

Original article / Article original

Development and validation of a motor function classification in patients with neuromuscular disease: The NM-Score

Création et validation d'une classification en grade de sévérité dans les maladies neuromusculaires : la classification NM-Score

C. Vuillerot^{a,*,b,c,d}, P. Rippert^{a,e}, S. Roche^{b,f}, C. Bérard^a, F. Margirier^a, C. de Lattre^a, I. Poirot^a, A. Berruyer^a, V. Tiffreau^g, M. Fournier-Mehouas^h, F. Bouhourⁱ, J.-A. Urtizberea^j, A. Renders^k, R. Ecochard^{b,f}, Le groupe d'étude NM-Score, A. Le Flem^l, A. Barrière^l, A.P. Rouyer^l, S. Fontaine^l, J.-P. Vadot^l, E. Luc Pupat^l, Y. Chartier^l, D. Vincent-Genod^l, F. Girardot^l, V. Manel^l, F. Aubert^l, G. Rode^m, D. Denisⁿ, V. Germa^o, S. Quijano^p, N. Pelligrini^p, M.C. d'Anjou^q, L. Féasson^q, S. Chabrier^q, A. Furby^q, C. Goyet^q, M.C. Delmas^r, M. Campech^r, F. Robert^r, H. Hovart^s, J.-M. Cuisset^s, I. Badoil^t, C. Fafin^u, V. Tanant^u, S. Sacconi^u, J.-P. Gayraud^v, A. Carpentier^w, S. Vanderschueren^w, I. Bourdeauducq^w, D. Salicio-Castillo^x, A.M. Cobo^x, M.C. Commare^y, V. Farigoule^y, C. Huzar^y, B. Berger^z, V. Humbertclaude^{aa}, F. Rumeau^{bb}, E. Viehweger^{cc}, C. Payet-Laury^{cc}, I. Penisson-Besnier^{dd}, V. Kinet^{ee}, D. Laridant^{ee}, V. Spehrs-Ciaffi^{ff}, G. Bassez^{gg}, N. Goemans^{hh}, D. Pichancourtⁱⁱ, L. Jezequelⁱⁱ, N. Vedrenneⁱⁱ

^a L'Escale, service de MPR pédiatrique, hôpital Femme-Mère-Enfant, hospices civils de Lyon, 59, boulevard Pinel, 69677 Bron cedex, France

^b Université de Lyon, 69000 Lyon, France

^c Université Lyon-I, 69100 Villeurbanne, France

^d CNRS, UMR 5558, laboratoire de biométrie et biologie évolutive, équipe biostatistique santé, 69310 Pierre-Bénite, France

^e Pôle information médicale évaluation recherche, hospices civils de Lyon, 69003 Lyon, France

^f Service de biostatistique, hospices civils de Lyon, 162, avenue Lacassagne, 69424 Lyon cedex 03, France

^g EA 4488 « activités physiques, muscles, santé », centre de référence des maladies neuromusculaires, hôpital Swynghedauw, université de Lille-2, CHRU de Lille, 59037 Lille cedex, France

^h Service MPR, pôle neurosciences, clinique, CHU de Nice, 151, route de Saint-Antoine-de-Ginestiere, BP 3079, 06202 Nice cedex 03, France

ⁱ Service d'électromyographie et pathologies neuromusculaires, hôpital neurologique, groupement hospitalier Lyon-Est, hospices civils de Lyon, 59, boulevard Pinel, 69677 Bron cedex, France

^j Centre de référence neuromusculaire, GNMH, hôpital Marin, AP-HP, 64700 Hendaye, France

^k Centre de référence des maladies neuromusculaires, cliniques universitaires Saint-Luc, université catholique de Louvain, 1200 Bruxelles, Belgium
^l HFME, 69677 Lyon, France

^m Hôpital H.-Gabriel, 69677 Lyon, France

ⁿ Fondation Richard, 69677 Lyon, France

^o Hôpital Necker, 75015 Paris, France

^p Hôpital R.-Poincaré, 92380 Garches, France

^q Hôpital Bellevue, 42100 Saint-Étienne, France

^r CHU de La Réunion, 97405 Saint-Denis, Reunion

^s Hôpital Swynghedauw, 59037 Lille, France

^t SAMSAH, 69355 Lyon, France

^u CHU de l'Archet, 06200 Nice, France

^v Centre P.-Dottin, 31522 Toulouse, France

^w Centre M.-Sautelet, 59650 Villeneuve-d'Asq, France

^x Hôpital Marin, 64700 Hendaye, France

^y Hôpital Mère-Enfant, 38700 Grenoble, France

* Corresponding author.

E-mail address: carole.vuillerot@gmail.com, carole.vuillerot@chu-lyon.fr (C. Vuillerot).

^z Hôpital Michallon, 38700 Grenoble, France^{aa} CHU de Montpellier, 34295 Montpellier, France^{bb} CHU Nancy, 54000 Nancy, France^{cc} Hôpital de la Timone, 13005 Marseille, France^{dd} CHU Angers, 49100 Angers, France^{ee} UCL Saint-Luc, Bruxelles, Belgium^{ff} CHUV, Lausanne, Switzerland^{gg} CHU H.-Mondor, 94010 Créteil, France^{hh} University Hospitals Leuven, Louvain, Belgiumⁱⁱ CHR Pierre-Le-Damany, 22300 Lannion, France

Received 28 February 2013; accepted 3 October 2013

Abstract

Objective. – To develop a classification for neuromuscular disease patients in each of the three motor function domains (D1: standing and transfers; D2: axial and proximal function; D3: distal function).

Materials and methods. – A draft classification was developed by a study group and then improved by qualitative validation studies (according to the Delphi method) and quantitative validation studies (content validity, criterion validity and inter-rater reliability). A total of 448 patients with genetic neuromuscular diseases participated in the studies.

Results. – On average, it took 6.3 minutes to rate a patient. The inter-rater agreement was good when the classification was based on patient observation or an interview with the patient (Cohen's kappa = 0.770, 0.690 and 0.642 for NM-Score D1, D2 and D3 domains, respectively). Stronger correlations (according to Spearman's coefficient) with the respective “gold standard” classifications were found for NM-Score D1 (0.86 vs. the Vignos Scale and –0.88 vs. the Motor Function Measure [MFM]-D1) and NM-Score D2 (–0.7 vs. the Brooke Scale and 0.64 vs. MFM D2) than for NM-Score D3 (0.49 vs. the Brooke scale and –0.49 vs. MFM D3).

Discussion/conclusions. – The NM-Score is a reliable, reproducible outcome measure with value in clinical practice and in clinical research for the description of patients and the constitution of uniform patient groups (in terms of motor function).

© 2013 Published by Elsevier Masson SAS.

Keywords: Neuromuscular disease; Outcome measure; Motor function; Activities of daily living

Résumé

Objectifs. – L'objectif de cette étude est le développement d'une classification rapide et reproductible en grades de sévérité pour les patients porteurs d'une maladie neuromusculaire dans chacun des 3 domaines de fonction motrice (D1 : position debout et transferts ; D2 : motricité axiale et proximale ; et D3 : motricité distale).

Patient et méthodes. – Plusieurs phases de validation qualitative (méthode Delphi) et quantitative (validité de contenu, concurrente et inter-observateur) ont été nécessaires après la proposition d'une première classification par un groupe d'étude. Respectivement pour chacune des phases, 161 et 229 patients porteurs de maladies neuromusculaires d'origine génétique ont été inclus dans ces 2 phases.

Résultats. – La durée moyenne de passation de la classification est de 6,3 minutes (DS : 3,7). La fiabilité inter-observateur est satisfaisante lorsqu'elle est utilisée à partir de l'interrogatoire ou de l'observation du patient en situation écologique (kappa = 0,770, 0,690 et 0,642 pour NM-Score D1, D2 et D3 respectivement). De meilleures corrélations avec les échelles de références (mesurées par le coefficient de Spearman) sont retrouvées pour le NM-Score D1 (vs Vignos = 0,86, vs MFM D1 = –0,88) et D2 (vs Brooke = –0,7, vs MFM D2 = 0,64) que pour le NM-Score D3 (vs Brooke = 0,49, vs MFM D3 = –0,49).

Discussion/conclusions. – La classification NM-Score est une classification fiable et reproductible, pouvant être utilisée dans la pratique clinique courante pour le suivi des patients ainsi qu'en recherche clinique pour la description des populations et la constitution de groupe d'atteinte fonctionnelle identique.

© 2013 Publié par Elsevier Masson SAS.

Mots clés : Maladies neuromusculaires ; Évaluation ; Fonction motrice ; Activités de la vie quotidienne

1. English version

1.1. Introduction

Over the last few decades, progress in research and clinical care has increased the life expectancy of patients with neuromuscular diseases (NMDs). Assessment of NMDs has become commonplace during clinical follow-up and as an

outcome measure in much-awaited clinical trials. A rigorous, appropriate metrological approach is essential when developing an assessment tool. It is no longer possible to make do with approximate measures or to use tools that are not appropriate for progressive diseases.

