



HAL
open science

Métrologie et évaluation fonctionnelle motrice dans les maladies neuromusculaires de l'enfance : Illustrations à partir de la Mesure de Fonction Motrice (MFM) et d'une classification en grades de sévérité d'atteinte fonctionnelle motrice (NM-Score)

Carole Vuillerot

► To cite this version:

Carole Vuillerot. Métrologie et évaluation fonctionnelle motrice dans les maladies neuromusculaires de l'enfance : Illustrations à partir de la Mesure de Fonction Motrice (MFM) et d'une classification en grades de sévérité d'atteinte fonctionnelle motrice (NM-Score). Sciences agricoles. Université Claude Bernard - Lyon I, 2012. Français. NNT : 2012LYO10081 . tel-01127566

HAL Id: tel-01127566

<https://theses.hal.science/tel-01127566>

Submitted on 7 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2012

THESE DE L'UNIVERSITE DE LYON

Délivrée par
L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE
Evolution Ecosystèmes Microbiologie Modélisation (E2M2)

DIPLOME DE DOCTORAT
(arrêté du 7 août 2006)

Soutenue publiquement le 28 juin 2012
par

Madame VUILLEROT Carole

**Métriologie et évaluation fonctionnelle motrice dans les maladies neuromusculaires
de l'enfance:**

**Illustrations à partir de la Mesure de Fonction Motrice (MFM) et d'une
classification en grades de sévérité d'atteinte fonctionnelle motrice (NM-Score)**

Directeur de thèse : Monsieur le Pr ECOCHARD René

JURY : Monsieur le Pr BENAÏM Charles
Madame de Dr BERARD Carole
Monsieur le Pr ECOCHARD René
Monsieur le Pr GAUTHERON Vincent (président du jury)
Madame le Pr SCHOTT PETHELAZ Anne Marie

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

M. François-Noël GILLY

Vice-président du Conseil d'Administration

M. le Professeur Hamda BEN HADID

Vice-président du Conseil des Etudes et de la Vie Universitaire

M. le Professeur Philippe LALLE

Vice-président du Conseil Scientifique

M. le Professeur Germain GILLET

Secrétaire Général

M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Directeur : M. le Professeur J. ETIENNE

Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux

Administrateur provisoire : M. le Professeur G. KIRKORIAN

UFR d'Odontologie

Directeur : M. le Professeur D. BOURGEOIS

Institut des Sciences Pharmaceutiques et Biologiques

Directeur : Mme la Professeure C. VINCIGUERRA.

Institut des Sciences et Techniques de la Réadaptation

Directeur : M. le Professeur Y. MATILLON

Département de formation et Centre de Recherche en Biologie Humaine

Directeur : M. le Professeur P. FARGE

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Directeur : M. le Professeur F. De MARCHI

Département Biologie

Directeur : M. le Professeur F. FLEURY

Département Chimie Biochimie

Directeur : Mme le Professeur H. PARROT

Département GEP

Directeur : M. N. SIAUVE

Département Informatique

Directeur : M. le Professeur S. AKKOUCHE

Département Mathématiques

Directeur : M. le Professeur A. GOLDMAN

Département Mécanique

Directeur : M. le Professeur H. BEN HADID

Département Physique

Directeur : Mme S. FLECK

Département Sciences de la Terre

Directeur : Mme la Professeure I. DANIEL

UFR Sciences et Techniques des Activités Physiques et Sportives

Directeur : M. C. COLLIGNON

Observatoire de Lyon

Directeur : M. B. GUIDERDONI

Polytech Lyon

Directeur : M. P. FOURNIER

Ecole Supérieure de Chimie Physique Electronique

Directeur : M. G. PIGNAULT

Institut Universitaire de Technologie de Lyon 1

Directeur : M. C. VITON

Institut Universitaire de Formation des Maîtres

Directeur : M. R. BERNARD

Institut de Science Financière et d'Assurances

Directeur : Mme la Professeure V. MAUME-DESCHAMPS

REMERCIEMENTS

AU DIRECTEUR DE THESE MONSIEUR LE PROFESSEUR RENE ECOCHARD

Professeur de Biostatistiques
Praticien hospitalier

Depuis 6 ans maintenant, j'ai l'honneur de travailler à vos côtés en bénéficiant de votre expérience dans le domaine de la mesure appliquée à l'humain. Je vous remercie sincèrement pour votre disponibilité tout au long de ce parcours parsemé de discussions passionnantes, importantes sources d'inspiration pour ce travail. Sans votre accompagnement et vos encouragements constants, je n'aurais pu faire aboutir ce projet de recherche. Soyez assuré, Monsieur, de toute ma reconnaissance et de mon admiration.

