Rasch analysis to evaluate the motor function measure for patients with facioscapulohumeral muscular dystrophy

Karlien Mul[a], Corinne G.C. Horlings[a], Catharina G. Faber[b],
Baziel G.M. van Engelen[a] and Ingemar S.J. Merkies[b,c]

Patient-relevant outcome measures for facioscapulohumeral muscular dystrophy (FSHD) are needed. The motor function measure (MFM) is an ordinal-based outcome measure for neuromuscular disorders, but its suitability to measure FSHD patients is questionable. Here, we performed Rasch analyses on MFM data from 194 FSHD patients to assess clinimetric properties in this patient group. Both the total scale and its three domains were analyzed (D1: standing position and transfers; D2: axial and proximal motor function; D3: distal motor function). Fit to the Rasch model, sample-item targeting, individual item fit, threshold ordering, sex- and age-based differential item functioning, response dependency and unidimensionality were assessed. Rasch analysis revealed multiple limitations of the MFM for FSHD, the most important being a large ceiling effect and suboptimal sample-item targeting, which were most pronounced for domains D2 and D3. There were disordered thresholds for most items, often resulting in items functioning in a dichotomous fashion. It was not possible to remodel the MFM into a Rasch-built interval scale. Remodeling of domain D1 into an interval scale with adequate fit statistics was achieved, but sample-item targeting remained suboptimal. Therefore, the MFM should be used with caution in FSHD patients, as it is not optimally suited to measure functional abilities in this patient group. *International Journal of Rehabilitation Research* 44: 38–44 Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc.

[a]Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, [b]Department of Neurology, School of Mental Health and Neuroscience, Maastricht University Medical Center+, Maastricht and [c]Department of Neurology, Curaçao Medical Center, Willemstad, Curaçao, The Netherlands

Correspondence to Karlien Mul, MD, PhD, Department of Neurology (910), Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands
Tel: +31 (0)24 3616600; e-mail: karlien.mul@radboudumc.nl

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is a hereditary muscle disorder that affects the muscles of the face, shoulder girdle and upper extremity and often in later stages muscles of the trunk and lower extremities [1]. Knowledge on the pathogenic mechanism is progressing and clinical trials on therapeutic interventions are both ongoing and expected in the upcoming years [2]. Therefore, valid, reliable, sensitive and clinically relevant outcome measures are warranted. A number of FSHD-specific measurement instruments are currently being developed [3–5], but many other more generic clinical outcome measures are already in use. Before any of these outcome measures can be used in clinical trials for FSHD, evidence should be collected to support its accuracy and suitability to measure treatment effects in this specific population, fulfilling modern scientific requirements [6,7].

One widespread used functional outcome measure for neuromuscular disorders is the motor function measure (MFM) [8]. The MFM measures the severity of motor deficits in three dimensions: standing position and transfers, axial and proximal motor function, and distal motor function. Although this scale was demonstrated to be valid and reliable according to the classical test theory methodology in a cohort comprising the most common neuromuscular diseases [8,9], it was not specifically designed for FSHD patients. Consequently, not all items on the scale may be equally suited for this patient group because of the specific distribution of muscle weakness [10]. Only limited work has been done so far to assess whether the scale captures the entire clinical spectrum of FSHD [10,11]. In addition, the MFM has important limitations as has been reported for ordinal-based classical test theory constructed metrics [12–14].

Analysis according to the Rasch model can be used to evaluate the suitability of the MFM as an outcome measure for FSHD patients. The Rasch model is a psychometric model for analyzing ordinal data. It is based on the assumption that a patient with a high overall ability (a less severely ill patient) will have a higher probability of fulfilling any single task compared to a patient with a lower overall ability (a more severely affected patient) [15]. Through Rasch analysis, important assumptions of

measurement theory can be tested to evaluate the quality of the measurement properties of the MFM at scale and items level [16]. Importantly, Rasch analysis can be used to transform ordinal scores into interval scores. In contrast to ordinal scores, that only provide a structured order, interval scores provide a continuous value that enables the use of parametric statistical testing [12,14,17].

Rasch analysis has been applied successfully in the neuromuscular field. Rasch-built functional outcome measures were newly developed for, among others, Pompe's disease [18], myotonic dystrophy type 1 [19] and inflammatory neuropathies [20], and other scales were successfully modified to create interval scales, such as the MRC-gradation [21], the North Star Ambulatory Assessment in Duchenne muscular dystrophy [22] and the MFM in congenital myopathies [23].