Loss of muscle strength is a symptom of all NMDs. Depending on the disease in question, this loss of strength affects different muscles to different extents. In clinical trials,

the therapeutic target is primarily the muscle fibre; it is important to consider not only changes in bodily functions such as strength (i.e. mechanical aspects) but also the latter's functional relevance (i.e. the impact on mobility and activities of daily living) [12,14]. Motor function assessment (i.e. measurement of the impact of the NMD on functional limitations) is well accepted by patient and therapists alike [5,10–12,14].

For the most commonly used motor function scales, assessment of a patient takes about 30 minutes and presupposes that the therapist has been specifically trained in this procedure. However, precise evaluation of motor function is not necessary in routine clinical practice (or in some research projects). For example, estimation of levels or grades of functional severity may be enough for the constitution of patient groups that are homogeneous in terms of motor function. In the field of NMD, a number of functional scores have been published. However, these scores either do not take account of the various domains of motor function (such as standing and transfers, axial/proximal motor function and the distal motor function) or are not sufficiently specific for the population being studied.

For example, the Brooke Upper Extremity Scale [3,4] and the Vignos Lower Extremity Scale [16] are specifically used to grade patients suffering from muscular dystrophy but focus solely on arm function and gait, respectively. Moreover, literature data on the validation of these two tools are scarce.

On the basis of the research developed by Palisano et al. [13] in the field of cerebral palsy, we have developed a classification (called the “NM-Score”) to rate the severity of motor function impairment in each of the three domains of motor function (D1: standing and transfers; D2: axial and proximal motor function; D3: distal motor function), as previously described for the development and validation of the Motor Function Measure (MFM) by Bérard et al. [1]. We sought to create a simple, quick-to-use tool whose primary goal is to classify patients by providing a general representation of their functional status in their usual environment via the written description of each severity grade in each domain of motor function.

Development of the NM-Score comprised the following steps:

- generation of an initial classification by a working group;
- content validation (according to the Delphi process);
- validation of the inter-rater agreement;
- criterion validation.

The classification was modified at each step, in order to obtain the final version presented here.

1.2. Materials and methods

Fig. 1 presents the different steps in the development of the NM-Score and shows the timeline, the methodology and the numbers of patients and professionals involved in each step. The study protocol was submitted to the local independent ethics committee (CCP Lyon Sud Est 2, Lyon, France) but was considered not to be subject to the French legislation on

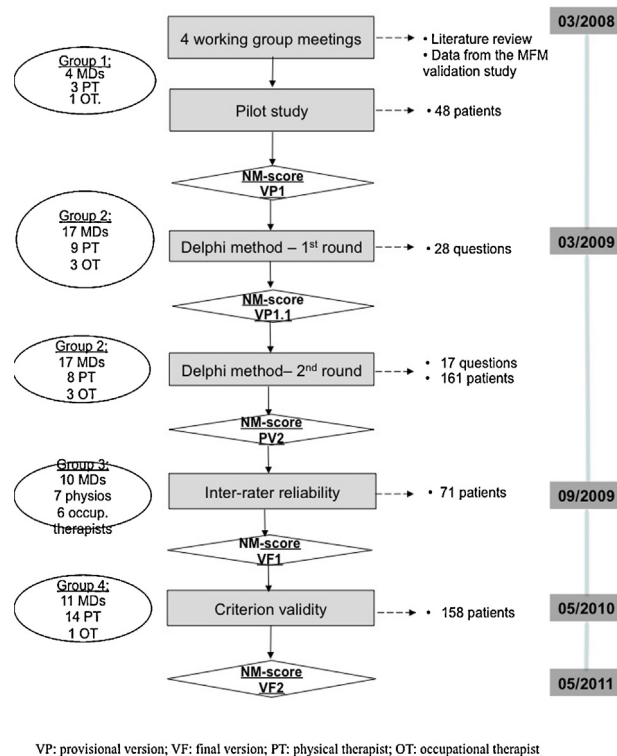


Fig. 1. A graphic representation of the different steps in the study.

biomedical research. Nevertheless, the investigator handed an information sheet to the patient prior to inclusion in the study.

1.2.1. Generation of an initial version of the NM-Score

We performed an exhaustive review of the literature on motor function assessment tools in the field of NMD. Eight experts in the treatment of NMDs (three physiotherapists, an occupational therapist and four physicians) developed the first version of the NM-Score (referred to as “NM-Score VP1”) in four meetings between March 2008 and February 2009. In their normal practice, four of the experts worked in the paediatric area and the other four worked with adults. The experts' mean time since qualification was 13.7 years (range: 2–25).

The different severity grades in the NM-Score VP1 classification were defined for each of the three motor function domains (D1: standing position and transfers; D2: axial and proximal motor function; D3: distal motor function) on the basis of the experts' clinical experience, the literature review and an analysis of data from the MFM's validation study of 303 NMD patients [1]. A pilot study was performed on 48 NMD patients, in order to help define each level of the rating scale.

For each of the three motor function domains, the experts suggested a five-grade classification (Grade 0: no impairment; Grade 1: slight impairment; Grade 2: moderate impairment; Grade 3: severe impairment; Grade 4: very severe impairment). Several so-called major criteria were used primarily to distinguish between two grades: the patient's usual ability to walk, run and jump (for D1), his/her usual ability to sit upright, perform activities involving proximal motor actions and hold his/her head up (for D2) and his/her usual ability to handle

objects (for D3). Several minor criteria (with less discriminant power) were used to improve the description of each major criterion. These minor criteria (identified during the patient study) concerned the need for technical aids and/or human assistance. A minor criterion was selected for use in improving the description of a grade when it was met by at least 80% of the patients in the corresponding grade.

1.2.2. Content validation

In order to analyze the NM-Score VP1's content validity, a Delphi-type consensus process [2,7,9,15] was applied. This step enabled us to gather opinions and comments on the relevance of the classification's content (and, in particular, the selected major and minor criteria) from several French-speaking experts in the field of NMD.

The Delphi process is an iterative, interactive procedure with two to four rounds. The selected experts were representative of their profession, had power to implement the findings and were not likely to be challenged as experts in the field (as suggested by Fink et al. [7]). After each round, the expert group is informed of the results, which are then used to build the questionnaire for the next round. The process ends when the group reaches a predefined level of agreement or when the exchanged information has enabled the process's objectives to be met.

Thirty-six French-speaking physiotherapists, occupational therapists and physicians (based in France, Belgium and Switzerland) with acknowledged expertise in NMD and who had not participated in the first phase of the study were invited to participate, and 29 accepted (9 physiotherapists, 3 occupational therapists and 17 physicians). Ten participants worked exclusively with adults, 14 worked exclusively with children and five worked with both adults and children. The experts' mean time since qualification was 11.8 years (range: 3–28).

In the first round in the Delphi process, a questionnaire and the NM-Score VP1 classification were e-mailed to the participating experts. The questionnaire was divided into four sections and included a total of 28 items. Each item also featured an open-ended question, so that the experts were encouraged to explain their answers. The first part (part A) was composed of four questions on the need for a new classification system. Parts B, C and D were respectively dedicated to the D1, D2, and D3 domains and included 24 questions on the definition and description of each grade and the inter-grade differences. Each item was scored on a nine-point Likert scale (with 1 indicating “strongly disagree”, 5 indicating “neither agree nor disagree” and 9 indicating “strongly agree”). Consensus for a given item was reached when there was either a high degree of agreement or no disagreement. Agreement was defined as a median agreement score of > 7 , with at least 80% of the experts giving a score of > 7 . Disagreement was defined as more than 30% of the individual scores falling in the ranges 1 to 3 and 6 to 9.

In the second round, the questionnaire included 17 items; the 11 items from the first questionnaire that had not obtained a consensus in the first round and 6 items concerning the

classification's ease of use and administration conditions. The experts were invited to use the NM-Score VP1 classification in their daily practice before completing the questionnaire.

1.2.3. Inter-rater reliability

This validation phase assessed the degree of agreement between scores reported by different therapists, i.e. the inter-rater reliability. Based on the explanations given by examiners in cases of disagreement, this phase was also intended to help us improve the classification and develop a new version (referred to as “NM-Score VF1”).

For each patient who met the inclusion criteria, two examiners (physicians, physiotherapists or occupational therapists) had to independently rate the patient according to the NM-Score on the basis of:

- an interview with the patient or their family;
- observation in an ecological context and/or;
- the patient's medical records.

A given pair of examiners was only allowed to grade five patients at most. If the examiners disagreed over the rating, they had to explain the reasons for their disagreement.

Six clinics and reference centres for NMDs (located in France and French-speaking Belgium) participated in the inter-rater reliability study. The study inclusion criteria were as follows: a patient with confirmed NMD, aged between 6 and 60, known to the two examiners for less than three months and having consulted in the three months preceding the rating visit. The 23 therapists having agreed to participate in the validation study (seven physiotherapists, six occupational therapists and 10 physicians) had to familiarize themselves with the NM-Score before rating any patients.

1.2.4. Criterion (concurrent) validity

The objective of this part of the study was to achieve criterion validation for the NM-Score against three so-called “fuzzy” reference scales (given the imperfect nature of the “gold standards” in this field): the MFM, the Brooke scale and the Vignos scale. In order to standardize these evaluations, all the therapists participating in this part of the study had already been trained in administration of the MFM. Furthermore, the criteria for the Brooke and Vignos classifications were explained on the case report form.