AU PRESIDENT DU JURY MONSIEUR LE PROFESSEUR VINCENT GAUTHERON

Professeur en médecine physique et réadaptation
Chef de service

Au niveau de la région, au niveau de la France, vous faites vivre la discipline de MPR pédiatrique. Vous m'avez très vite fait confiance et encouragé à poursuivre ma carrière universitaire dans ce domaine pour vous aider à développer notre spécialité. Je vous remercie de l'honneur que vous me faites en acceptant de juger ce travail.

AUX MEMBRES DU JURY MADAME LE DOCTEUR BERARD

Docteur en médecine physique et réadaptation
Praticien hospitalier

Vous m'avez fait découvrir et aimer la rééducation fonctionnelle pédiatrique et je vous en remercie. Votre grand respect humain envers les enfants et leur famille est un modèle pour moi. Votre enseignement de la médecine basé sur le respect des patients, la rigueur et l'humilité dans nos actions médicales m'a été d'une grande aide et m'accompagnera tout au long de ma carrière dans l'exercice de cette médecine si particulière. Votre bienveillance à mon égard, vos irremplaçables conseils et votre ténacité nous ont permis de faire aboutir beaucoup de projets. Veuillez trouver ici, Carole, le témoignage de toute ma reconnaissance et de mon admiration.

MADAME LE PROFESSEUR ANNE-MARIE SCHOTT PETHELAZ

Professeur de santé publique
Praticien hospitalier

Votre parcours professionnel est un modèle pour moi. Vos conseils avisés pour mener à bien mes projets scientifiques m'ont été d'une grande aide. Je vous remercie d'avoir accepté de juger ce travail.

MONSIEUR LE PROFESSEUR CHARLES BENAÏM

Professeur en médecine physique et réadaptation

Praticien hospitalier

Je vous remercie de l'honneur que vous me faites en acceptant de juger ce travail. Sachez que je suis extrêmement sensible à votre participation à l'évaluation de ce projet scientifique et que je suis honorée que vous ayez accepté de l'enrichir de votre savoir et de votre expérience dans ce domaine.

Je souhaite remercier très chaleureusement tous les professionnels du service de l'Escale pour leur accueil, leur soutien indéfectible dans tous les projets de recherche menés et à venir. Vous faites preuve d'un extraordinaire professionnalisme, dans les soins prodigués à ces enfants et l'accompagnement de leur famille dans ce domaine si particulier de la médecine. Isabelle, tu es l'initiatrice de ce travail, en me « confiant » à René Ecochard pour une année de master, tu m'as mise sur la route de la carrière universitaire et je t'en remercie.

Merci à Pascal, nouvel arrivant dans le service pour son aide inestimable dans ce travail.

Je souhaite également remercier le service de Biostatistiques des HCL et en particulier Stéphanie, Sylvain et Jean qui ont toujours été disponibles pour moi et mes demandes parfois (très) urgentes.

Un grand merci à tous les enfants et à leur famille ainsi qu'à tous les professionnels français et étrangers qui ont permis ce travail de recherche.

Enfin, je souhaite remercier ma famille et mes amis pour leur soutien sans faille et leur intérêt constant pour mon travail. Cette thèse, aboutissement de longues années d'études, je la dois beaucoup à mes parents qui m'ont toujours soutenu et toujours aidé. A mon père.

A tous mes amis, un grand merci pour leur soutien même si certains ont semblé quelque peu dérouté par le contenu de mes recherches (mais si Julien c'est promis cette fois ci il y a des données !). Ils ont pourtant toujours été là pour me soutenir et m'encourager à poursuivre dans les moments de moins bien. Je vais certainement en oublier, mais je pense à Manuel, Julien, Aurélie, Virginie, Cyril, Muriel, Nathalie, Isabelle, Marie, Anne-Sophie, Anne...