In this study, we apply Rasch analysis to assess the measurement properties of the MFM as an outcome measure assessing functional ability in FSHD patients.

Methods
Data collection
Data of 203 FSHD patients, collected in 2014–2015 for a cohort study at the neurology department of the Radboud University Medical Center, the Netherlands, were used [24,25]. Genetically confirmed patients aged 18 years and older were included. Nine nonpenetrant gene carriers (individuals without signs or symptoms of FSHD) were excluded, leading to a final group of 194 patients.

Ethical approval
This study was conducted according to the principles of the Declaration of Helsinki (version October 2013) and in accordance with the Medical Research Involving Human Subjects Act (WMO). The study was approved by the regional medical ethics committee. All patients signed informed consent.

Motor function measure
The 32-item MFM is an examiner-reported scale that consists of three different dimensions or domains – D1: standing position and transfers (13 items); D2: axial and proximal motor function (12 items); D3: distal motor function (7 items). Response categories per item are: 0 'does not initiate movement or starting position cannot be maintained'; 1 'partially completes the exercise'; 2 'completes the exercise with compensations, slowness or obvious clumsiness' and 3 'completes the exercise with a standard pattern'. The total score is expressed as a percentage of the maximum possible score (a lower score indicating clinically more affected).

Clinical severity rating
The FSHD clinical score is a widely accepted scale of the severity of muscle weakness [26]. Scores are assigned to five different body regions (face, shoulders, arms, hips,

lower legs and abdomen). A total score that ranges from 0 to 15 is calculated by summing all regional scores. Higher scores indicate more severe weakness.

Rasch analysis
Rasch analysis was performed using Rasch unidimensional measurement methods software (RUMM2030) [27]. A comprehensive description of Rasch analysis can be found elsewhere [28].

A sample size of approximately 200 is suggested to provide a stable model, with a 99% confidence that the estimated item difficulty is within ±0.5 logits of its stable value [29]. There were no missing data. Analyses of both the total MFM and of each of its three domains were performed [8,30].

The likelihood-ratio test was highly significant for the total MFM and for all three domains ($P < 0.001$), excluding the use of the rating scale model that assumes equivalent thresholds across all items. The partial credit model was, therefore, set as default. P values corrected according to Bonferroni were considered statistically significant [31].

For each analysis, the overall model fit was assessed by a chi-square item-trait interaction statistic. A nonsignificant probability value indicates no substantial deviation from the model. Second, possible issues contributing to the misfit in individual items were assessed. Individual item fit was checked by two statistics: fit residual value exceeding 2.5, significant chi-square probability value after Bonferroni adjustments or a combination of both, indicate the deviation from the Rasch model expectations. Next, threshold ordering was checked. The threshold is defined as the ability level between two adjacent response categories where either category is equally probable [32]. Disordered thresholds indicate inconsistent use of response options between subjects, which occurs when respondents or examiners have difficulty discriminating between response options or categories within an item. Differential item functioning (DIF), a form of item bias, occurs when different groups of patients with an equal level of disability, respond in a different manner to an individual item. In this study, DIF was checked for sex and different age groups (three groups with roughly the same number of participants: <45, 45–60 and >60 years). Response dependency was also examined, that is, items relating with each other after accounting for their contribution to the latent (intended) trait. As per the prevailing literature inter-item, residual correlations of ≥0.3 indicate local dependency. Distribution of patients within class intervals, tests for unidimensionality through principal components analysis (proportion of significant *t*-tests when plotting most positively items against most negatively loaded items) and internal consistency reliability were monitored throughout the analyses.

Table 1   Statistics for Rasch analysis of the total motor function measure and the three separate domains

Scale	Item location (mean, SD)	Item fit residual (mean, SD)	Person location (mean, SD)	Person fit residual (mean, SD)	Item trait interaction	PSI	Unidimensionality t-test % (95% CI)	Ceiling effect[a]
					df; P value			
Total MFM	0±2.96	−0.46±1.46	3.53±2.52	−0.26±0.67	64; 0.0000	0.94	18.1% (14.8–21.4)	14%
D1: standing position and transfers	0±2.18	−0.19±1.15	1.02±4.23	−0.35±0.56	26; <0.0001	0.96	4.1% (0.6–7.7)	18%
D2: axial and proximal motor function	0±1.89	−0.67±1.15	4.10±2.38	−0.44±0.64	24; 0.003	0.82	5.1% (1.5–8.8)	30%
D3: distal motor function	0±3.26	0.12±1.10	6.81±1.55	−0.10±0.34	14; 0.222	0.54	5.8% (1.6–10.0)	45%

MFM, motor function measure; PSI, person separation index.
[a]Proportion of participants achieving the maximum score; SD.

