by using Medcalc [21-23] (version 16.2, MedCalc Software bvba, Ostend, Belgium). A positive skewness value $\gamma_1 > 1.0$ or a negative skewness value $\gamma_1 < -1.0$ indicates substantial skewness, and may indicate a floor and ceiling effect respectively.

Item Difficulty and Item Discrimination Index

The item difficulty index and item discrimination index were used to evaluate the effectiveness of individual test items. As the item scores of PUL for DMD are polytomous, item difficulty was expressed as the item mean of the point scale. A good item will have an item mean close to half of the maximum. The minimum and maximum item mean bounds represent what was considered as the cutoff point for the item mean score being too low (i.e., difficult item) and too high (i.e., easy item) respectively [24]. A factor of 0.3 and 0.6 was used to compute the minimum and maximum item mean bound respectively. The number of categories for the items and whether the item responses begin at 0 must be considered when setting the minimum/maximum item mean bounds. For an item rated on a 4-point scale (0-3), a mean score ≤0.9 (3×0.3) based on a maximum score of 3 would be considered to be a difficult item and a mean score of ≥1.8 (3×0.6) was considered as an easy item. Items with a mean score between 0.9 and 1.8 are considered to have acceptable difficulty level. For an item rated on a 3-point scale (0, 1, 2), a mean score of ≤0.6 (2×0.3) was considered as a difficult item and a mean score of ≥1.2 (2×0.6) as an easy item. For an item rated on a 5-point scale (0-4), a difficult item has a mean score ≤1.2 (4×0.3) whereas an easy item should have a mean score ≥2.4 (4×0.6).

The item discrimination index was expressed as the correlation between the item and the total score (Pearson's product-moment correlation), which could range from 0 to 1. An item discrimination index value <0.4 indicated that the item was ineffective and may require further examination to determine if it can be edited in some way before discarding [25].

Reliability Analysis

The internal consistency of PUL for DMD was assessed using Cronbach's alpha (>0.8: excellent, 0.7-0.8: moderate, <0.7: poor) with subscale analysis. Intraclass correlation coefficients (ICC) were used to estimate the intra-rater ($ICC_{3,1}$) and inter-rater ($ICC_{2,1}$) reliability (poor: <0.40, adequate: 0.40≤ICC≤0.75, excellent:>0.75) [26]. Using the intra-rater reliability results, the standard error of measurement (SEM) value was derived from the formula: $SEM = Sx\sqrt{(1-r_{xx})}$, where Sx was the standard deviation of the test total score and r_{xx} was the reliability coefficient. The minimal detectable changes at the 95% confidence interval (MDC_{95}) were computed using the formula: $MDC_{95} = 1.96 \times SEM \times \sqrt{2}$ [27].

Validity Analysis

Known-groups validity was established by assessing whether PUL for DMD could significantly differentiate between ambulatory and non-ambulatory subjects. A Receiving Operating Characteristic curve (ROC) analysis was carried out to assess the ability of the test to classify the above patient groups, generating the area under the curve (AUC) values (outstanding discrimination: AUC ≥0.9; excellent discrimination: AUC=0.8-0.9; acceptable discrimination: AUC=0.7-0.8) [28]. The positive and negative likelihood ratios (+LR, -LR) were also determined. A LR value of 1.0 indicates the test is useless in discrimination. +LR > 2 and −LR < 0.5 can be considered as clinically important [27].

Construct validity was assessed by comparing the PUL for DMD with the forced vital capacity, and the Hammersmith motor scale and Pearson correlation coefficient was used to assess the relationship. A value of 0.2 would represent a very weak or no relationship, 0.2 to 0.4 would represent a weak relationship, 0.4-0.6 would represent a moderate relationship, 0.6-0.8 would represent a strong relationship and 0.8 to 1.0 would represent a very strong relationship [27].

Results

Twenty-three Chinese patients with DMD aged 5 to 32 participated in the reliability study and known groups validity analysis. Complete sets of data were obtained from all participants and there was no missing data. Two patients were studied both in the ambulatory and later at the non-ambulatory stage with a total of 25 measurements performed. Twelve of them were on oral prednisolone, while 11 of them were not.

There was no significant floor or ceiling for total PUL scores, as the proportion of participants who received the lowest possible score was <20% and none of the scales demonstrated a positive skewness value in the score distribution (Table 1). The proportion of difficult items was not significant for both the elbow level as well as the wrist and hand level dimension. Item total correlation or item discrimination index showed no ineffective item within PUL. (Table 2).