Pour conclure, je souhaite bien évidemment remercier mon extraordinaire compagnon Pascal et nos deux petits amours Charlie et Violette. C'est votre présence à mes côtés qui m'a donné les ailes nécessaires à la réalisation de ce travail, merci.

RESUME EN FRANÇAIS

Les progrès de la recherche et de la prise en charge des maladies neuromusculaires de l'enfance ont prolongé la survie des patients. L'évaluation s'impose donc pour le suivi des patients et aussi en recherche clinique car les premiers essais cliniques tant attendus commencent à paraître. Une métrologie rigoureuse et adaptée est alors indispensable parce qu'il n'est possible ni de se contenter d'une quantification approximative ni d'utiliser des outils non adaptés à des pathologies évolutives. Nous résumons l'état des connaissances sur la métrologie appliquée à l'évaluation fonctionnelle motrice des patients atteints de maladies neuromusculaires et proposons une revue de la littérature sur les outils disponibles avec des analyses précises de leurs propriétés métrologiques. La Mesure de Fonction Motrice, développée à partir de 1998, présente des qualités intéressantes en termes de validité et de fiabilité. Nous avons analysé sa sensibilité au changement dans différentes populations de patients adultes et enfants. Nous proposons ensuite, une classification en grades de sévérité d'atteinte fonctionnelle motrice, le NM-score. Les études de validation ont confirmé son intérêt, sa facilité d'utilisation, sa validité et sa reproductibilité. Le NM-Score permet de décrire précisément et de façon discriminante les patients en termes de fonction motrice pour la position debout et les transferts, la motricité axiale ou proximale et la motricité distale. S'intéresser à l'évaluation et à la mesure en médecine, c'est faire preuve d'une rigueur indispensable aux décisions de soins touchant des personnes vulnérables aux besoins spécifiques.

TITRE EN ANGLAIS

Metrology and motor function assessment in childhood neuromuscular diseases: Illustration with the Motor Function Measure (MFM) and a classification in levels of motor function decline severity (NM-Score)

RESUME EN ANGLAIS

Advances in the research and treatment of childhood neuromuscular diseases have led to longer patient survivals. Evaluation is thus required not only in clinical practice for patient follow-up but also in medical research because the results of long-awaited clinical trials are beginning to emerge. A rigorous and appropriate metrology is then necessary because rough estimates or the use of improper assessment tools are no more satisfactory. We summarize here the current knowledge on the metrology applied to motor function assessment of patients with neuromuscular diseases. We propose a review of the literature on the tools available to monitor motor function with detailed analyses of their metrological properties. Developed since 1998, the Motor Function Measure presents interesting properties in terms of validity and reliability. We analyzed its sensitivity to change in different patient populations of adults and children. We then propose, the NM-Score, a classification in levels of severity of motor function decline. Validation studies have confirmed the interest of this score as well as its ease of use, validity, and reproducibility. The NM-Score is able to describe the patients precisely and discriminantly in terms of motor function for standing position and transfers, axial / proximal motor function and distal motor function. Being interested in evaluation and measurement in medicine is a sign of rigor necessary for decision-making regarding vulnerable persons with special need.

DISCIPLINE

Epidémiologie clinique

MOTS-CLES

Maladies neuromusculaires

Echelle d'évaluation

Métriologie

Critère d'évaluation

Essais thérapeutiques

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE :

UMR CNRS 5558 Laboratoire de Biométrie et Biologie Évolutive Equipe Biostatistiques santé
UCB Lyon 1 - Bât. Grégor Mendel 43 bd du 11 novembre 1918 69622 VILLEURBANNE cedex