Table 2   Individual item statistics for the total motor function measure scale ordered by item location

Item		Domain	Location	Fit residuals	Chi square prob	Thresholds	Ceiling effect[a]	DIF	Response dependency with
22	Place finger on 8 drawings	3	−6.76	−0.09	0.89	Disordered	98		Item 18: 0.390
18	Go round edge of cd	3	−5.40	−0.74	0.37	Disordered	89		Item 22: 0.390
17	Pick up and hold 10 coins	3	−4.67	0.27	0.10		93		Item 16: 0.537
20	Tear sheet of folded paper	3	−3.90	0.86	0.09		82		
16	Extend the elbow	2	−3.86	−0.58	0.34		81		Item 17: 0.537
19	Draw a series of loops	3	−3.76	3.29	0.00[b]		66		
14	Raise head from flexion	2	−3.48	−0.18	0.89	Disordered	98		Item 13: 0.301
23	Place hands on table	2	−2.06	−0.24	0.77		87		
15	Place hands on head	2	−1.83	0.99	0.07		33		Item 9: 0.453
21	Pick up ball and turn hand	3	−1.40	1.11	0.29		62		
13	Maintain seated position on chair	2	−1.03	−0.69	0.59	Disordered	88		Item 14: 0.301
1	Hold head for 5 s	2	−0.31	1.29	0.000[b]	Disordered	89		Item 2: 0.527
2	From supine, raise the head	2	−0.25	0.35	0.58	Disordered	86		Item 1: 0.527
3	Flex hip and knee >90°	2	−0.11	0.32	0.36	Disordered	82		
25	Maintain standing position	1	0.14	−2.22	0.02	Disordered	73		
5	Move hand to shoulder	2	0.39	0.49	0.32		49		Item 9: 0.353
10	Lean forward and touch ball	2	1.17	1.41	0.00[b]	Disordered	75		
12	Sit down from standing	1	1.18	−1.87	0.29	Disordered	57		Item 24: 0.393
9	Maintain seated position on mat	2	1.35	−1.26	0.59	Disordered	39		Item 5: 0.353; item 15: 0.453
4	Dorsiflex the foot	3	1.47	1.90	0.00[b]	Disordered	59		
24	Stand up from the mat	1	1.60	−2.93[b]	0.02	Disordered	49		Item 12: 0.393
6	Raise the pelvis	1	1.65	−1.89	0.47		56		
27	Touch the floor	1	1.73	−1.66	0.10	Disordered	63		
7	Roll to prone and free arms	2	1.91	−2.95[b]	0.15	Disordered	37		
26	Raise the foot	1	2.59	−2.05	0.02	Disordered	48		
29	Take 10 steps on a line	1	2.70	−0.99	0.10	Disordered	43		
8	From supine, sit up	1	2.72	0.53	0.04		17		
32	Squat from standing	1	3.25	−1.03	0.11	Disordered	42		
11	Stand up from the mat	1	3.36	−1.47	0.70		20		
28	10 steps on both heels	1	3.55	−1.24	0.04	Disordered	35		
30	Run 10 m	1	3.93	−2.27	0.00[b]		27	Age UN	Item 31: 0.314
31	Hop 10 times in place	1	4.14	−1.03	0.13	Disordered	28		Item 30: 0.314

Disordered thresholds: at least one of the response options is never the most likely option to be chosen.
DIF, differential item functioning; UN, uniform.
[a]Percentage of participants achieving the maximum score on the item.
[b]Significant deviation from the Rasch model.

Results
Study population
Ninety-five (49.0%) of 194 participants were male and the mean age was 51.6 years (±15.7 years). Patient-reported mean disease duration was 26.8 years (±17.5 years). The total clinical spectrum of FSHD was represented, from patients using a wheelchair ($n=46$; 23.7%) to minimally affected individuals. The mean FSHD clinical score was 7.2 ± 4.4 (range 0–15). The mean score on the total MFM was $78.3\% \pm 23.0$ (range 13.5–100%). Mean scores on the three domains were $63.6\% \pm 35.6$ for D1, $86.3\% \pm 20.3$ for D2 and $91.8\% \pm 11.4$ for D3.