Fourteen clinics and NMD reference centres participated in the criterion validity study. The included patients had to be aged between 6 and 60 and have a confirmed NMD. All the therapists participating in the criterion validation study had to familiarize themselves with the NM-Score before starting the assessment (by carefully reading the rating instructions and looking at the classification as a whole).

The therapists rated the patients according to the NM-Score on the day of the MFM assessment (but before the latter was administered).

1.2.5. Statistical analyses

Quantitative variables were described as the mean, standard deviation (SD), range and quartiles. Qualitative variables were described as the number and percentage.

The inter-rater agreement was analyzed for each field (D1, D2 and D3) and each rating condition (medical records, an interview and/or an observation in an ecological context) by using Cohen's kappa coefficient (κ), which indicates whether the agreement is better than that expected by chance. Depending on the value of κ , the degree of agreement can be interpreted as fair (< 0.4), moderate (0.4 to 0.59), substantial (0.6 to 0.8) and almost perfect (> 0.8) [6].

In order to analyze the NM-Score's criterion validity, we chose "fuzzy" gold standard scales for each domain (D1, D2 and D3). We assessed the degree of agreement between the NM-Score (in five ranked classes and for each of the D1, D2 and D3 domains) and the value of the MFM (a quantitative score ranging from 0 to 100). The degree of agreement between the NM-Score severity grade and the Vignos and Brooke scores was also studied. We calculated various measures of agreement, such as Spearman's correlation coefficient.

1.3. Results

1.3.1. Development of the initial classification and a qualitative, Delphi-like analysis of the content validity

In the first round of the Delphi process, a consensus was reached for 17 of the 28 items in NM-Score VP1. All the experts agreed on the necessity of a new classification system for patients suffering from NMD based on their performance of activities of daily living. Twenty-two of the 28 experts thought that the classification had potential applications in both routine clinical practice and clinical research. All the experts considered that one of the major strengths of the NM-Score was its reliance on a description of the patient's performance in all motor function domains (rather than walking ability alone).

After the first round of the Delphi process, the classification was revised using the experts' comments and was sent back with a questionnaire for a second round. A consensus was reached for 16 of the 17 items. The only non-consensual point concerned the age from which the classification could be administered; 13 of the 28 experts thought that the classification could not be correctly applied to children as young as four. In the second round of the Delphi process, the VP2 classification was used by the experts to rate a total of 161 patients. The

largest number of patients rated by any one expert was 24 and all the experts agreed that the classification was easy to use.

1.3.2. Quantitative analyses

1.3.2.1. The inter-rater reliability study. Seventy-one patients (47 males and 24 females, aged between 6 and 50) were included in this study of inter-rater reliability. The mean $\pm$ SD age was 18.7 ± 13 . Sixty-six percent were aged under 18 on the day of the evaluation.

For 36 of the 71 patients, the two evaluations were performed on the same day. In 76% of cases, the time interval between the two evaluations was less than two weeks (the longest interval being 3 months). In terms of the rating conditions, 32 evaluations were based solely on observation of the patient, 20 were based solely on medical records and 9 were based solely on an interview with the patient and/or the patient's family. The other evaluations involved two or more of these situations. The κ coefficients [95% confidence interval (CI)] for D1, D2 and D3 were respectively 0.72 [0.57 to 0.83], 0.62 [0.47 to 0.75] and 0.56 [0.39 to 0.71] (Table 1). For D1, D2 and D3, the percentages of inter-rater agreement were respectively 82.9%, 78.1% and 75.6% and the κ coefficients [95% CI] were respectively 0.77 [0.59 to 0.90], 0.69 [0.48 to 0.85] and 0.64 [0.43 to 0.82]. These values were higher when evaluations were based on patient observation and/or a patient interview.

Nineteen cases of disagreement were analyzed. In 15 cases, at least one therapist had not correctly read the introduction or the instructions for use. In particular, the therapist at fault did not apply the major criteria before referring to the minor criteria when rating the patient.

1.3.2.2. The criterion validity study. A total of 158 patients were included in this validation phase. The great majority of these patient were able to walk ($n = 124$). There were more males ($n = 102$) than females in the study population; this was due to the presence of a large number of patients suffering from the male-only diseases Duchenne muscular dystrophy ($n = 27$) and Becker muscular dystrophy ($n = 10$). The other diagnoses were myopathy and congenital muscular dystrophy ($n = 26$), infantile spinal muscular atrophy ($n = 22$), peripheral neuropathies ($n = 21$), myotonic dystrophy type 1 ($n = 17$), facioscapulohumeral muscular dystrophy ($n = 12$), girdle myopathies ($n = 8$) and other NMDs (15). The mean $\pm$ SD time needed to administer the NM-Score was 6.3 ± 3.7 minutes per patient

Table 1

Percentage agreement between raters and inter-rater reliability (as judged by Cohen's kappa [κ] for each motor function domain [D1, D2 and D3]).

Domain	Overall study population ($n = 71$)		Patients graded on the basis of an observation and/or an interview ($n = 41$)	
	% agreement	κ (95% CI) ^a	% agreement	κ (95% CI) ^a
D1	78.87	0.72 (0.57–0.83)	82.93	0.77 (0.59–0.90)
D2	71.83	0.62 (0.47–0.75)	78.05	0.69 (0.48–0.85)
D3	70.42	0.56 (0.39–0.71)	75.61	0.64 (0.43–0.82)

^a The confidence interval (CI) was calculated using the Bootstrap method.

Table 2

Distribution of the Vignos scores by NM-Score D1 grade and distribution of the Brooke scores by NM-Score D2 and D3 grades.

NM-Score D1	Number of patients	Vignos score					
		Mean (SD)	Minimum	p25	p50	p75	Maximum
0	8	1 (0)	1	1	1	1	1
1	57	1.6 (0.7)	1	1	1	2	3
2	38	2.76 (1.3)	1	2	2	3	6
3	16	5.75 (2.24)	2	3.5	6.5	7	9
4	29	8.93 (0.26)	8	9	9	9	9
Total	158	3.75 (3.05)	1	1	2	6.5	9

NM-Score D2	Number of patients	Brooke score					
		Mean (SD)	Minimum	p25	p50	p75	Maximum
0	44	1.2 (0.51)	1	1	1	1	3
1	65	1.6 (0.8)	1	1	1	2	5
2	17	2.65 (1.0)	1	2	3	3	4
3	11	3.9 (1.0)	2	3	4	5	5
4	6	5.2 (0.4)	5	5	5	5	6
Total	143	1.9 (1.3)	1	1	1	3	6

NM-Score D3	Number of patients	Brooke score					
		Mean (SD)	Minimum	p25	p50	p75	Maximum
0	42	1.4 (0.7)	1	1	1	2	3
1	52	1.4 (0.7)	1	1	1	2	4
2	50	2.6 (1.3)	1	1	2	3.5	5
3	7	4.1 (1.2)	2	3	5	5	5
4	2	5.5 (0.7)	5	5	5.5	6	6
Total	143	1.9 (1.3)	1	1	1	3	6

p25: first quartile; p50: second quartile; p75: third quartile.

($n = 158$). In order to study the classification's criterion validity, 148, 143 and 158 patients were scored against the Vignos, Brooke and MFM scales, respectively.

The distribution of the Vignos scores for each NM-Score D1 grade and the distribution of the Brooke scores for each NM-Score D2 and D3 grade are shown in Table 2. The distribution between the various grades was relatively balanced for D1 and D2 but was less balanced for D3 (for which there were relatively few grade 4 patients). Since the dispersion of the Vignos score was low for grades 0 and 4 in NM-Score D1, a floor effect and a ceiling effect can be suspected for at least one of the two scales. However, the small sample size (8 patients) for grade 0 in NM-Score D1 makes it difficult to confirm the presence of a floor effect. The means Vignos score increased steadily from NM-Score D1 grade 0 to grade 4. Only grade 3 was associated with high dispersion of the Vignos score (mean: 5.75; range: 2–9), suggesting that the level of impairment in grade 3 patients was heterogeneous. For NM-Score D2 and D3, the dispersion of the Brooke score was high and was essentially the same when comparing grades 0 and 1 in each domain (NM-Score D2 grade 0 [1–3] and grade 1 [1–5]; NM-Score D3 grade 0 [1–3] and grade 1 [1–4]). For the other grades, the mean values and distributions of the Brooke score increased steadily from grade 2 to grade 4 in the NM-Score D2 and D3 domains.

The distributions of the D1, D2 and D3 MFM scores by NM-Score D1, D2 and D3 grade are presented as box plots in Fig. 2. For the NM D2 score and (above all) the D3 score (but not the NM-Score D1 score), there were no significant differences

between grades 0 and 1 in terms of the mean Brooke score or mean MFM score. The dispersion of the MFM D3 score for the NM-Score D3 grades 3 and 4 was high, although the sample sizes were small (7 and 3 patients, respectively); this must be taken into account when interpreting the box plots. Moreover, for grade 4, the first quartile and the median were equivalent; this is not visible on the box plots.

The criterion validity for each NM-Score domain with respect to the reference scales was estimated by calculating Spearman's coefficient and its 95% CI (Table 3). The best correlations were observed for NM-Score D1 and NM-Score D2, which were more strongly correlated with the MFM score than with the functional scores.