TABLES DES MATIERES

ABREVIATIONS	15
---------------------------	-----------

INTRODUCTION.....	17
--------------------------	-----------

CHAPITRE 1 : CONTEXTE ET DEMARCHES DE LA RECHERCHE 21

A L'ÉVALUATION EN MÉDECINE.....	21
<u>A.1</u> JUSTIFICATIONS DE L'ÉVALUATION EN MÉDECINE.....	21
<u>A.2</u> DIFFÉRENCIATION DE L'ÉVALUATION DES MOYENS DES PROCÉDURES ET DES RESULTATS.....	24
<u>A.3</u> LE PRINCIPAL ÉCHAPPE À L'ÉVALUATION	25
B L'ÉVALUATION DANS LE DOMAINE DES MALADIES NEUROMUSCULAIRES DE L'ENFANCE	26
<u>B.1</u> L'ÉVALUATION EN MÉDECINE PHYSIQUE ET READAPTATION.....	26
<u>B.2</u> L'ÉVALUATION EN RECHERCHE CLINIQUE.....	27
C DÉFINIR LE CONCEPT AVANT DE TENTER DE LE MESURER	30
<u>C.1</u> QUELS SONT LES CONCEPTS À MESURER DANS LES MALADIES NEUROMUSCULAIRES ?	30
<u>C.2</u> DÉFINITION DE LA PLACE DE L'ÉVALUATION FONCTIONNELLE MOTRICE DANS LES MALADIES NEUROMUSCULAIRES	34
<u>C.3</u> ET LA QUALITÉ DE VIE.....	35
D CHOISIR UN BON OUTIL D'ÉVALUATION POUR OBTENIR UNE MESURE VALIDE DE LA FONCTION MOTRICE	37
<u>D.1</u> LES DIFFÉRENTS TYPES D'OUTILS.....	37
<u>D.1.1</u> LES SCORES FONCTIONNELS	37
<u>D.1.2</u> LES TESTS CHRONOMETRES	37
<u>D.1.3</u> LES ÉCHELLES DE MESURE	38
<u>D.2</u> UNE MÉTHODOLOGIE ROBUSTE DE VALIDATION : LA THÉORIE DES QUESTIONNAIRES	38
<u>D.2.1</u> LA THÉORIE CLASSIQUE DES TESTS VS LA THÉORIE DE RÉPONSES AUX ITEMS	38
<u>D.2.2</u> LES QUALITÉS MÉTROLOGIQUES D'UN INSTRUMENT DE MESURE [RODARY 2001][FERMANIAN 2005][ANDERSEN 2000].....	40
<u>D.2.2.1</u> La Fidélité (« reliability »)	40

<u>D.2.2.2</u>	La validité.....	42
<u>D.2.2.3</u>	Sa sensibilité au changement (Responsiveness) [Beaton 2001].....	43
<u>D.2.2.4</u>	L'utilité ou applicabilité en pratique	43

CHAPITRE 2 : DEVELOPPEMENT ET TRAVAUX 47

A	REVUE DE LA LITTERATURE DES DIFFERENTS INSTRUMENTS	47
<u>A.1</u>	OBJECTIFS.....	47
<u>A.2</u>	METHODES	47
<u>A.2.1</u>	DE LA RECHERCHE BIBLIOGRAPHIQUE.....	47
<u>A.2.2</u>	DE L'ANALYSE DES QUALITES METROLOGIQUES	48
<u>A.3</u>	RESULTATS DE LA REVUE DE LA LITTERATURE	48
<u>A.3.1</u>	TABLEAU RECAPITULATIF	49
<u>A.3.2</u>	DESCRIPTION DETAILLEE DE CHAQUE OUTIL	49
<u>A.4</u>	CONCLUSIONS DE LA REVUE DE LA LITTERATURE	69
B	ETUDES DE VALIDATION DE LA MFM : LA SENSIBILITE AU CHANGEMENT.....	72
<u>B.1</u>	LA SENSIBILITE AU CHANGEMENT SUR LA POPULATION GLOBALE.....	72
<u>B.2</u>	LA SENSIBILITE AU CHANGEMENT PAR PATHOLOGIE	73
<u>B.2.1</u>	DANS LA DYSTROPHIE MUSCULAIRE DE DUCHENNE.....	73
<u>B.2.2</u>	DANS L'AMYOTROPHIE SPINALE INFANTILE	74
C	LA CREATION D'UNE CLASSIFICATION EN GRADES DE SEVERITE DANS LE DOMAINE DES MNM DE L'ENFANCE : LA CLASSIFICATION NM-SCORE.....	75
<u>C.1</u>	JUSTIFICATIONS SCIENTIFIQUES DU PROJET.....	75
<u>C.2</u>	ETUDE PRELIMINAIRE A PARTIR DES DONNEES DE L'ETUDE DE VALIDATION INITIALE DE LA MFM.....	77
<u>C.3</u>	MATERIELS ET METHODES	78
<u>C.3.1</u>	TABLEAU SYNOPTIQUE DE L'ETUDE	78
<u>C.3.2</u>	DEVELOPPEMENT DE LA CLASSIFICATION.....	79
<u>C.3.2.1</u>	Construction d'une première classification.....	79
<u>C.3.2.2</u>	Etude de validation de contenu : Méthode Delphi	79
<u>C.3.3</u>	ETUDE DE VALIDATION INTER OBSERVATEUR.....	81
<u>C.3.4</u>	ETUDE DE LA VALIDITE CONTRE CRITERE	82
<u>C.3.5</u>	VALIDITE DE CONSTRUIT	83
<u>C.4</u>	RESULTATS.....	85
<u>C.4.1</u>	DEVELOPPEMENT DE LA CLASSIFICATION ET VALIDATION INTER-OBSERVATEUR.....	85
<u>C.4.1.1</u>	L'accord au sein du groupe sur le cahier des charges précis de la classification.....	85