Rasch analysis of the total motor function measure
The MFM did not meet the requirements to fit to the Rasch model for patients with FSHD (Table 1) and inspection of individual item function revealed numerous poorly fitting items (Table 2). A major concern was the poor targeting of items (Fig. 1). The mean person location was higher than the mean item location (3.525 vs. 0.000), indicating that the person's abilities were higher than the difficulty of the items. This fits with the person-threshold location distribution map showing a lack of more difficult items that are needed to evaluate patients with a high ability level (Fig. 1). It is also illustrated by high ceiling

Fig. 1
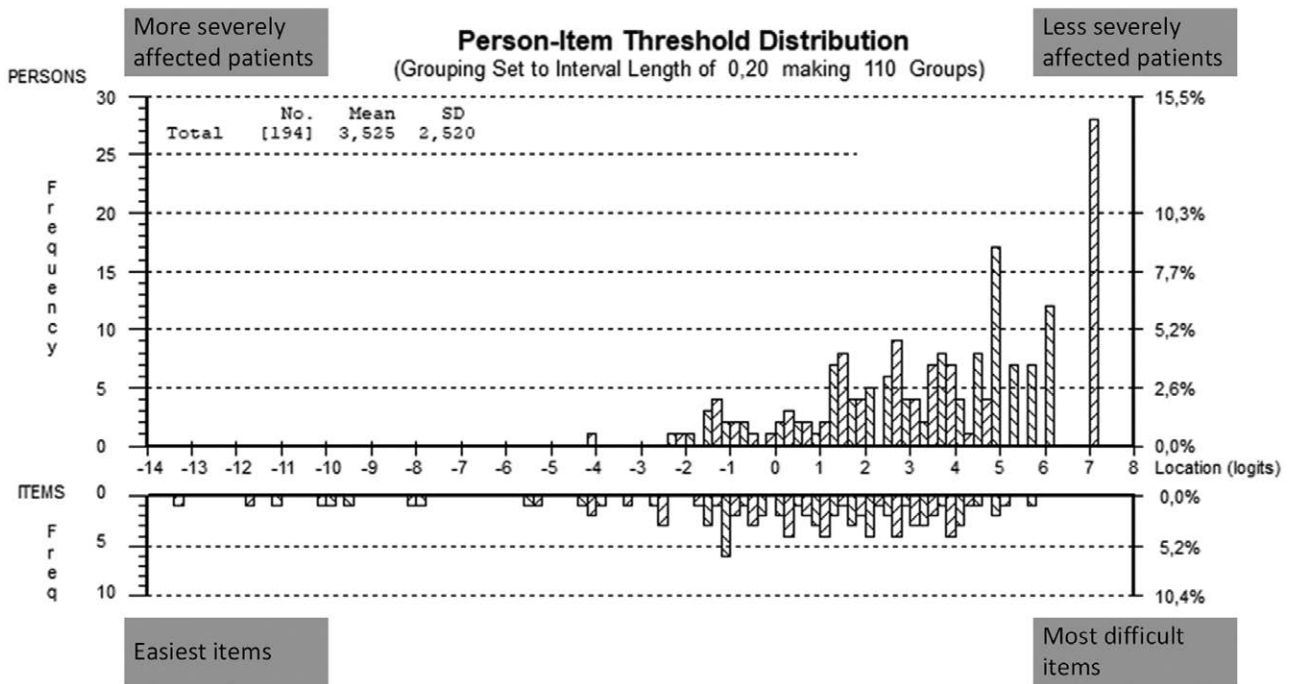
Person-threshold location distribution of the total motor function measure. Upper part of the graph shows distribution of person abilities, and lower part shows distribution of item difficulty.

Table 3 Statistics for Rasch analysis of domain D1 after remodeling

Scale	Item location (mean, SD)	Item fit residual (mean, SD)	Person location (mean, SD)	Person fit residual (mean, SD)	Item trait interaction	PSI	Unidimensionality t-test % (95% CI)	Ceiling effect*
					df; *P* value			
Rasch-built D1: standing position and transfers	0±2.78	−0.25±0.58	1.45±4.94	−0.27±0.50	24; 0.22	0.96	7.6% (3.9–11.3)	24%

PSI, person separation index. *Proportion of participants achieving the maximum score.

effects on the total score (14% of patients achieving the maximum score) and on multiple individual items (Table 2). Another major concern was the high proportion of items with disordered thresholds (20/32 items, 63%). For most of these items, there was a relative infrequency of response option 1 'partially completes the exercise'. Other violations of the Rasch model assumptions were DIF, response dependency and multidimensionality of the scale. Only one item showed uniform DIF: item 30 (run 10 m) was scored differently in individuals with an equal level of disability but from different age groups. Eight-item pairs showed response dependency (Table 2). In all except one pair, both items were part of the same domain. We tried to remodel the total MFM according to the Rasch model to improve the scale specifically for use in FSHD patients. However, after numerous attempts, we did not succeed in remodeling the MFM to a scale that sufficiently fit all Rasch model requirements, and

still being suitable for clinical application. Collapsing the response categories from four to three, taking into account the distribution of response options per category, did not restore the disordered thresholds for all items. Reducing the number of response categories increased the ceiling effect and the misfit in sample-item targeting even more. This resulted in a decreased discriminative power of the scale.