1.4. Discussion

The NM-Score is a valid, reliable tool for classifying patients into five categories as a function of the degree of functional impairment in the three domains of motor function (D1: standing and transfers; D2: axial and proximal function; D3: distal function: D1).

This new classification has been welcomed by the many practitioners and researchers who have already used it. The French version of the NM-Score can be downloaded free of charge from the MFM website (<http://www.mfm-nmd.org/accueil.aspx>).

The NM-Score is a simple, quick-to-administer tool for describing a patient's motor functions. Detailed training is not

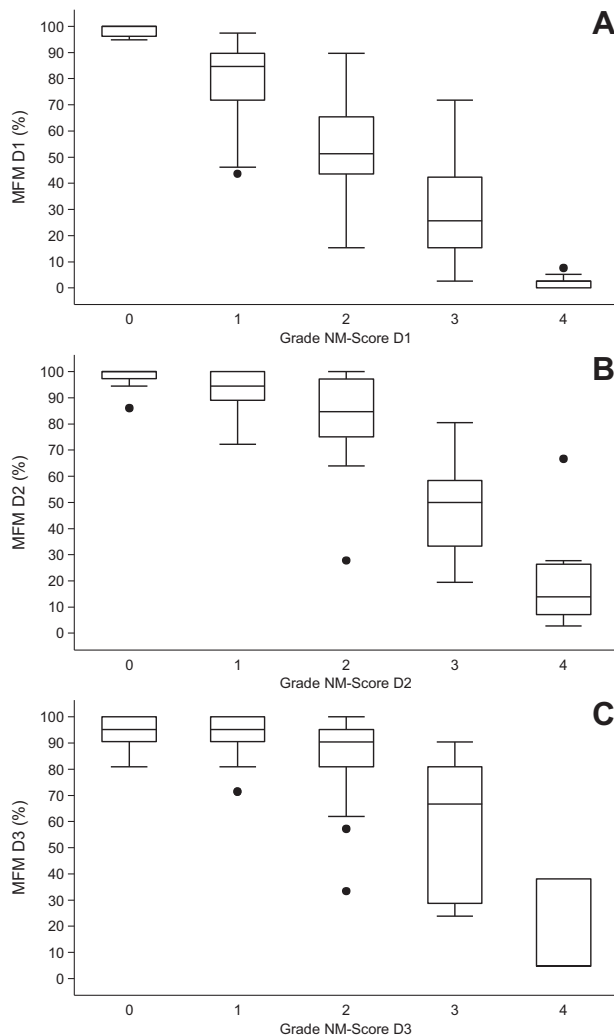


Fig. 2. Box plots of the distributions of the MFM D1 (A), D2 (B) and D3 scores (C) according to the corresponding NM-Score D1, D2 and D3 grades.

Table 3

Spearman's coefficients for the correlation between each domain of the NM-Score and the reference scales (criterion validity).

Reference scale	Spearman's coefficient	95% confidence interval	
Vignos vs. NM-Score D1	0.86	0.81	0.89
MFM D1 vs. NM-Score D1	−0.88	−0.91	−0.83
Brooke vs. NM-Score D2	0.64	0.53	0.73
MFM D2 vs. NM-Score D2	−0.70	−0.77	−0.60
Brooke vs. NM-Score D3	0.49	0.36	0.61
MFM D3 vs. NM-Score D3	−0.49	−0.60	−0.36

required for use of the NM-Score; users are simply told to read the introduction and instructions carefully and to get to know the tool by doing a few trial ratings. We consider that this classification facilitates discussion between the physician and the patient by focusing on activities of daily living and by identifying factors that aggravate or facilitate motor function in an ecological context. This classification also enables healthcare professionals to use a common vocabulary when

they talk about their patients. It notably takes account of the patient's opinions and those of his/her family, since the interview probes these aspects.

The NM-Score describes the patient in terms of a severity grade that is not restricted to his/her ability to walk. In fact, the NM-Score is based on a clinical description of motor function, which yields a comprehensive view of the patient's functions in all domains. This classification can be likened to the multicriterion TNM tumour staging classification. Depending on the patient's disease and/or age, a specific pattern may be observed. For example, the NM-Score's D1 and D2 domains are most affected (regardless of the patient's age) by infantile spinal muscular atrophy or facioscapulohumeral muscular dystrophy, where the impairment is primarily proximal in nature. In Duchenne muscular dystrophy, changes in NM-Score D1 during early-stage disease are followed by changes in NM-Score D2 and then in NM-Score D3.

During the Delphi process, the experts all agreed that there is a need for a classification system based on the patient's functional limitations in activities of daily living (rather than on their abilities in clinical tests). Most of the experts considered that our classification was useful for clinical research (in the constitution of groups of patients with a similar level of motor function) and clinical follow-up. Two rounds of the Delphi process were needed to obtain a consensus on all the descriptions suggested for each grade. About 30% of the experts thought that the NM-Score could not be used with sufficient accuracy in children under the age of six. In order to avoid difficulties related to the variability of normal psychomotor development in the child, we suggest that the NM-Score can be applied from the age of 6 onwards.

The NM-Score's inter-rater reliability was good when the tool was administered as part of an interview and/or during observation of the patient in an ecological context. This condition is required for sufficient reliability (Cohen's $\kappa > 0.60$) in all three domains (with κ values of 0.77, 0.69 and 0.64 for D1, D2, and D3, respectively) and enables the use of the NM-Score in multicentre therapeutic trials. By way of a comparison, Palisano et al. [13] studied the Gross Motor Function Classification System and reported a κ of 0.75 in children aged 2 and over, whereas Finkel et al. [8] reported a weighted κ of 0.61 [95% CI: 0.59 to 0.63] for the Test of Infant Motor Performance. Comparisons with other tools are complicated by the fact that ICCs were used to estimate inter-rater reliability. For instance, the ICC for the Brooke score is between 0.96 and 0.99 [3,4]. By definition, Cohen's κ is well suited to ordinal variables, such as those in the Brooke, Vignos and NM-Score classifications. The ICC is considered as the most suitable for quantitative variables and is also more dependent on the sample's variability than Cohen's κ is; the greater the variability, the closer the ICC is to 1. Hence, with the same parameters, Cohen's κ will give a lower value than the ICC. Our analysis of inter-rater disagreements during the reliability study showed that in most cases, at least one of the two raters had not complied with initial use of the major criteria when rating the patient. Thus, in the final version of the classification, we moved the minor criteria to a separate page so

that they were only used when the rater hesitated between two grades (established using the major criteria).

In terms of criterion validity, NM-Score D1 was very strongly correlated (Spearman's coefficient: 0.86) with the Vignos score. The low dispersion of the Vignos scores for grade 4 prompted us to suspect a small floor effect.

The Brooke classification measures proximal motor function (grades 1, 2 and 3) and/or distal motor function (grade 3, 4, 5 and 6) of the upper extremities. In fact, the Brooke classification was more strongly correlated with NM-Score D2 (Spearman's coefficient: 0.64) than NM-Score D3 (Spearman coefficient: 0.49) but both values were lower than for D1. Given that the statistical significance and the breadth of the coefficient's 95% CI are strongly linked to the sample size, interpretation of these finding is complicated by the fact that there were few patients with a severe impairment in D2 or D3.

The correlation between the MFM and the NM-Score was substantial for D1 and D2 and moderate for D3. Generally, in NMD, the D3 motor function domain is the last to be impaired (except in purely distal NMDs); this is borne out by the patients' distribution by domain and by grade in the NM-Score and probably limits our ability to interpret the results. In terms of the results, the NM-score's ability to discriminate between grades 0 and 1 in the D2 and D3 domains appears to be too low, even though we wanted to ensure that grade 0 (corresponding to normal abilities) could be differentiated from even slight impairment (grade 1), such as that due to fatigability in the sitting position during the day or a lack of precision in a certain movement mentioned in the patient interview.

Translation of the NM-Score into English would enable the international dissemination of this new tool.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

The authors wish to thank the patients and family members who (as always) participated in the project with great enthusiasm. We also thank all the physicians and therapists in the NM-Score working group for enabling the successful conclusion of this project. We thank the Lyon Public Hospitals Group (Hospices Civils de Lyon) for financial and logistic support as part of the Projet Hospitalier de Recherche Clinique Inter-Régional 2008 programme.

2. Version française

2.1. Introduction

Au cours des dernières décennies, les progrès de la recherche et de la prise en charge des patients atteints d'une maladie neuromusculaire (MNM) ont permis une augmentation de l'espérance de vie de ces patients. L'évaluation s'est ainsi imposée lors du suivi clinique des patients, et dans le domaine

de la recherche clinique pour répondre aux premiers essais cliniques tant attendus. Une métrologie rigoureuse et adaptée est indispensable lors du développement d'outil d'évaluation. Il n'est pas possible de se contenter d'une quantification approximative ou d'utiliser des outils non adaptés à des pathologies évolutives.