<u>C.4.1.2</u>	Le choix et la définition des critères dits majeurs pour classer les patients dans chacun des grades pour chaque domaine de la classification (D1, D2, D3).....	86
<u>C.4.1.3</u>	Etude expérimentale.....	86
<u>C.4.1.4</u>	Méthode Delphi et Fidélité inter-observateur	87
<u>C.4.2</u>	VALIDATION CONTRE CRITERE	88
<u>C.4.2.1</u>	Description de la population	88
<u>C.4.2.2</u>	NM-Score versus gold standards.....	89
<u>C.4.2.3</u>	Validité contre critère : étude de corrélations	90
<u>C.4.3</u>	VALIDATION DES CONSTRUITS	91

CHAPITRE 3 : DISCUSSION ET PERSPECTIVES DE RECHERCHE..... 95

A	DISCUSSION.....	95
<u>A.1</u>	MFM.....	95
<u>A.2</u>	NM-SCORE.....	97
B	PERSPECTIVES	100
<u>B.1</u>	ANALYSE DE LA MFM PAR LES MODELES DE LA THEORIE DE REPONSE AUX ITEMS	100
<u>B.1.1</u>	EVALUATION DE LA VALIDITE DES HYPOTHESES DES MODELES DE REPONSES AUX ITEMS	100
<u>B.1.2</u>	AJUSTEMENT AUX DONNEES D'UN MODELE DE REPONSES AUX ITEMS	101
<u>B.1.3</u>	DIFFERENCES ENTRE GROUPES DANS LE FONCTIONNEMENT DES ITEMS ET DE L'ECHELLE.....	103
<u>B.2</u>	LA VALIDATION TRANSCULTURELLE DE LA CLASSIFICATION NM-SCORE	103
<u>B.2.1</u>	L'EQUIVALENCE.....	104
<u>B.2.2</u>	APPROCHES GENERALES D'ADAPTATION TRANSCULTURELLE	104
<u>B.2.3</u>	METHODOLOGIE PROPOSEE POUR L'ADAPTATION ET LA VALIDATION TRANSCULTURELLE	105

CONCLUSIONS..... 107

REFERENCES..... 109

ANNEXES..... 121

Annexe 1

Article 1 : Self perception of quality of life by adolescents with neuromuscular diseases

Annexe 2

Article 2 : Responsiveness of the Motor Function Measure in neuromuscular diseases

Annexe 3

Article 3 : Monitoring changes and predicting loss of ambulation in Duchenne Muscle Dystrophy with the Motor Function Measure

Annexe 4

Article 4 : Use of the Motor Function Measure to Assess Motor Function in patients with Spinal Muscular Atrophy

Annexe 5

Article 5 : Development and reliability of NM-Score: a system to classify motor function in patients with neuromuscular diseases