Rasch analysis of the three domains

In addition to the analysis of the total MFM, analyses were performed on its three domains. Statistics for analyses of the three domains are shown in Table 1.

The D1 domain (standing position and transfers) did not fit the Rasch model. The mean person location was higher than the mean item location (1.020 vs. 0.000) and there was a ceiling effect (18% of patients achieved the maximum score). Six items had disordered thresholds.

Fig. 2
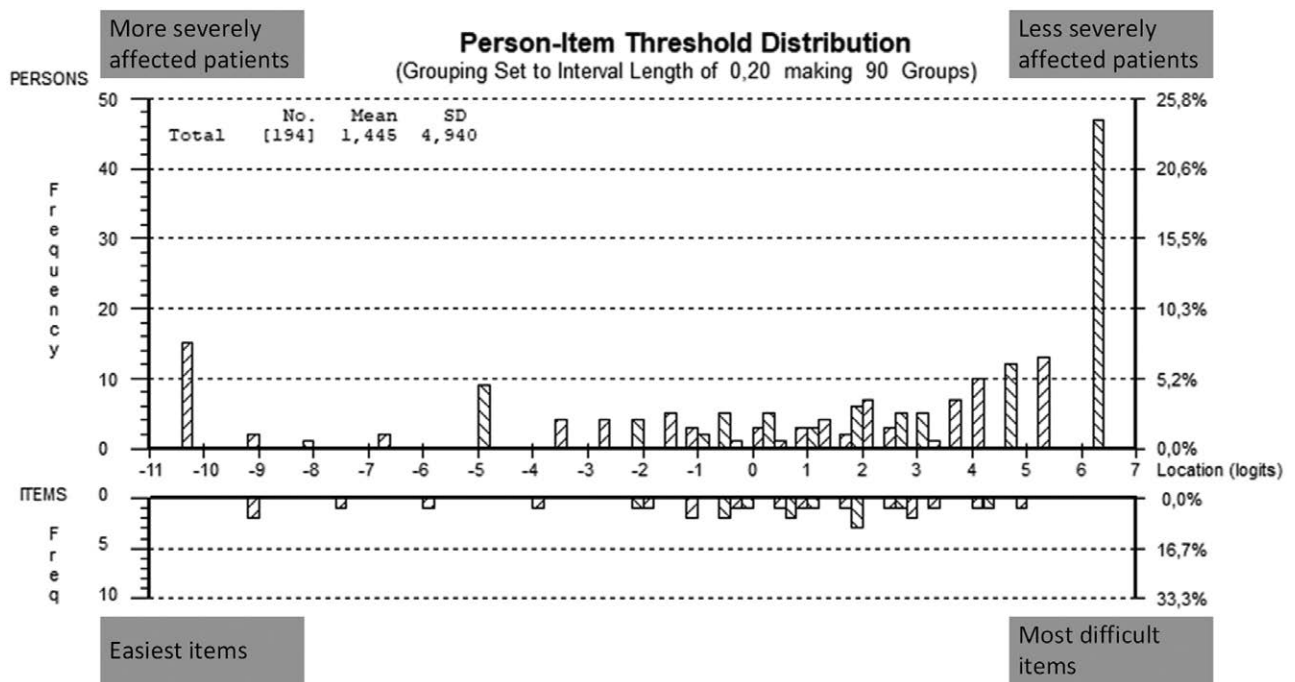
Person-threshold location distribution of the remodeled D1 domain (standing position and transfers). Upper part of the graph shows distribution of person abilities, and lower part shows distribution of item difficulty.

Items 8 (from supine, sit up) and item 30 (run 10 m) showed individual item misfit. Item 30 also showed DIF on age, with lower scores for individuals >60 years with an equal level of disability compared to younger individuals. No response dependency was found.