There is excellent intra-and-inter-rater reliability for total PUL and the three level scores with intraclass correlation coefficient (ICC) ranged from 0.92 to 0.99. Good internal consistency was found for the total and the level scores with Cronbach's alpha ranging from 0.97 to 0.99. The standard error of measurement and minimal detectable changes for total PUL are 2.78 and 2.49 respectively and ranged from 0.90 to 2.84 and 7.43 to 7.87 respectively for the three level scores. (Table 3).

The mean total PUL score was significantly lower in non-ambulatory group (37.06) as compared to ambulatory group (70.25) ($p<0.01$). ROC analysis was used to determine the optimal cut-off score and area under curve (AUC) values of the total PUL and three level dimensions in classifying ambulatory and non-ambulatory group. The AUC values were greater than 0.9 ($p<0.01$), indicating excellent known-groups validity for total PUL score. The +LR ratios were >2 and – LR ratios were <0.5, indicating that the discrimination was clinically important for total score. (Table 4)

Thirty-three patients participated in the construct validity study with the patient's age, their forced vital capacity and their Hammersmith Motor Scale performance correlated with their PUL Module scores performed on the same day. Total PUL scores were found significantly correlated inversely with age ($p<0.05$), i.e. the older the child the lower the PUL scores for the whole DMD group and the non-ambulatory subgroup, in which 25 assessments were performed in 23 patients with 2 patients studied both in the ambulatory and later at the non-ambulatory stage. Age at evaluation (years): mean=15.4, standard deviation=7.5, median =13.0). On the other hand, for the ambulatory state, positive correlation was noted for mid-level and distal-level scores with patient's age ($p<0.05$) suggesting these two scores increase with age in ambulant children with DMD. (Table 5 and Figure 1). The total and the three levels PUL scores of DMD patients measured at longitudinal follow-up with at least three consecutive time points (age at initial evaluation (years): mean=11.9, standard deviation=3.5, median=11.5) also showed significant correlation with age change ($p<0.05$) and possible slower upper limb motor decline rate for those still on steroid (Figure 2).

A statistically significant correlation was also found between total PUL score with FVC% ($p<0.05$) (number of patients=25) (age at evaluation (years): mean=13.7 standard deviation=6.1, median=11.7) and Hammersmith Motor Scale ($p<0.05$) (number of patients=25) (age at evaluation (years): mean=12.2, standard deviation=5.3, median=10.9) with Pearson correlation coefficient of 0.869 and 0.844 respectively confirming very strong correlation (Table 5 and Figure 3).

Discussion

The PUL scale for DMD was specifically developed to measure the upper limb performance for DMD of a spectrum of severity [8]. This study was the first to investigate its psychometric properties, reliability and validity when used in

Chinese DMD patients. Our findings confirmed that PUL scale for DMD is a highly reliable and valid scale for Chinese patients with DMD. The scoring pattern corresponds well with changes in disease progression from ambulatory to non-ambulatory stage. No floor or ceiling effect was observed in our patients. It is useful for disease monitoring as a continuum with no scale revision is required. Item analysis also showed no particularly difficult items for the three level assessments. The tool is easy to administer with an excellent intra-and-inter-rater reliability. Our study confirmed PUL scale for DMD is a reliable upper limb performance measure for Chinese patients with DMD.

We used version 1.3 of PUL as it was the available version from the start of our study. There is an updated version 2.0 in 2018[15]. Upon content comparison, some fine hand function items of wrist and hand dimension like thumb grip are deleted. This may reduce the sensitivity of detecting changes in hand function of patients in their advanced disease course. On the other hand, high level shoulder dimension test items in terms of bilateral shoulder movement are incorporated. This may better reflect patient's performance during early ambulation status. The grading system is being simplified for version 2.0. This can reduce total time taken to complete the whole test but earlier version will have a more comprehensive description of the patient's performance to detect gradual and subtle changes over time as the disease progress. Score comparison between different versions can be further investigated looking into above aspects.

Our study also found that PUL scale is effective in detecting early upper limb shoulder involvement in DMD patients still in the ambulatory stage with a mean shoulder score of 13.50 (SD 2.33) (Table 4) followed by a declining trend with age (Figure 1). In fact, subclinical involvement in latissimus dorsi, serratus anterior, shoulder rotators and deltoid had been detected in patients with full score in shoulder domain by upper limb muscle MRI [29] . Another study with 164 patients with DMD [30] had shown that the onset of shoulder involvement occurred in the ambulatory phase for patients with DMD and the shoulder scoring could range from 2 to 16 (mean 13.98, SD 3.04).