Annexe 6

Classification NM-Score

FIGURES

FIGURE 1 : L'unité motrice	17
FIGURE 2 : Roue de Deming. Place de l'évaluation dans la réalisation des programmes d'EPP : Observer, comparer, améliorer, mesurer, pérenniser... ..	23
FIGURE 3 : Dessin issu de l'article de D Dupage, 2008. Qualité et santé. Première partie : qualité des moyens ou qualité des résultats ?.....	24
FIGURE 4 : Courbe d'évolution de la Mesure de Fonction Motrice chez un patient atteint d'une dystrophie musculaire de Duchenne.	27
FIGURE 5 : Modélisation de la mesure.....	29
FIGURE 6 : Classification internationale des déficiences, incapacités et handicap (CIDIH ou CIH1).....	31
FIGURE 7 : Classification internationale du fonctionnement du handicap et de la santé (CIFou CIH2).....	31
FIGURE 8 : Passage d'une échelle quantitative (MFM) à une description en grades de sévérité d'atteinte fonctionnelle motrice (NM-Score).....	76
FIGURE 9 : Lien entre mesure et construit du NM-Score.....	76
FIGURE 10 : Répartition des patients en fonctions des scores MFM dans chaque groupe de sévérité (atteinte légère, moyenne, sévère et très sévère)	77
FIGURE 11 : Score moyen MFM (+ déviation standard) pour chaque groupe de sévérité (atteinte légère, moyenne, sévère et très sévère)	78
FIGURE 12 : Représentation graphique de la méthodologie utilisée pour l'analyse de la validité des construits du NM-Score	84
FIGURE 13 : Box plots des scores MFM (D1, D2 et D3) en fonction des NM-Score D1, D2 et D3	90

TABLEAUX

TABLEAU 1 : Les différents concepts à mesurer en fonction de la CIF	33
TABLEAU 2 : Critères d'évaluation des propriétés métrologiques d'outils de mesure en santé selon Andersen	45
TABLEAU 3 : Les différents outils d'évaluation de la fonction motrice selon la CIF	49
TABLEAU 4 : Score de Brooke	50
TABLEAU 5 : Score de Vignos.....	51
TABLEAU 6 : 6-minute walk test	52
TABLEAU 7 : Abilhand for neuromuscular diseases.....	53
TABLEAU 8 : Activlim.....	54
TABLEAU 9 : CHOP INTEND.....	55
TABLEAU 10 : EgenKlassifikation scale.....	56
TABLEAU 11 : ExpandedHammersmithfunctionalmotorscale	57
TABLEAU 12 : Functional independence measure for children	58
TABLEAU 13 : Gross motor function measure.....	59

TABLEAU 14 : Hammersmithfunctionalmotorscale	60
TABLEAU 15 : Hammersmith motor ability scale	61
TABLEAU 16 : mesure de fonction motrice-32	62
TABLEAU 17 : Mesure de fonction motrice-20	63
TABLEAU 18 : Modified Hammersmith functional motor scale.....	64
TABLEAU 19 : Muscular dystrophy functional rating scale	65
TABLEAU 20 : North star ambulatory assessment	66
TABLEAU 21 : Test of infant motor performance.....	67
TABLEAU 22 : UpperLimb module	68
TABLEAU 23 : Tableau synoptique de l'étude NM-Score	78
TABLEAU 24 : Phase de développement du NM-Score : Résultats de l'étude expérimentale pour le NM-Score D1	87
TABLEAU 25 : Phase de validation contre critère : Distribution des 158 patients dans chaque grade de l'échelle NM-Score pour les 3 domaines D1, D2 et D3.....	88
TABLEAU 26 : Phase de validation contre critère : Distribution du score de Vignos par grades NM-Score D1	89
TABLEAU 27 : Phase de validation contre critère : Distribution du score de Brooke par grades NM score D2 et D3	89
TABLEAU 28 : Phase de validation contre critère : Coefficients de Spearman	91
TABLEAU 29 : Facteurs d'échelle pour les indicateurs de construits D1, D2 et D3.....	91
TABLEAU 30 : Tableau croisé des NM-Score D1/D2, D1/D3 et D2/D3	93