Le symptôme commun à l'ensemble des MNM est la diminution de la force musculaire. Cette diminution de force peut toucher, selon la pathologie, différents muscles à des degrés variables. La cible thérapeutique dans les essais cliniques est de ce fait prioritairement la fibre musculaire et il est important de considérer, en particulier dans les essais thérapeutiques, les altérations des fonctions organiques comme la force (aspects mécaniques) avec leurs pertinences fonctionnelles, c'est-à-dire leurs impacts sur le mouvement et les retentissements de la faiblesse musculaire dans la vie quotidienne [12,14]. L'évaluation fonctionnelle motrice, qui mesure le retentissement de la maladie neuromusculaire sur les limitations fonctionnelles, est bien acceptée par le patient et les thérapeutes [5,10–12,14].

Proposer la passation d'une échelle évaluant la fonction motrice à un patient suppose la disponibilité d'un thérapeute formé à sa passation et un temps de passation d'environ une demi-heure quelle que soit l'échelle utilisée. Or, en pratique clinique courante ainsi que dans certains projets d'étude, une évaluation précise de la fonction motrice n'est pas nécessaire. Par exemple, pour la constitution de groupes de patients homogènes en termes de fonction motrice, une appréciation en niveaux ou grades de sévérité de l'atteinte fonctionnelle peut être suffisante. Dans le domaine des MNM, des scores fonctionnels existent déjà dans la littérature. Cependant, ces scores ne prennent pas en compte les différentes composantes de la fonction motrice (comme la position debout et les transferts, la motricité axiale/proximale et la motricité distale) ou ne sont pas suffisamment adaptés à la population étudiée. Par exemple, les scores de Brooke [3,4] et de Vignos [16], qui sont spécifiquement utilisés chez les patients atteints de dystrophie musculaire, ne s'intéressent qu'à la fonction des membres supérieurs pour le premier, et à la déambulation pour le second. De plus, peu de données sont disponibles dans la littérature concernant la validation de ces deux outils.

En s'inspirant des travaux de Palisano et al. [13] dans le domaine de la paralysie cérébrale, nous proposons de développer une classification en grades de sévérité d'atteinte fonctionnelle motrice, appelée NM-Score, pour chacun des 3 domaines de fonction motrice (D1 : position debout et transferts ; D2 : motricité axiale et proximale ; et D3 : motricité distale), précédemment décrit lors du développement et la validation de la Mesure de Fonction Motrice par Bérard et al. [1]. Nous souhaitons créer un outil simple et rapide à utiliser dont l'objectif principal est de classer les patients par analogie entre leur état fonctionnel dans leur environnement et la description clinique littérale correspondant à chaque grade de sévérité.

Les différentes étapes du développement de la classification NM-Score comprennent :

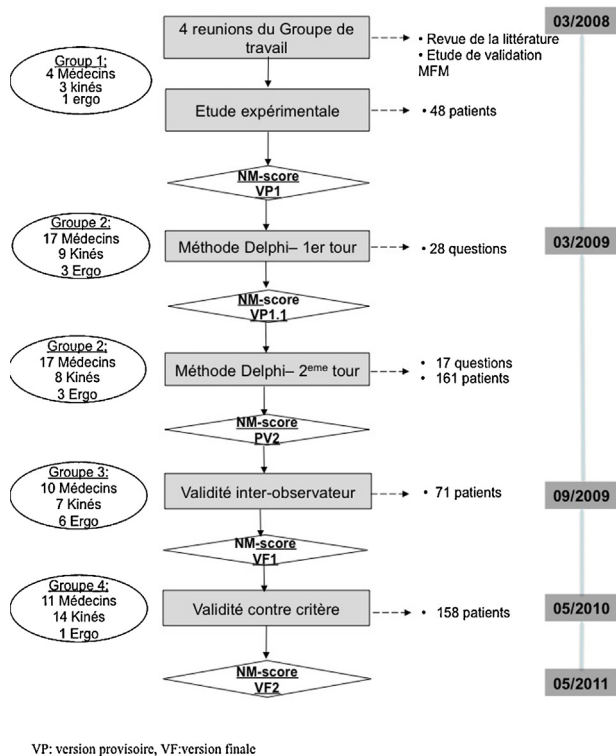


Fig. 1. Graphique représentant les différentes étapes de l'étude.

- la proposition d'une première classification par un groupe d'étude ;
- une étude de sa validité de contenu par une méthode Delphi ;
- une étude de validation inter-observateur ;
- une étude de validation contre critère.

À chaque étape, la classification est modifiée pour aboutir à la version finale présentée dans cet article.

2.2. Patients et méthodes

La Fig. 1 reprend les différentes étapes du développement de la classification NM-Score avec, pour chacune, un calendrier précis, la méthodologie utilisée et le nombre de patients et de professionnels impliqués. Cette étude, soumise au CCP Lyon Sud Est 2, a été considérée en dehors du champ d'application de la loi de bioéthique française. Une note d'information a été remise à chaque patient par leur thérapeute avant leur inclusion.

2.2.1. Construction d'une première version de la classification NM-Score

Une revue de la littérature exhaustive portant sur les différents outils d'évaluation de la fonction motrice disponibles dans le domaine des MNM a été réalisée. Huit experts (3 kinésithérapeutes, 1 ergothérapeute et 4 médecins) dans le domaine des MNM ont participé à l'élaboration de la première version du NM-Score (appelée NM-Score VP1) au cours de quatre réunions entre mars 2008 et février 2009. Quatre experts avaient une expérience exclusivement pédiatrique et 4 une expérience en milieu adulte. Leurs expériences variaient de 2 à 25 ans (moyenne : 13,7 ans).

Les différents niveaux de sévérité de NM-Score VP1 ont été définis pour chacun des trois domaines de la fonction motrice (D1 : position debout et transferts ; D2 : motricité axiale et proximale ; D3 : motricité distale) à partir de l'expérience clinique des experts, de la revue de la littérature et de l'analyse des données de l'étude de validation de la MFM provenant de 303 patients atteints de MNM [1]. Une étude expérimentale a été menée sur 48 patients atteints de MNM pour aider à définir chaque niveau du système de classification.

Pour chacun des trois domaines de la fonction motrice, les experts ont proposé un système de classification des fonctions motrices en cinq niveaux de gravité (normal : niveau 0 ; atteinte légère : niveau 1 ; atteinte moyenne : niveau 2 ; atteinte sévère : niveau 3 ; atteinte très sévère : niveau 4). Certains critères dits majeurs doivent être utilisés prioritairement pour classer les patients entre deux niveaux. Ces critères majeurs sont pour le domaine D1 : capacités habituelles de marche, course et sauts, pour le domaine D2 : capacités habituelles de tenue assise, de motricité proximale et de tenue de tête, et pour le domaine D3 : capacités habituelles de manipulation des objets. D'autres critères, appelés critères mineurs, permettent d'améliorer la description de chaque critère majeur avec un pouvoir discriminant plus modeste. Ces critères mineurs, identifiés lors de l'étude expérimentale, concernent les besoins en aides techniques ou humaines. Un critère mineur était retenu pour améliorer la description d'un niveau lorsqu'au moins 80 % des patients de l'étude expérimentale classés dans ce grade répondaient à ce critère mineur.

2.2.2. Analyse de la validité de contenu de la classification

Afin d'analyser la validité de contenu de la classification NM-Score VP1 proposée, une méthode de consensus de type méthode Delphi [2,7,9,15] a été utilisée. Cette étape a permis de récolter l'avis et les commentaires de plusieurs experts francophones dans le domaine des MNM sur la pertinence du contenu de la classification, en particulier des critères retenus.

La méthode Delphi est une procédure itérative et interactive comportant 2 à 4 tours. Les participants constituant le groupe d'experts sont choisis pour être « représentatifs des connaissances et/ou des perceptions actuelles, relativement impartiaux mais intéressés et impliqués dans la problématique abordée », comme conseillé par Fink et al. [7]. Après chaque tour, le groupe d'experts est informé des résultats et ces résultats sont ensuite utilisés pour construire le questionnaire pour le tour suivant. Le processus se termine lorsque le groupe atteint un niveau prédéfini d'accord ou lorsque des informations suffisantes ont été échangées pour atteindre les objectifs du processus.

Trente-six kinésithérapeutes, ergothérapeutes et médecins francophones (France, Belgique et Suisse) reconnus pour leur expertise dans les MNM et n'ayant pas participé à la première phase de l'étude ont été invités à participer ; 29 ont accepté (9 kinésithérapeutes, 3 ergothérapeutes et 17 médecins). Dix participants travaillaient en milieu adulte, 14 en pédiatrie, et 5 dans les deux milieux. Leurs expériences variaient de 3 à 28 ans (moyenne : 11,8 ans).

Pour le premier tour de la méthode Delphi, le questionnaire a été envoyé conjointement avec la classification NM-Score VP1 par e-mail. Le questionnaire a été divisé en quatre parties et comprenait 28 énoncés. Une réponse ouverte était associée à chaque énoncé pour encourager les experts à expliciter leurs réponses. La première partie (partie A) se composait de quatre questions portant sur la nécessité d'un nouveau système de classification. Les parties B, C et D, dédiées aux domaines de la fonction motrice D1, D2 et D3 respectivement, comportaient 24 questions portant sur la définition et la description de chacun des niveaux et la distinction entre les niveaux. Chaque énoncé était évalué à l'aide d'une échelle de Likert en neuf points ; 1 indiquant « fortement en désaccord », 5 « indifférence » et 9 « fortement d'accord » avec l'énoncé. Le consensus pour chaque énoncé est atteint s'il existe un haut degré d'accord ou en l'absence de désaccord. L'accord est défini par un score médian de l'accord entre experts ≥ 7 , avec au moins 80 % des experts ayant un score ≥ 7 . Le désaccord est défini par plus de 30 % des scores individuels des experts répartis entre 1 et 3 et entre 6 et 9.