Rasch analysis was applied to remodel the D1 domain into an interval scale for FSHD patients. First, disordered thresholds were restored for items 27, 28, 29 and 32 by collapsing the number of response categories from four to three (0-1-2-3 to 0-1-1-2), taking into account the frequency distribution of the categories. One of the items that showed individual item misfit was removed (item 8, fit residuals 3.773 and chi square probability 0.0000). After removal, no DIF was found for age or sex and there were no items with response dependency. Two items were left with disordered thresholds: items 24 and 31. For both items, we assessed which rescoring option resulted in the best overall model, and subsequently rescored both items from 0-1-2-3 to 0-1-1-2. The new 12-item domain D1 showed an acceptable trend towards unidimensionality and fitted the Rasch model expectations (Table 3). However, fitting the D1 scale to the Rasch model, came at the cost of an increase in the mean person location from 1.020 to 1.445. Consequently, the ceiling effect increased to 24% of patients achieving the maximum score. The limited number of items' thresholds

resulted in a suboptimal sample-item targeting (Fig. 2). The range of patient abilities that could be measured by the items on domain D1 was too narrow and there were gaps between threshold locations of more than 1 logits.

Both domain D2 (axial and proximal motor function) and domain D3 (distal motor function) had very poor sample-item targeting with large ceiling effects (30% respectively 45% of participants achieved the maximum score) and very high mean person locations (4.096 and 6.806, respectively). As such, nearly all items were too easy for most patients and these domains did not provide information on the actual abilities of these patients. Indeed, person separation indexes were moderate and low (0.82 and 0.54, respectively) indicating that domains D2 and D3 were not able to sufficiently discriminate between individuals with different ability levels.

Both domains also did not fulfill other Rasch model assumptions. There were many items with disordered thresholds (8/12 items in D2 and 3/7 items in D3). Domain D2 also had two items with individual item misfit (items 9 maintain seated position on the mat and 16 extend the elbow). There was response dependency between items 1 and 2 (raise the head from supine and hold the head for 5 s) and between items 14 (maintain seated position on a chair) and 2 and 9 (hold the head

for 5 s, raise the head from flexion while seated). Domain D3 did not show individual item misfit, DIF or response dependency.

As domains D3 and D2 had very large ceiling effects and, therefore, inadequate item-person distribution, remodeling of these domains did not yield useful results for scale improvement. Either collapsing response categories to restore disordered thresholds or removing misfitting items only reduced the score range and enlarged the ceiling effect and sample-item misfit.

Discussion

We used Rasch analysis to evaluate the MFM as a functional outcome measure in a large cohort of FSHD patients. The MFM was specifically designed to measure neuromuscular disorders. It has the advantages of being reliable, easy to perform and is suited to measure non-ambulatory patients, but our analyses revealed important limitations for its use in FSHD patients [8].

The tasks of the total MFM were relatively easy for FSHD patients. Especially domains D2 (axial and proximal motor function) and D3 (distal motor function) showed large ceiling effects, which is in line with previous studies [10,11]. Considering the items on these domains, this is not completely surprising. More than half of the items on domain D2 focus on maintaining sitting positions, keeping the head upright and contractures, and domain D3 contains many items on distal arm function. These functions are generally only limited in the most severely affected FSHD patients and contractures are rare. More items on shoulder function would be appropriate to measure FSHD patients.

Rasch analysis revealed many items with disordered thresholds, indicating difficulty for the examiner to discriminate between different response options. As a result, many of these items functioned as if they were dichotomous items: for easier items patients were able to complete the exercise either with or without compensatory movements, for difficult items patients were either able or not able to perform the exercise. This fits with the clinical observation that FSHD patients compensate for their slowly progressive weakness for a long time, then cross a certain threshold and suddenly lose function. For some of the items collapsing from 4 to 3 response categories did not restore the disordered thresholds and we did not succeed in building a model for the total MFM or the separate domains. For the total MFM, this is not surprising, as unidimensionality is one of the requirements of the Rasch model and an outcome measure composed of three different domains is in itself contradictory with the idea of finding a unidimensional measure. The remodeled domain D1 scale still showed major limitations, most importantly the sample-item mistargeting. As for domain D1, a trend towards unidimensionality was achieved; perhaps, this part of the MFM could be further improved

by adding more (difficult) items belonging to the same domain. Another approach used to remodel the MFM to be more suited to measure specific patient groups such as FSHD was confirmatory factor analysis [10]. Although this approach bears the advantage of the ability to weigh items according to their discriminant ability in any specific disorder, it does not take into account other essential clinimetric properties such as the linearity of a scale.