ABREVIATIONS

ADL	Activity daily living (activités de la vie quotidienne)
AFC	Analyse factorielle confirmatoire
AIMS	Alberta infant motor scale
ALFR	Amyotrophic lateral sclerosis functional rating scale
ASI	Amyotrophie spinale infantile
CIDIH	Classification internationale des déficiences incapacités et handicap
CIF	Classification internationale du fonctionnement
CIH	Classification internationale des handicaps
CIM-10	Classification internationale des maladies, 10 ^{ème} révision
DFSH	Dystrophie fascio scapulo humérale
DIF	Differential item functioning
DMC	Dystrophie musculaire congénitale
DMD	Dystrophie musculaire de Duchenne
DMS	Dystrophie myotonique de Steinert
EBM	Evidence Based medicine (Médecine fondée sur les preuves)
EK-Scale	Egen Klassifikation scale
EMG	Electromyogramme
EPP	Evaluation des pratiques professionnelles
WeeFIM	Functional independence measure for children (Mesure d'indépendance fonctionnelle)
GMFM	Gross motor function measure (Evaluation motrice fonctionnelle globale)
HFMS	Hammersmith functional motor scale
HFMSSE	Expanded Hammersmith functional motor scale
HMAS	Hammersmith motor ability scale
ICC	Intra class correlation coefficient (coefficient de corrélation intra classe)
INQoL	Individualized neuromuscular quality of life questionnaire
JHFT	Jebsen hand function test
MAUULF	Melbourne assessment of unilateral upper limb function
MFM-32	Mesure de la fonction motrice
MFM-20	Mesure de la fonction motrice pour les moins de 7ans
MNM	Maladie neuromusculaire
MoHFMS	Modified Hammersmith functional motor scale (MoHFMS)
Mov ABC	Movement assessment battery for children
MDFRS	Muscular dystrophy functional rating scale
NP	Neuropathies périphériques
NSAA	North star ambulatory assessment
OMS	Organisation mondiale de la santé
PQLI	Physical quality of life index

PSI	Person Separation Index
QDV	Qualité de vie
SF 36	Short Form 36 of the outcomes medical questionnaire
SIP	Sickness Impact Profil
TCT	Théorie classique des tests (classical Test theory)
TRI	Théorie des réponses aux items (item response theory)
TIMP	Test of infant motor performance
UL-Module	Upper limb module
VSP-A	Vécu Santé perçu par les adolescents

INTRODUCTION

Le terme de maladies neuromusculaires (MNM) regroupe les affections touchant les muscles et le système nerveux. Dans ces affections, l'unité motrice (constituée par le muscle, la jonction nerf/muscle et le nerf, Figure 1) fonctionne mal et le muscle ne peut pas se contracter normalement. Il existe plus de 200 maladies différentes avec des spécificités (symptômes, âge de survenue...), correspondant à une atteinte différente de l'unité motrice.

Figure 1 : L'unité motrice

La plupart des MNM sont d'origine génétique. Il existe, cependant, d'autres causes possibles (dysfonctionnement endocrinien, perturbation du système immunitaire, mécanisme de neuro-dégénérescence...).

Les MNM dites de l'enfance débutent la plupart du temps chez l'enfant ou l'adolescent ; les plus communes sont la dystrophie musculaire de Duchenne (DMD), l'amyotrophie spinale infantile (ASI), les neuropathies périphériques ainsi que les myopathies et dystrophies musculaires congénitales (DMC). Sont exclues des MNM de l'enfance toutes les MNM apparaissant exclusivement à l'âge adulte comme la sclérose latérale amyotrophique par exemple.

Le seul signe commun à toutes les MNM est le déficit de la force musculaire, quelles que soient la sévérité et la localisation de l'atteinte. Les symptômes, l'âge de survenue et la sévérité de l'atteinte diffèrent selon les maladies. Aussi les fonctions motrices, respiratoires, cardiaques et digestives peuvent être diversement perturbées. La fin du 19^e siècle, marquée par l'essor de la neurologie, a vu éclore de nombreuses descriptions de pathologies entraînant des déficits moteurs. A l'origine, toutes ces maladies semblent essentiellement restreintes aux observations rares et isolées des découvreurs qui leur laissent leurs noms. Progressivement les descriptions cliniques se sont précisées et enrichies à partir de l'observation attentive des patients (de l'âge de début de la maladie, de la localisation et de l'évolution de la faiblesse musculaire), et à partir des résultats de nouvelles techniques (EMG, microscopie électronique). L'apparition de nouvelles techniques de mesure mises au point (EMG, myométrie, spirométrie) et de nouveaux champs d'étude, comme la génétique et la biologie moléculaire ont ainsi permis de parfaire la classification des MNM et de commencer à décrire l'évolution de ces maladies en fonction de différentes mesures comme la force musculaire, la fonction motrice ou la capacité respiratoire

[4][81]... Il s'agit des premières études d'histoire naturelle des maladies permettant d'obtenir des descriptions précises de l'évolution des maladies en l'absence de traitement.