Pour le second tour, le questionnaire comprenait 17 énoncés ; 11 énoncés du 1^{er} questionnaire qui n'ont pas obtenu de consensus au premier tour et 6 énoncés portant sur la facilité d'utilisation et les conditions d'administration de la classification. Les experts ont été invités à utiliser la classification NM-Score VP1 dans leur pratique quotidienne avant de répondre au questionnaire.

2.2.3. Étude de validation inter-observateur

Cette phase de validation avait pour but de mesurer le degré d'accord entre les cotations de différents professionnels pour mesurer la fiabilité inter-observateur de la classification. En se basant sur les explications données par les examinateurs en cas de désaccord lors d'une cotation, cette phase devait aussi nous aider à améliorer la classification afin d'aboutir à une nouvelle version, appelée NM-Score VF1.

Pour chaque patient remplissant les critères d'inclusion, deux examinateurs, médecin, kinésithérapeute ou ergothérapeute, devaient, chacun de leur côté et sans discussion au préalable entre eux, coter le patient grâce à la classification NM-Score à partir :

- de l'interrogatoire du patient ou de sa famille et/ou ;
- de l'observation en situation écologique et/ou ;
- du dossier médical du patient.

Un même duo de co-examineur ne pouvait coter que 5 patients au maximum. En cas de désaccord de cotation, une demande de confrontation était envoyée aux thérapeutes concernés afin de déterminer l'origine du désaccord.

Six consultations ou centres de références pour les MNM francophones (France et Belgique) ont participé à l'étude de validation inter-observateur. Les patients inclus devaient être âgés de 6 à 60 ans et porteurs d'une MNM, connus des deux examinateurs depuis au moins 3 mois et leur dernière consultation ou leur dernier bilan devait dater de moins de 3 mois le jour de la cotation. Les 23 thérapeutes ayant accepté de participer à cette étude de validation (7 kinésithérapeutes,

6 ergothérapeutes et 10 médecins) devaient, dans un premier temps, se familiariser avec la classification NM-Score avant de débiter les cotations.

2.2.4. Étude de validation contre critère (ou validité concurrente)

L'objectif de cette étude était la validation contre critère de la classification en utilisant des *gold standards* imparfait (score MFM, scores de Brooke et de Vignos). Pour une standardisation des évaluations, tous les thérapeutes de cette étude avaient été au préalable formés à la passation de la MFM et disposaient sur le cahier d'observation des classifications de Brooke et de Vignos.

Quatorze consultations ou centres de références pour les MNM ont participé à l'étude contre critère. Les patients inclus devaient être âgés de 6 à 60 ans et porteurs d'une MNM. Tous les thérapeutes ayant accepté de participer à l'étude de validation contre critère devaient, dans un premier temps, se familiariser avec la classification NM-Score avant de débiter les cotations : ils devaient au préalable lire attentivement les consignes de cotation et l'intégralité de la classification.

La cotation des patients par le thérapeute selon la classification NM-Score était réalisée le jour de la passation de la MFM et avant le début de la passation de la MFM.

2.2.5. Analyses statistiques

Les variables quantitatives ont été décrites par leurs moyennes, déviations standards (DS), minimum, maximum et quartiles. Les variables catégorielles ont été décrites par leurs effectifs et leurs pourcentages.

La fiabilité inter-observateur a été analysée pour chaque domaine (D1, D2 et D3) et selon les conditions de cotations (dossier médical $\pm$ interrogatoire $\pm$ observation en situation écologique) en utilisant le coefficient kappa de Cohen (κ) qui permet de déterminer si l'accord observé est différent d'un accord dû au hasard. Les coefficients kappa peuvent être interprétés comme faibles ($< 0,4$), modérée ($0,4$ à $0,59$), bon ($0,6$ à $0,8$) ou excellent ($> 0,8$) [6].

Pour analyser la validité contre critère de la classification NM-Score, nous avons choisi pour chaque domaine (D1, D2 et D3) des échelles de référence appelées *fuzzy gold standard* étant donné le caractère imparfait des *gold standards* dans ce domaine. Une analyse de la concordance a été réalisée entre la classification NM-Score (en 5 classes ordinales, sur chacun des trois domaines D1, D2 et D3) et la valeur de l'échelle de référence MFM (quantitative, prenant ses valeurs entre 0 et 100). La concordance entre le grade de sévérité et les scores de Vignos et de Brooke a aussi été étudiée. Nous avons utilisé des mesures de concordance, telles que la corrélation estimée par le coefficient de Spearman.

2.3. Résultats

2.3.1. Construction d'une première classification et analyse qualitative de la validité de contenu par méthode Delphi

Lors du premier tour de la méthode Delphi, un consensus a été atteint pour 17 des 28 énoncés concernant la classification

NM-Score VP1. Tous les experts se sont accordés sur la nécessité d'un nouveau système de classification pour les patients atteints de MNM basé sur leurs performances dans la vie quotidienne. Vingt-deux experts sur les 28 estiment que la classification a des applications potentielles à la fois dans la pratique clinique courante et dans la recherche clinique. Tous les experts pensent que l'un des grands intérêts de la classification NM-Score est de s'appuyer sur la description des performances du patient dans tous les domaines de la fonction motrice, plutôt que seulement sur les capacités de marche.

Après le premier tour de la méthode Delphi, la classification a été révisée à partir des commentaires envoyés par les experts et renvoyée conjointement avec le questionnaire du second tour. Un consensus a été atteint pour 16 des 17 énoncés. Le seul point non consensuel concernait l'âge à partir duquel la classification pourrait être administrée ; 13 experts sur les 28 pensaient que la classification ne peut pas être correctement utilisée chez des enfants dès 4 ans. Lors du deuxième tour du processus Delphi, la classification VP2 a été utilisée par les experts pour classer un total de 161 patients, le nombre maximal de patients classé par un expert était de 24, et tous les experts s'accordaient sur la facilité d'utilisation de la classification.

2.3.2. Analyses quantitatives

2.3.2.1. Première étude. Soixante et onze patients, 47 hommes et 24 femmes, âgés de 6 à 50 ans ont été inclus dans cette étude de validité inter-observateur. Leur âge moyen était de 18,7 ans (DS : 13). Soixante-six pour cent avaient moins de 18 ans le jour de l'évaluation.

Concernant le délai entre les 2 évaluations, pour 36 des 71 patients, elles ont été effectuées le même jour et dans 76 % des cas, le délai était de moins de 15 jours (le maximum étant de 3 mois).

Concernant les conditions des différentes cotations, 32 cotations ont été réalisées à partir de l'observation seule du patient, 20 uniquement à partir du dossier médical, et 9 uniquement par l'interrogatoire du patient et/ou de sa famille. Les autres évaluations ont été effectuées en utilisant plusieurs de ces options. Les coefficients kappa pour D1, D2 et D3 étaient respectivement de 0,72 (IC 95 % : 0,57 à 0,83), 0,62 (IC 95 % : 0,47 à 0,75) et 0,56 (IC 95 % : 0,39 à 0,71) (Tableau 1). Les pourcentages d'accords entre cotateurs (respectivement pour D1, D2 et D3 : 82,9, 78,1 et 75,6 %) ainsi que les coefficients κ (respectivement pour D1, D2 et D3 : 0,77 [IC 95 % : 0,59 à

0,90], 0,69 [IC : 95 % 0,48 à 0,85] et 0,64 [IC : 95 % 0,43 à 0,82]) étaient améliorés lorsque les évaluations étaient réalisées à partir de l'observation et/ou de l'interrogatoire du patient.

Dix-neuf cas de désaccord ont été analysés. Dans 15 cas, au moins un thérapeute n'avait pas lu correctement l'introduction ou la notice d'utilisation. En particulier, la priorité des critères majeurs par rapport aux critères mineurs pour la cotation des patients n'était pas respectée.

2.3.2.2. Deuxième étude. Un total de 158 patients a été inclus dans cette phase de validation. Il s'agit essentiellement de patients marchants (124). Une surreprésentation masculine de l'échantillon (102 patients), en lien avec un nombre important de patients atteints de dystrophie musculaire de Duchenne (27) ou de Becker (10) touchant exclusivement le sexe masculin, est observée. Les autres diagnostics représentés sont : myopathie et dystrophie musculaire congénitale (26), amyotrophie spinale infantile (22), neuropathies périphériques (21), dystrophie myotonique de Steinert (17), dystrophie musculaire facio-scapulo-humérale (12), myopathies des ceintures (8), autres MNM (15). La durée moyenne de passation de la classification NM-Score chez ces 158 patients est de 6,3 minutes (DS : 3,7). De manière à étudier la validité concurrente de la classification, 148 patients ont été évalués par le score de Vignos, 143 patients par le score de Brooke et 158 par la MFM.