This study serves as an example of the importance of critically assessing the properties of an outcome measure in the light of the target population to optimize the design of clinical trials from modern metric perspectives. We show that a scale that is validated in a cohort with patients with various diagnoses is not necessarily optimally suited to measure each of the subgroups within the validation cohort separately. Although in theory the items on motor ability should be equally difficult among different neuromuscular diseases, perceived item difficulty often varies across diagnostic groups depending, for example, on their specific distribution of weakness [33,34]. Disease-specific outcome measures are often designed to optimize their person separation reliability which increases the discriminative power and the ability to measure small differences [35–37]. The latter can be of great importance in slowly progressive diseases, such as FSHD to increase responsiveness over time. Therefore, disease-specific outcome measures capturing subtle differences between diagnostic groups are, in our view, preferred over more generic outcome measures. Although this study included a large FSHD cohort comprising the whole clinical severity spectrum, it was a single-center study. For any analysis on the metric properties of an outcome measure, the results will depend on the characteristics of the included cohort. Since the scores on the MFM, FSHD clinical score and the proportion of wheelchair-bound patients in our study are similar to the results that have been described in the literature for FSHD patients, our cohort seems representative of the total FSHD population [8,26,30,38]. Another limitation of this study is that patients completed the MFM once and consequently test-retest Rasch-stability (e.g. DIF by time) of the measure was not assessed. Rasch analysis revealed multiple limitations of the MFM for FSHD and should, therefore, be used with caution in this patient population. The detailed insights into the metric abilities of the MFM are important for correct interpretation of test results, but can also be useful in developing new scales. For FSHD, there is a high need for the development interval scales on functional abilities for clinical trials.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

1 Mul K, Lassche S, Voermans NC, Padberg GW, Horlings CG, van Engelen BG. What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. *Pract Neurol* 2016; **16**:201–207.

2 Tawil R, Shaw DW, van der Maarel SM, Tapscott SJ. Clinical trial preparedness in facioscapulohumeral dystrophy: outcome measures and patient access: 8-9 April 2013, Leiden, The Netherlands. *Neuromuscul Disord* 2014; **24**:79–85.

3 Tawil R, Padberg GW, Shaw DW, van der Maarel SM, Tapscott SJ, Participants FW. Clinical trial preparedness in facioscapulohumeral muscular dystrophy: clinical, tissue, and imaging outcome measures 29-30 May 2015, Rochester, New York. *Neuromuscul Disord* 2016; **26**:181–186.

4 LoRusso S, Johnson NE, McDermott MP, Eichinger K, Butterfield RJ, Carraro E, *et al*; ReSolve Investigators and the FSHD CTRN18. Clinical trial readiness to solve barriers to drug development in FSHD (ReSolve): protocol of a large, international, multi-center prospective study. *BMC Neurol* 2019; **19**:224.

5 Eichinger K, Heatwole C, Iyadurai S, King W, Baker L, Heininger S *et al*. Facioscapulohumeral muscular dystrophy functional composite outcome measure. *Muscle Nerve* 2018. doi: 10.1002/mus.26088

6 Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989; **70**:857–860.

7 Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; **57**:1358–1362.

8 Bérard C, Payan C, Hodgkinson I, Fermanian J; MFM Collaborative Study Group. A motor function measure for neuromuscular diseases. Construction and validation study. *Neuromuscul Disord* 2005; **15**:463–470.

9 Vuillerot C, Payan C, Girardot F, Fermanian J, Iwaz J, Berard C *et al*. Responsiveness of the motor function measure in neuromuscular diseases. *Arch Phys Med Rehabil* 2012; **93**:2251–2256 e2251.

10 Guillot T, Roche S, Rippert P, Hamroun D, Iwaz J, Ecochard R, Vuillerot C; MFM Study Group. Is going beyond rasch analysis necessary to assess the construct validity of a motor function scale? *Arch Phys Med Rehabil* 2018; **99**:1776–1782.e9.

11 de Lattre C, Rippert P, Hmaroun D, Sacconi S, Poirot I, Vuillerot C. The motor function measure (MFM) in the facio scapulo humeral dystrophy (FSHD) population: description and responsiveness. *Ann Phys Rehabil Med* 2016; **59S**:e84–e85.

12 Merbitz C, Morris J, Grip JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; **70**:308–312.

13 DeVellis RF. Classical test theory. *Med Care* 2006; **44** (11 Suppl 3):S50–S59.

14 Svensson E. Guidelines to statistical evaluation of data from rating scales and questionnaires. *J Rehabil Med* 2001; **33**:47–48.