The FVC is typically progressively reduced in patients with DMD and serial FVC measurement has been found to be a reliable monitoring in assessing pulmonary function and disease progression [12, 13]. In our study, the FVC% was strongly correlated with the total scores of PUL for our DMD patients. The finding has been supported by the Lee HN et al. previous study [31] suggesting that DMD patients with strong upper limb performance have better pulmonary function and vis versa.

The Hammersmith functional motor scale (HFMS) is frequently used and the scale measures the overall motor function in both ambulant and non-ambulant patients with DMD that could be used as an outcome measure for clinical trials for Duchenne muscular dystrophy. In our study, we found a strong and statistically significant correlation with the total scorings of the PUL for DMD. Under the Hammersmith motor scale, the trunk and neck control, as well as lower limb control, are included in the assessment. A better upper limb function is expected to be associated with a better performance in the truncal strength and control. Compared with previous finding of the non-linear correlation of 0.49 between PUL and 6 Minute Walk Test (6MWT) on ambulant DMD boys [30], our findings showed Hammersmith functional motor scale have a higher and more linear correlation with PUL and can be applied to both ambulant and non-ambulant patients with DMD.

In summary our study findings support PUL scale is a reliable measure with good validity in the disease monitoring, reflecting the upper limb performance and also correlate well with the lung function and generalized motor performance status.

Limitations and future research directions

The current limited number of recruited patients, though not of a large number, is adequate to give a statistically significant result. We correlated PUL scale for DMD with only FVC representing the lung function. In the future, the correlation could be extended to include different lung function assessment parameters such as inspiratory muscle strength.

The current PUL scale is of version 1.3. This year in 2019, an updated version 2.0 has been established for patients with DMD [10] and the study confirmed that both PUL versions measure the same construct so the current study findings can be generalized to the use of PUL version 2.0 in addition we understand that the update version could be better in detecting the serial changes when checked in 12 months or 24 months interval.

Conclusion

PUL is a highly reliable, valid and easy to use tool to measure upper limb motor performance with good scoring distribution and no ineffective items for Chinese patients with DMD. The scoring pattern corresponds well with patient's age, pulmonary function measured by FVC % predicted and the motor progression measured by Hammersmith Motor Functional Scale in Chinese children with DMD. Following the standard of care requiring regular evaluation, PUL should be incorporated as part of the assessment panel in the monitoring of disease progress and as treatment outcome measure.

Abbreviations

AUC = area under the curve

DMD = Duchenne muscular dystrophy

FVC% = forced vital capacity % predicted

PUL = Performance of Upper Limb

ROC = Receiving Operating Characteristic curve

Acknowledgement

We would like to thank Ms. Elena Mazzone of the Department of Paediatric Neurology of the Catholic University of Rome, for her valuable advice and support when we first set up the PUL study for our patients. We would also like to thank all the participating families and their parents. Without their support this study will not be possible.

References

1. Bushby K, Finkel R, Birnkrant DJ, Case LE, Clemens PR, Cripe L et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. 2010;9(1):77-93.
2. Chan SH, Lo IF, Cherk SW, Cheng WW, Fung EL, Yeung WL et al. Prevalence and characteristics of chinese patients with Duchenne and Becker muscular dystrophy: a territory wide collaborative study in Hong Kong. Child neurology open. 2015;2(2):2329048X15585345.
3. Liang W-C, Wang C-H, Chou P-C, Chen W-Z, Jong Y-J. The natural history of the patients with Duchenne muscular dystrophy in Taiwan: a medical center experience. Pediatrics & Neonatology. 2018;59(2):176-83.