Les distributions du grade de Vignos par grade NM-Score D1 et du grade de Vignos par grades NM-Score D2 et D3 sont représentées dans le Tableau 2. La répartition des patients dans les différents grades est relativement équilibrée pour D1 et D2, moins pour D3 où peu de patients sont classés dans le grade 4. La dispersion du score de Vignos pour les grades 0 et 4 de la classification NM-Score D1 étant faible, un effet plancher et un effet plafond pour au moins l'une des deux échelles peut être suspecté. Cependant, la faible taille de l'effectif (8 patients) pour le grade 0 du domaine D1 de l'échelle NM-Score rend difficile l'affirmation d'un effet plancher. Les moyennes des scores de Vignos augmentent régulièrement du grade 0 au grade 4 du NM-Score D1. Seul le grade 3 montre une forte dispersion du score de Vignos (moyenne : 5,75, minimum : 2, maximum : 9) témoignant d'une grande variabilité de l'atteinte des patients classés dans ce grade. Pour NM-Score D2 et D3, la dispersion du score de Brooke est forte et quasi comparable entre les 2 niveaux pour les grades 0 et 1 (NM-Score D2 grade 0 [1–3] grade 1 [1–5], NM-Score D3 grade 0 [1–3] grade 1 [1–4]). Pour les autres grades, les moyennes et distributions des scores de

Tableau 1

Pourcentage d'accord entre cotateurs et fiabilité inter-observateur par le coefficient Kappa (κ) pour chaque domaine de la fonction motrice (D1, D2 et D3).

Domaine	Population totale (n = 71)		Patients cotés à partir de l'observation et/ou de l'interrogatoire (n = 41)	
	% d'accord	κ (IC 95 %) ^a	% d'accord	κ (IC 95 %) ^a
D1	78,87	0,72 (0,57–0,83)	82,93	0,77 (0,59–0,90)
D2	71,83	0,62 (0,47–0,75)	78,05	0,69 (0,48–0,85)
D3	70,42	0,56 (0,39–0,71)	75,61	0,64 (0,43–0,82)

^a Intervalle de confiance (IC) calculé par la méthode du Bootstrap.

Tableau 2

Distribution du score de Vignos par grades NM-Score D1 et du score de Brooke par grades NM-Score D2 and D3.

NM-Score D1	Nombre de patients	Score de VIGNOS					
		Moyenne (DS)	Minimum	p25	p50	p75	Maximum
0	8	1 (0)	1	1	1	1	1
1	57	1,6 (0,7)	1	1	1	2	3
2	38	2,76 (1,3)	1	2	2	3	6
3	16	5,75 (2,24)	2	3,5	6,5	7	9
4	29	8,93 (0,26)	8	9	9	9	9
Total	158	3,75 (3,05)	1	1	2	6,5	9

NM-Score D2	Nombre de patients	Score de BROOKE					
		Moyenne (DS)	Minimum	p25	p50	p75	Maximum
0	44	1,2 (0,51)	1	1	1	1	3
1	65	1,6 (0,8)	1	1	1	2	5
2	17	2,65 (1,0)	1	2	3	3	4
3	11	3,9 (1,0)	2	3	4	5	5
4	6	5,2 (0,4)	5	5	5	5	6
Total	143	1,9 (1,3)	1	1	1	3	6

NM-Score D3	Nombre de patients	Score de BROOKE					
		Moyenne (DS)	Minimum	p25	p50	p75	Maximum
0	42	1,4 (0,7)	1	1	1	2	3
1	52	1,4 (0,7)	1	1	1	2	4
2	50	2,6 (1,3)	1	1	2	3,5	5
3	7	4,1 (1,2)	2	3	5	5	5
4	2	5,5 (0,7)	5	5	5,5	6	6
Total	143	1,9 (1,3)	1	1	1	3	6

p25 : premier quartile ; p50 : deuxième quartile ; p75 : troisième quartile.

Brooke augmentent régulièrement du grade 2 au grade 4 du NM-Score D2 et D3.

Les distributions des scores MFM D1, D2 et D3 par grades NM-Score D1, D2 et D3 sont représentées sous la forme de box plot sur la Fig. 2. Pour les NM-scores D2 et surtout D3 contrairement au NM-score D1, la différence entre les grades 0 et 1 n'est pas significative en termes de moyenne de score de Brooke et de score MFM. La dispersion du score D3 de la MFM pour les grades 3 et 4 de l'échelle NM-Score D3 est importante, cependant les effectifs de patients sont petits (respectivement 7 et 3). Ceci est à prendre en compte dans l'interprétation des box plots. De plus, pour le grade 4, le premier quartile (p25) et la médiane (p50) sont égaux, ce qui n'est pas visible sur le graphique des box plots.

La validité contre critère a été estimée entre chaque domaine de la classification NM-Score et les échelles de références par le calcul du coefficient de Spearman et de son intervalle de confiance à 95 % (Tableau 3). Les meilleures corrélations sont retrouvées pour le NM-Score D1 et NM-Score D2 avec une corrélation plus importante avec l'échelle MFM qu'avec les scores fonctionnels.

2.4. Discussion

La classification NM-Score est un outil valide et fiable permettant de classer les patients en 4 catégories selon leur grade de sévérité fonctionnelle dans les 3 domaines de la fonction motrice : D1 : position debout et transferts ; D2 : motricité axiale et proximale ; et D3 : motricité distale.

Cette nouvelle classification a été bien accueillie par les nombreux praticiens et chercheurs qui l'ont déjà utilisée. La version française du NM-Score peut être librement téléchargée sur le site de la MFM (<http://www.mfm-nmd.org/accueil.aspx>).

La classification NM-Score est un outil facile et rapide à utiliser pour décrire les performances motrices des patients. Aucune formation n'est nécessaire à son utilisation ; les utilisateurs sont seulement priés de lire attentivement l'introduction et les instructions de l'utilisateur et de faire quelques essais de cotations avant une utilisation fiable pour une meilleure connaissance de l'outil. Nous pensons que cette classification facilite les discussions entre médecin et patient en s'axant plus sur le quotidien de manière à identifier les facteurs aggravant ou facilitant la fonction en situation écologique.

Cette classification permet aux professionnels d'avoir un langage commun lorsqu'ils discutent de leurs patients et elle permet également de tenir compte de l'avis du patient ou de sa famille car elle peut être renseignée par l'interrogatoire.

La classification NM-Score fournit une description du patient sous la forme d'un grade de sévérité qui n'est pas limité à la capacité de marche. La classification NM-Score est basée sur la description clinique de la fonction motrice permettant de fournir « une image complète » de la fonction des patients dans tous les domaines. Cette classification peut être assimilée à la classification TNM utilisée en cancérologie pour typer les atteintes tumorales en fonction de différents critères. En fonction de la pathologie ou de l'âge, un « pattern » spécifique

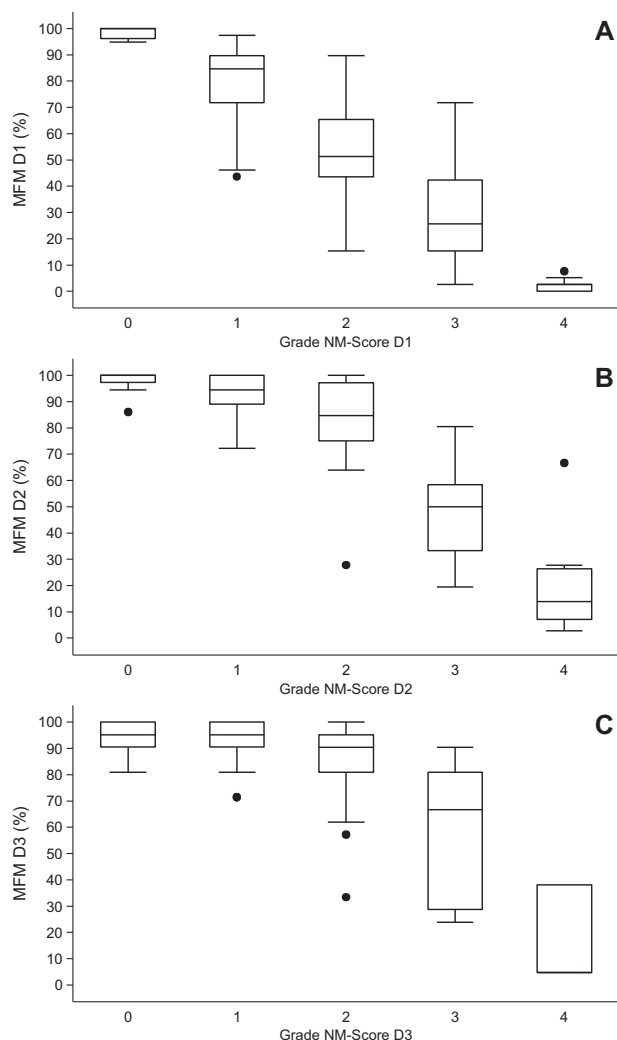


Fig. 2. Box plots des distributions des scores MFM D1 (A), D2 (B) et D3 (C) par grades NM-Score D1, D2 et D3 correspondant.

Tableau 3

Coefficients de Spearman entre chaque domaine de la classification NM-Score et les échelles de références (validité contre critère).