15 Rasch G. Probabilistic models for some intelligence and attainment tests. *Danish Institute for Educational Research, Copenhagen* 1960; **xii**:184.

16 Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007; **46**:1–18.

17 Stucki G, Daltroy L, Katz JN, Johannesson M, Liang MH. Interpretation of change scores in ordinal clinical scales and health status measures: the whole may not equal the sum of the parts. *J Clin Epidemiol* 1996; **49**:711–717.

18 van der Beek NA, Hagemans ML, van der Ploeg AT, van Doorn PA, Merkies IS. The Rasch-built Pompe-specific activity (R-PAct) scale. *Neuromuscul Disord* 2013; **23**:256–264.

19 Hermans MC, Faber CG, De Baets MH, de Die-Smulders CE, Merkies IS. Rasch-built myotonic dystrophy type 1 activity and participation scale (DM1-Activ). *Neuromuscul Disord* 2010; **20**:310–318.

20 van Nes SI, Vanhoutte EK, van Doorn PA, Hermans M, Bakkers M, Kuitwaard K, *et al*. Rasch-built Overall Disability Scale (R-ODS) for immune-mediated peripheral neuropathies. *Neurology* 2011; **76**:337–345.

21 Vanhoutte EK, Faber CG, van Nes SI, Jacobs BC, van Doorn PA, van Koningsveld R, *et al*.; PeriNomS Study Group. Modifying the medical research council grading system through rasch analyses. *Brain* 2012; **135**:1639–1649.

22 Mayhew A, Cano S, Scott E, Eagle M, Bushby K, Muntoni F.; North Star Clinical Network for Paediatric Neuromuscular Disease. Moving towards meaningful measurement: rasch analysis of the North star ambulatory assessment in duchenne muscular dystrophy. *Dev Med Child Neurol* 2011; **53**:535–542.

23 Vuillerot C, Rippert P, Kinet V, Renders A, Jain M, Waite M, *et al*.; CDM MFM Study Group. Rasch analysis of the motor function measure in patients with congenital muscle dystrophy and congenital myopathy. *Arch Phys Med Rehabil* 2014; **95**:2086–2095.

24 Mul K, Vincenten SCC, Voermans NC, Lemmers RJLF, van der Vliet PJ, van der Maarel SM, *et al*. Adding quantitative muscle MRI to the FSHD clinical trial toolbox. *Neurology* 2017; **89**:2057–2065.

25 Mul K, Voermans NC, Lemmers RJLF, Jonker MA, van der Vliet PJ, Padberg GW, *et al*. Phenotype-genotype relations in facioscapulohumeral muscular dystrophy type 1. *Clin Genet* 2018; **94**:521–527.

26 Lamperti C, Fabbri G, Vercelli L, D'Amico R, Frusciante R, Bonifazi E, *et al*. A standardized clinical evaluation of patients affected by facioscapulohumeral muscular dystrophy: the FSHD clinical score. *Muscle Nerve* 2010; **42**:213–217.

27 Andrich D, Sheridan B, Luo G. Rasch models for measurement: *RUMM2030*. RUMM Laboratory, Perth, 2010.

28 Vanhoutte EK, Hermans MC, Faber CG, Gorson KC, Merkies IS, Thonnard JL.; PeriNomS Study Group. Rasch-ionale for neurologists. *J Peripher Nerv Syst* 2015; **20**:260–268.

29 Wright BD, Tennant A. Sample size again. *Rasch Meas Transactions* 1996; **9**:468.

30 Benaïm C, Sacconi S, Fournier-Mehouas M, Tanant V, Desnuelle C. Validity of the motor function measurement scale when routinely used in the follow-up of adult outpatients in a neuromuscular center. *Rev Neurol (Paris)* 2010; **166**:49–53.

31 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; **310**:170.

32 Andrich D, Luo G. Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. *J Appl Meas* 2003; **4**:205–221.

33 Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, *et al*. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004; **42**:I37–I48.

34 Arnould C, Vandervelde L, Batcho CS, Penta M, Thonnard JL. Can manual ability be measured with a generic ABILHAND scale? A cross-sectional study conducted on six diagnostic groups. *BMJ Open* 2012; **2**:e001807.

35 Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol* 2003; **56**:52–60.

36 Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, *et al*. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med* 2005; **60**:1571–1582.

37 Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989; **27**:S217–S232.

38 Statland JM, Tawil R. Risk of functional impairment in Facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2014; **49**:520–527.