4. Passamano L, Taglia A, Palladino A, Viggiano E, D'AMBROSIO P, Scutifero M et al. Improvement of survival in Duchenne Muscular Dystrophy: retrospective analysis of 835 patients. Acta Myologica. 2012;31(2):121.
5. Kieny P, Chollet S, Delalande P, Le Fort M, Magot A, Pereon Y et al. Evolution of life expectancy of patients with Duchenne muscular dystrophy at AFM Yolaine de Kepper centre between 1981 and 2011. Annals of physical and rehabilitation medicine. 2013;56(6):443-54.
6. Moxley III RT, Pandya S, Ciafaloni E, Fox DJ, Campbell K. Change in natural history of Duchenne muscular dystrophy with long-term corticosteroid treatment: implications for management. Journal of child neurology. 2010;25(9):1116-29.
7. Matthews E, Brassington R, Kuntzer T, Jichi F, Manzur AY. Corticosteroids for the treatment of Duchenne muscular dystrophy. Cochrane Database of Systematic Reviews. 2016(5).
8. Mayhew A, Mazzone ES, Eagle M, Duong T, Ash M, Decostre V et al. Development of the P erformance of the U pper L imb module for D uchenne muscular dystrophy. Developmental Medicine & Child Neurology. 2013;55(11):1038-45.
9. Pane M, Mazzone ES, Fanelli L, De Sanctis R, Bianco F, Sivo S et al. Reliability of the Performance of Upper Limb assessment in Duchenne muscular dystrophy. Neuromuscular Disorders. 2014;24(3):201-6.
10. Mayhew AG, Coratti G, Mazzone ES, Klingels K, James M, Pane M et al. Performance of Upper Limb module for Duchenne muscular dystrophy. Developmental Medicine & Child Neurology. 2019.
11. McDonald CM, Abresch RT, Carter GT, Fowler JW, Johnson ER, Kilmer DD et al. Profiles of neuromuscular diseases. Duchenne muscular dystrophy. American journal of physical medicine & rehabilitation. 1995;74(5 Suppl):S70-92.
12. Phillips MF, Quinlivan RC, Edwards RH, Calverley PM. Changes in spirometry over time as a prognostic marker in patients with Duchenne muscular dystrophy. American journal of respiratory and critical care medicine. 2001;164(12):2191-4.
13. Sharma GD. Pulmonary function testing in neuromuscular disorders. Pediatrics. 2009;123(Suppl 4):S219-21.
14. Pane M, Fanelli L, Mazzone ES, Olivieri G, D'Amico A, Messina S 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. Neuromuscular Disorders. 2015;25(10):749-53.
15. Pane M, Coratti G, Brogna C, Mazzone ES, Mayhew A, Fanelli L et al. Upper limb function in Duchenne muscular dystrophy: 24 month longitudinal data. PloS one. 2018;13(6):e0199223.
16. Mayer O, Finkel R, Rummey C, Benton M, Glanzman A, Flickinger J et al. Characterization of pulmonary function in Duchenne muscular dystrophy. Pediatric pulmonology. 2015;50(5):487-94.
17. Mayer OH, Aliverti A, Meier T. Breathe Duchenne: what natural history studies tell us about the progression of pulmonary morbidity in Duchenne muscular dystrophy. Neuromuscular disorders: NMD. 2018;28(11):910.
18. Parreira SLS, Resende MBD, Peduto MDC, Marie SKN, Carvalho MS, Reed UC. Quantification of muscle strength and motor ability in patients with Duchenne muscular dystrophy on steroid therapy. Arquivos de neuro-psiquiatria. 2007;65(2A):245-50.
19. Parreira SLS, Resende MBD, Zanoteli E, Carvalho MS, Marie SK, Reed UC. Comparison of motor strength and function in patients with Duchenne muscular dystrophy with or without steroid therapy. Arquivos de neuro-psiquiatria. 2010;68(5):683-8.
20. Main M, Kairon H, Mercuri E, Muntoni F. The Hammersmith functional motor scale for children with spinal muscular atrophy: a scale to test ability and monitor progress in children with limited ambulation. European

Journal of Paediatric Neurology. 2003;7(4):155-9.

21. Tsang CS, Liao L-R, Chung RC, Pang MY. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. Physical therapy. 2013;93(8):1102-15.
22. Holmes W, Shea J. Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. Quality of Life Research. 1997;6(6):561-71.
23. McHorney CA, Ware Jr JE, Lu JR, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical care. 1994:40-66.
24. Guyer R, Thompson N. User's Manual for Iteman 4.3. Woodbury, MN: Assessment Systems Corporation. No part of this publication …; 2013.
25. Crocker L, Algina J. Introduction to classical and modern test theory. ERIC; 1986.
26. Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. Psychometrika. 1978;43(2):259-62.
27. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Pearson/Prentice Hall Upper Saddle River, NJ; 2009.
28. Hosmer DW, Lemeshow S. Applied logistic regression. Wiley New York; 2000.
29. Brogna C, Cristiano L, Tartaglione T, Verdolotti T, Fanelli L, Ficociello L et al. Functional levels and MRI patterns of muscle involvement in upper limbs in Duchenne muscular dystrophy. PloS one. 2018;13(6):e0199222.
30. Pane M, Mazzone ES, Sivo S, Fanelli L, De Sanctis R, D'Amico A et al. The 6 minute walk test and performance of upper limb in ambulant duchenne muscular dystrophy boys. PLoS currents. 2014;6.
31. Lee HN, Sawnani H, Horn PS, Rybalsky I, Relucio L, Wong BL. The performance of the upper limb scores correlate with pulmonary function test measures and Egen Klassifikation scores in Duchenne muscular dystrophy. Neuromuscular Disorders. 2016;26(4-5):264-71.