Échelles de références	Coefficient de Spearman	Intervalle de confiance à 95 %	
VIGNOS vs NM-Score D1	0,86	0,81	0,89
MFM D1 vs NM-Score D1	−0,88	−0,91	−0,83
BROOKE vs NM-Score D2	0,64	0,53	0,73
MFM D2 vs NM-Score D2	−0,70	−0,77	−0,60
BROOKE vs NM-Score D3	0,49	0,36	0,61
MFM D3 vs NM-Score D3	−0,49	−0,60	−0,36

peut être observé. Par exemple, dans l'amyotrophie spinale infantile ou la dystrophie musculaire facio-scapulo-humérale ou l'atteinte est essentiellement proximale, les NM-Scores D1 et D2 sont les plus altérés quel que soit l'âge des patients. Dans la dystrophie musculaire de Duchenne, alors que le NM-Score D1 est altéré dès les premières années de la maladie, les NM-Scores D2 et D3 sont altérés de façon plus tardive (le NM-Score D2 étant altéré plus précocement que le NM-Score D3).

Durant le processus Delphi, les experts ont été unanimes sur la nécessité d'un système de classification basé sur les limitations fonctionnelles des patients dans leur vie quotidienne et non sur leurs capacités en situation de test. La majorité des experts ont indiqué que notre classification est utile à la fois à des fins de recherche clinique pour constituer des groupes de patients de fonction motrice comparable, ainsi que pour le suivi clinique des patients. Deux tours de processus Delphi ont été nécessaires afin d'aboutir à un consensus sur l'ensemble des descriptions proposées pour chaque grade. Près de 30 % des experts estiment que le NM-Score ne peut être utilisé avec une précision suffisante chez les enfants de moins de 6 ans. Afin d'éviter les difficultés en lien avec la variabilité du développement psychomoteur normal de l'enfant nous proposons d'utiliser le NM-Score à partir de l'âge de 6 ans.

La fiabilité inter-observateur de la classification NM-Score est bonne lorsqu'elle est utilisée à partir de l'interrogatoire et/ou de l'observation du patient en situation écologique. Cette condition d'administration est nécessaire pour une fiabilité suffisante ($\kappa > 0,60$) pour tous les domaines (0,77, 0,69 et 0,64 pour respectivement les domaines D1, D2 et D3) et autorise l'utilisation du NM-Score dans des essais thérapeutiques multicentriques. Pour comparaison, Palisano et al. [13] ont rapporté pour le « Gross Motor Function Classification System » (GMF-CS) un κ de 0,75 chez les enfants âgés de 2 ans et plus, et Finkel et al. [8] rapportent un κ pondéré de 0,61 (IC 95 % : 0,59 à 0,63) pour le « Test of Infant Motor Performance » (TIMP). Les comparaisons avec les autres outils sont rendues difficiles par l'utilisation des coefficients intra-classe (ICC) pour estimer leurs fiabilités inter-observateurs. À titre d'exemple, l'ICC est estimé entre 0,96 et 0,99 pour le score de Brooke [3,4]. Par nature, le coefficient κ est bien adapté à des variables ordinales, comme les classifications de Brooke, de Vignos, et le NM-Score. L'ICC, qui est considéré comme plus adapté à des variables quantitatives, est aussi plus dépendant de la variabilité de l'échantillon que le κ (plus la variabilité est élevée, plus l'ICC est proche de 1). Ainsi, à paramètres identiques, le κ donnera une valeur inférieure à l'ICC. L'analyse des désaccords entre les évaluateurs appariés lors de l'étude inter-observateur a montré que, dans la plupart des cas, un des 2 cotateurs au moins n'avait pas respecté l'utilisation prioritaire des critères majeurs pour la cotation. Nous avons donc proposé dans la version finale de la classification d'isoler sur une seconde page les critères mineurs pour limiter leurs utilisations seulement en cas de doute entre 2 grades (à partir des critères majeurs).

Concernant la validité contre critère, le NM-Score D1 est très bien corrélé (coefficient de Spearman = 0,86) au score de Vignos avec une suspicion d'un petit effet plancher du fait de la dispersion faible des scores de Vignos pour le grade 4.

Le score de Brooke est un score fonctionnel qui permet de mesurer les performances motrices des membres supérieures. En fonction des grades de la classification de Brooke, ce score mesure la motricité proximale (grade 1, 2 et 3) et/ou la motricité distale (grade 3, 4, 5 et 6) des membres supérieurs. Ses corrélations avec le NM-Score, bien qu'étant meilleure en D2 (coefficient de Spearman = 0,64) qu'en D3 (coefficient de

Spearman = 0,49), sont inférieures au coefficient obtenu pour D1. La significativité statistique et la largeur des intervalles de confiance des coefficients de corrélation étant très en lien avec l'effectif de la population, l'interprétation de ces corrélations est rendue délicate par des effectifs insuffisants de patients ayant une atteinte très sévère en D2 ou D3.

Lorsque l'on s'intéresse aux corrélations entre la MFM et le NM-Score, les coefficients de corrélations sont satisfaisants pour D1 et D2, moyens pour D3. D'une manière générale, dans les MNM, le domaine D3 de la fonction motrice est le plus longtemps préservé (sauf pour les MNM purement distales) ; la répartition des patients par domaine et par grade de la classification le montre bien, et constitue probablement une limite dans l'interprétation des résultats. Au regard des résultats, le caractère discriminant entre les grades 0 et 1 des NM-Score D2 et D3 apparaît insuffisant même si nous avons souhaité garder un niveau 0 représentant les capacités normales d'un sujet, à différencier d'une atteinte même minime (grade 1) comme par exemple une fatigabilité dans la tenue assise au cours de la journée ou un manque de précision du geste qui peuvent être retrouvés à l'interrogatoire approfondi du patient.

Une traduction en langue anglaise autorisant une diffusion internationale de notre outil est en cours.

Déclaration d'intérêts

Les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

Remerciements

Les auteurs tiennent à remercier en premier lieu les patients et leur famille qui comme toujours ont participé avec un grand enthousiasme à ce projet. Nous remercions également tous les médecins et rééducateurs du groupe de travail NM-Score qui ont permis à ce travail d'aboutir. Nous remercions les Hospices Civils de Lyon pour leur soutien financier et logistique de ce Projet Hospitalier de Recherche Clinique Inter-Régional 2008.

References

- [1] Bérard C, Payan C, Hodgkinson I, Fermanian J, the MFM collaborative study group. A motor function measure scale for neuromuscular diseases. Construction and validation study. *Neuromuscul Disord* 2005;15:463–70.
- [2] Beretta R. A critical review of the Delphi technique. *Nurs Res* 1996;3:79–89.
- [3] Brooke MH, Griggs RC, Mendell JR, Fenichel GM, Shumate JB, Pellegrino RJ. Clinical trial in Duchenne dystrophy. I. The design of the protocol. *Muscle Nerve* 1981;4:186–97.
- [4] Brooke MH, Fenichel GM, Griggs RC, et al. Clinical investigation in Duchenne dystrophy: 2. Determination of the “power” of therapeutic trials based on the natural history. *Muscle Nerve* 1983;6:91–103.
- [5] Bosboom WM, Vrancken AF, van den Berg LH, Wokke JH, Iannaccone ST. Drug treatment for spinal muscular atrophy types II and III. *Cochrane Database Syst Rev* 2009;21:CD006282.
- [6] Colton T. Statistics in medicine. Boston: Little, Brown and company; 1974 : 372.
- [7] Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984;74:979–83.
- [8] Finkel RS, Hynan LS, Glanzman AM, Owens H, Nelson L, Cone SR, et al. The test of infant motor performance: reliability in spinal muscular atrophy type I. *Pediatr Phys Ther* 2008;20:242–6.
- [9] Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Medical Care* 2000;38:28–42.
- [10] Iannaccone ST, American Spinal Muscular Atrophy Randomized Trials (AmSMART) Group. Outcome measures for pediatric spinal muscular atrophy. *Arch Neurol* 2002;59:1445–50.
- [11] Iannaccone ST, Hynan LS. Reliability of 4 outcome measures in pediatric spinal muscular atrophy. *Arch Neurol* 2003;60:1130–6.
- [12] Mercuri E, Mayhew A, Muntoni F, Messina S, Straub V, Van Ommen GJ, et al. Towards harmonisation of outcome measures for DMD and SMA within TREAT-NMD; report of three expert workshops: TREAT-NMD/ENMC workshop on outcome measures, 12th–13th May 2007, Naarden, The Netherlands; TREAT-NMD workshop on outcome measures in experimental trials for DMD, 30th June–1st July 2007, Naarden, The Netherlands; conjoint Institute of Myology TREAT-NMD meeting on physical activity monitoring in neuromuscular disorders, 11th July 2007, Paris, France. *Neuromuscul Disord* 2008;18:894–903.
- [13] Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol* 1997;39:214–23.
- [14] Scott E, Mawson SJ. Measurement in Duchenne muscular dystrophy: considerations in the development of a neuromuscular assessment tool. *Dev Med Child Neurol* 2006;48:540–4.
- [15] Sumsion T. The Delphi technique: an adaptive research tool. *Br J Occup Ther* 1998;61:153–6.
- [16] Vignos PJ, Spencer GE, Archibald KC. Management of progressive muscular dystrophy of childhood. *J Am Med Assoc* 1963;184:89–96.