Tables

Table 1. Distribution of PUL Score.

PUL Level	Mean	Median	SD	Minimum	Maximum	Skewness (γ_1)	Floor Effect (%)[a]	Ceiling Effect (%)[b]
Total	47.68	63.00	27.71	0	74	-0.92*	16	8
Shoulder	6.24	3.00	6.43	0	16	NA	NA	NA
Elbow	22.48	31.00	14.19	0	34	NA	NA	NA
Wrist & Hand	18.80	24.00	9.49	0	24	NA	NA	NA

Abbreviations: NA, not applicable; PUL, Performance of the Upper Limb module.

*Significant $p<0.05$.

[a]Floor effect: proportion of participants who obtained the highest possible score of the scale.

[b]Ceiling effect: proportion of participants who obtained the lowest possible score of the scale.

Table 2. Item Analysis of PUL.

PUL Level	Item Number	Scale rating range	Item mean	Item scale correlation	Item total correlation/ Discrimination Index
Shoulder	B	5	1.64	0.99	0.79
	C	5	1.60	0.99	0.79
	D	5	1.64	0.99	0.81
	E	5	1.52	0.98	0.76
Elbow	F	4	1.96	0.97	0.94
	G	4	2.08	0.97	0.96
	H	6	3.16	0.98	0.96
	I	6	3.44	0.99	0.96
	J	6	3.32	0.98	0.93
	K	5	2.60	0.97	0.96
	L	5	2.36	0.93	0.93
	M	2	0.80	0.81	0.87
	N	5	2.76	0.87	0.86
Wrist & Hand	O	5	2.92	0.88	0.80
	P	4	2.24	0.89	0.89
	Q	5	3.04	0.98	0.91
	R	4	2.36	0.99	0.91
	S	4	2.40	0.97	0.84
	T	3	1.64	0.97	0.85
	U	3	1.60	0.98	0.87
	V	4	2.28	0.92	0.88

Abbreviations: PUL, Performance of the Upper Limb module.

Table 3. Reliability Analysis of PUL.

PUL Level	Item mean range (Intra-rater)	Intra-rater reliability ICC (3,1)	Item mean range (Inter-rater)	Inter-rater reliability ICC (2,1)	Cronbach's alpha	Standard error of measurement	Minimal detectable changes
Total score	46.5 - 46.9	0.99	44.3 - 47.9	0.95	NA†	2.78	2.49
Shoulder	6.6 - 6.8	0.98	5.6 - 6.5	0.92	0.99	0.9	7.87
Elbow	21.6 - 22.2	0.96	21.5 - 21.6	0.98	0.97	2.84	7.43
Wrist & Hand	17.2 - 18.3	0.92	18.1 - 18.3	0.99	0.97	2.68	7.71

Abbreviations: PUL, Performance of the Upper Limb module.

†NA: The Cronbach’s alpha was generated for each of the subscale instead.

Table 4. Known-Groups Validity: Distinguishing Between Ambulatory and Non-Ambulatory Group.

	Mean ±SD[a]	Mean ±SD[a]	Mean Difference (95% CI)	P value	Area Under Curve (95%CI)	Cut-off score	Sensitivity (95% CI)	Specificity (95% CI)	+LR	-LR
Comparison	Ambulatory (N=8)	Non ambulatory (N=17)								
Total	70.25 ± 2.32	37.06 ± 27.81	33.19 (12.6, 53.8)	<0.01*	0.97* (0.81, 1.0)	>65	100 (63.1, 100)	94.12 (71.3, 99.9)	14.87	0.13
Shoulder	13.50 ± 2.33	2.82 ± 4.57	-10.44 (-14.1, -6.8)	<0.01*	0.96* (0.80, 0.99)	>8	80.6 (63.1, 100)	82.35 (56.6, 96.2)	7.44	0.14
Elbow	32.88 ± 1.25	17.59 ± 14.89	-15.29 (-26.3, -4.3)	<0.01*	0.88* (0.69, 0.98)	>29	100 (63.1, 100)	70.59 (44.0, 89.7)	2.55	0.35
Wrist & Hand	23.88 ± 0.35	16.41 ± 10.78	7.46 (-0.51, 15.4)	<0.01*	0.69* (0.48, 0.86)	>19	100 (63.1, 100)	41.18 (18.4, 67.1)	1.65	0.27

Abbreviations: N, number of assessments; [a]SD, standard deviation (95%); CI, 95% confidence interval; +LR, positive likelihood ratio; -LR, negative likelihood ratio.

*Statistical significance <0.01.

Table 5. Correlation Between PUL Scores and Age, FVC% And Hammersmith Motor Ability Score.

Measure	Pearson correlation coefficient (r)				
	Combined (N = 25)	Ambulatory (N = 8)	Non-ambulatory (N = 17)	Steroid (N = 14)	Non-steroid (N = 11)
PUL total score (correlates with age)	-0.912*	0.189	-0.896*	-0.551*	-0.842*
PUL High level (shoulder) score (correlates with age)	-0.784*	-0.209	-0.703*	-0.672*	-0.675*
PUL Mid-level (elbow) score (correlates with age)	-0.893*	0.761*	-0.884*	-0.530	-0.834*
PUL Distal level (wrist and hand) score (correlates with age)	-0.805*	0.712*	-0.799*	0.560*	-0.724*
-	Combined (N = 25)				
PUL total score (correlates with FVC%)	0.869*				
PUL total score (correlates with Hammersmith motor ability score)	0.844*				

Abbreviations: N, number of assessments; PUL, Performance of the Upper Limb module; FVC%, forced vital capacity % predicted.

* P-value < 0.05 for statistical significance.

Figures

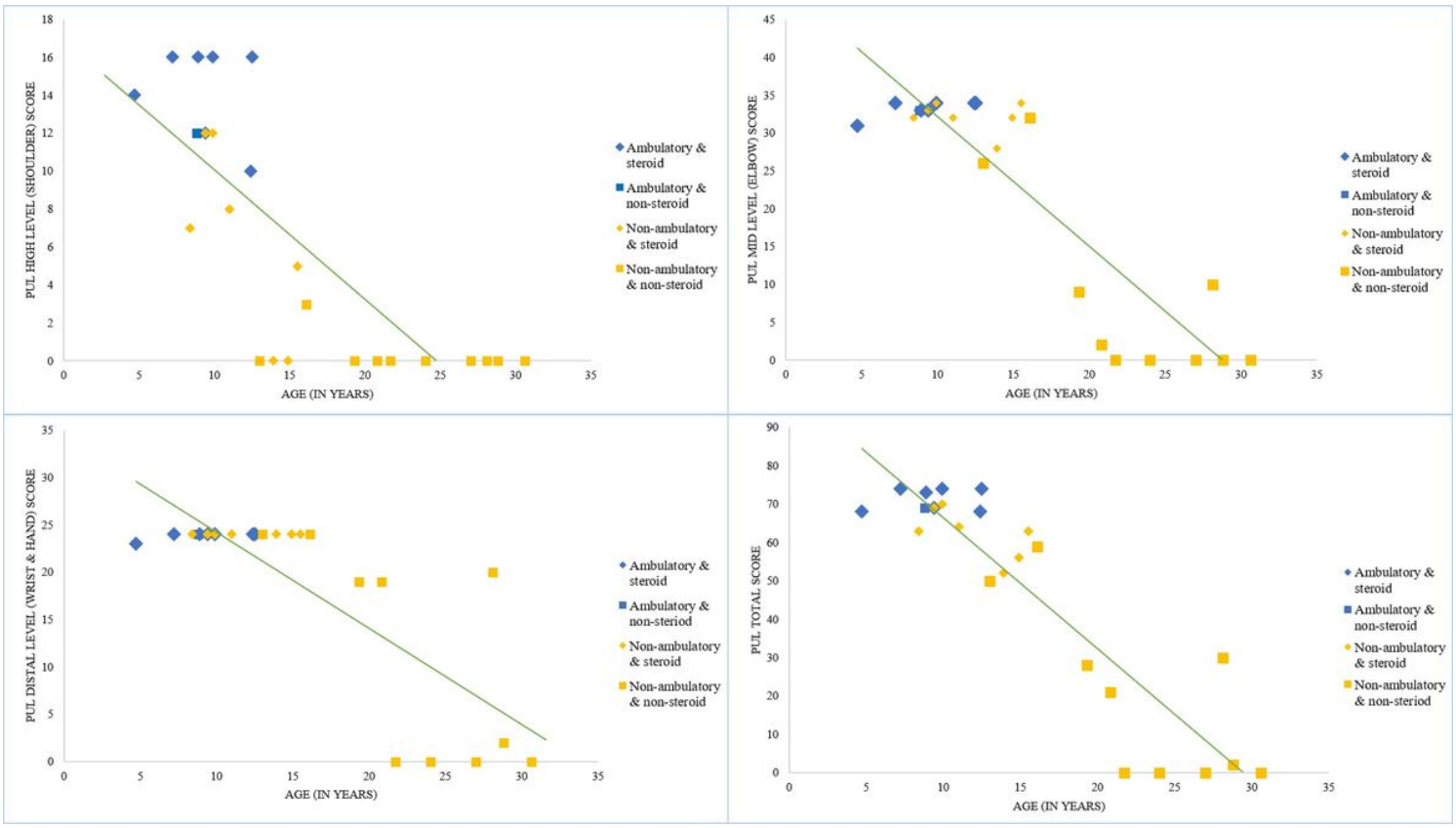

Figure 1

Scatterplot of the Performance of the Upper Limb (PUL) scores against age. Solid lines indicating linear regression. Number of patients=23. Number of assessments=25. Pearson correlation coefficients are r=-784, -0.893, -0.805 and -0.912 for correlations between age and PUL high level score, mid-level score, distal level score and total score respectively. All P-values are P<0.05.

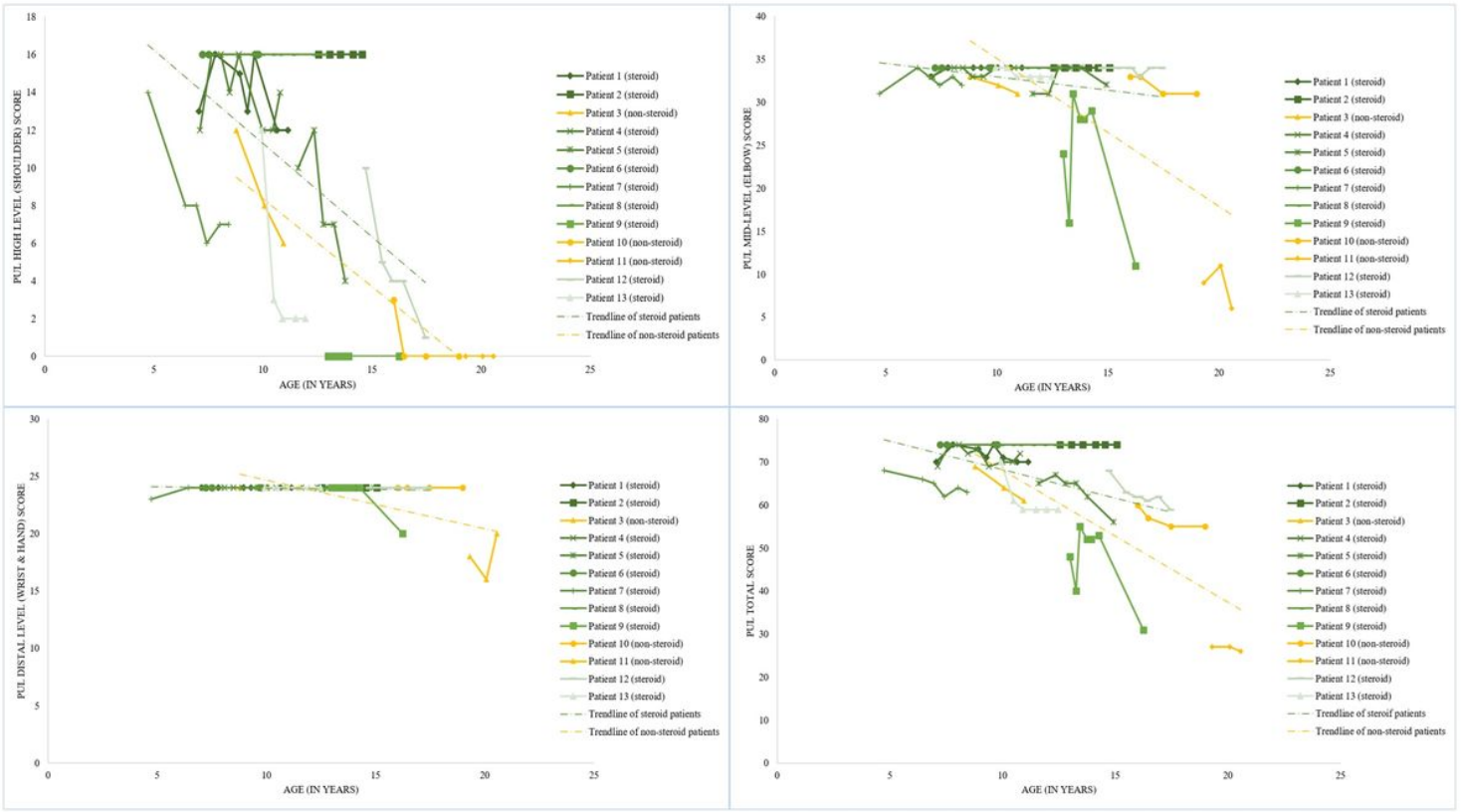

Figure 2

Scatterplot of the Performance of the Upper Limb (PUL) scores against age (serial time points). Solid lines indicating linear regression. Number of patients=13. Number of assessments=77. Pearson correlation coefficients are r=-0.603, -0.796, -0.452 and -0.636 for correlations between age and PUL high level score, PUL mid-level score, PUL distal level score and PUL total score respectively. All P-values are P<0.05.

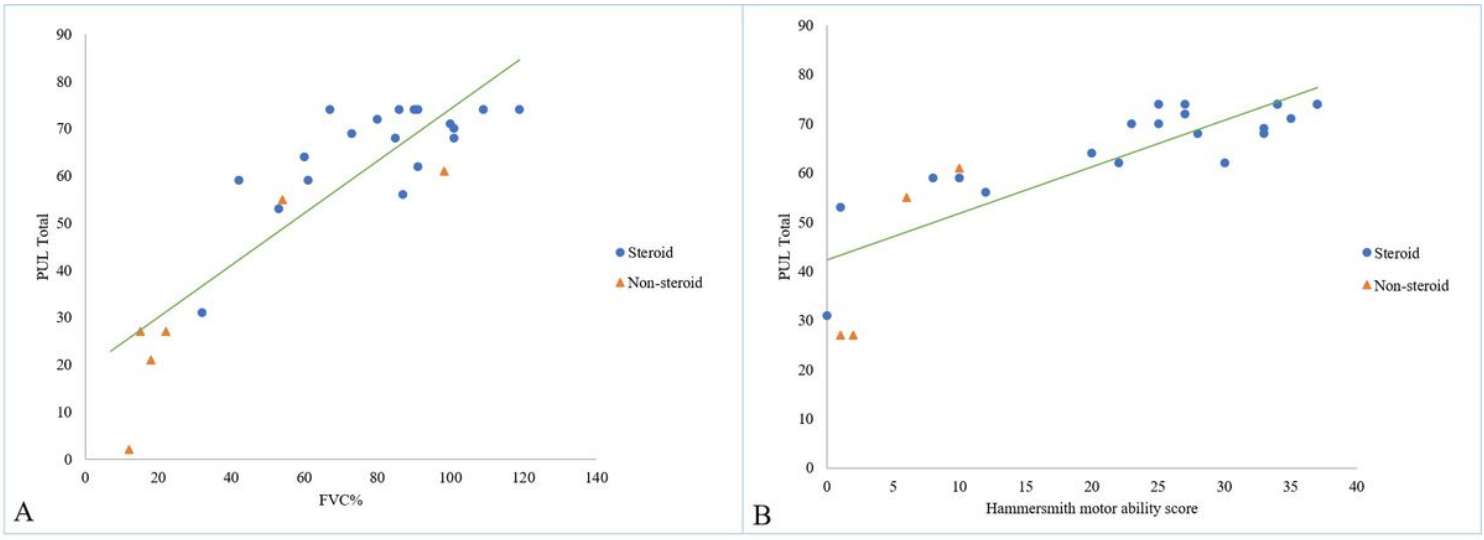

Figure 3

Figure 3a. Scatterplot of the Performance of the Upper Limb (PUL) total score against Forced Vital Capacity percent predicted (FVC%). Solid line indicating linear regression. Number of patients=25. Number of assessments=25. Pearson correlation coefficient is r=0.869 and P-value is P<0.05. Figure 3b. Scatterplot of the Performance of the

Upper Limb (PUL) total score against Hammersmith motor ability score. Solid line indicating linear regression. Number of patients=25. Number of assessments=25. Pearson correlation coefficient is r=0.844 and P-value is P<0.05.