Le profil évolutif naturel de ces maladies débutant dans l'enfance ou l'adolescence consiste en une réduction, progressive, de la force musculaire aboutissant à une dégradation des capacités fonctionnelles des patients, ainsi qu'une réduction progressive, qualitative et quantitative, de l'espace de vie dans ses dimensions individuelles, familiales et sociales. Même si quantité (espérance de vie) et qualité de vie de ces patients se sont améliorées ces dernières décennies, elles demeurent des maladies incurables nécessitant un suivi et une prise en charge pluridisciplinaire prolongés en médecine physique et réadaptation (MPR). L'objectif de la MPR est de restaurer et compenser les incapacités, et de favoriser et encourager les aptitudes personnelles et sociales des malades tout en remplaçant l'individu dans son environnement et en prenant en compte ses particularités [6]. Pour cela, il faut pouvoir analyser les interactions du patient avec son environnement et quantifier ses incapacités et aptitudes par la mesure d'indicateurs pertinents.

Dans le champ des MNM, comme ailleurs en médecine, une mesure est la plupart du temps effectuée pour aider à la décision pour le diagnostic et les indications thérapeutique. Par exemple, si le médecin mesure la fonction motrice de son patient c'est pour décider ou non de lui prescrire une aide technique en cas d'aggravation. Or si le résultat de mesure est erroné, il y aura une forte probabilité pour que la décision le soit aussi. Si finalement, la mesure de fonction motrice est erronée et que le patient progresse, la prescription d'aide technique n'est pas adaptée et inutile. Un résultat erroné peut également avoir des conséquences fâcheuses dans la vie de tous les jours du patient. En effet, le patient et sa famille sont maintenant très inquiets face à l'évolution annoncée de la maladie par leur médecin.

Il est donc primordial d'avoir des résultats de mesure fiables car toute erreur en médecine peut avoir des conséquences importantes. Intervient alors la métrologie : la « Science des mesures et ses applications » [59]. Elle « comprend tous les aspects théoriques et pratiques des mesurages, quels que soient l'incertitude de mesure et le domaine d'application » [59].

Un résultat de mesure fiable correspond au besoin en matière de maîtrise des risques liés aux erreurs de mesure et à leurs conséquences (négatives bien sûr). Un résultat de mesure fiable s'obtient par la maîtrise du processus de mesure, c'est-à-dire la maîtrise des facteurs ayant une influence sur la qualité des résultats délivrés. Ces facteurs peuvent être la nature de l'outil utilisé pour effectuer la mesure, la personne réalisant la mesure, la méthode utilisée pour effectuer la mesure, le type d'échantillon sur lequel est effectuée la mesure, etc...

En recherche clinique tout particulièrement, une métrologie rigoureuse est indispensable car l'on ne peut, en l'occurrence, ni se passer de quantification, ni se contenter d'une quantification approximative [107]. L'apparition des thérapies géniques et les grands espoirs qu'elles suscitent chez les patients et leur famille a été un argument supplémentaire pour encourager l'évaluation dans ce domaine. Nous citerons Bushby et Beckmann, à propos de la nécessité urgente de disposer de données cliniques fiables issues de groupes génétiquement homogènes, « We have now entered the world of the « après gène ». We need to have tools to seek answers to new, more focused questions on pathophysiological and clinical aspects » [27].

Depuis quelques années maintenant, les différentes équipes internationales impliquées dans le domaine des MNM ont pris conscience de l'importance du choix avisé des critères de jugement (ou « outcomes measures »). Des réseaux se sont constitués, soit pour l'ensemble des maladies (par exemple : le réseau « Cornemus » qui fédère les structures et équipes françaises impliquées dans les MNM) soit spécifiques à une maladie (par exemple : « The International Coordinating Committee » ou ICC dans le cadre de l'amyotrophie spinale infantile (ASI), « Cure-CMD » et le réseau européen « Treat-CMD » dans le cadre des DMC. Récemment, l'intérêt de ces réseaux pour les outils d'évaluation a fortement augmenté. Cet aspect a été ainsi le centre d'intérêt de plusieurs réunions de travail internationales [19][23][28][99]. A l'issue de chacune de ces réunions, tout le monde s'accorde sur la nécessité de développer des outils d'évaluation valides dans différentes pathologies afin de pouvoir mesurer de façon sensible et précise les effets des traitements dans le cadre d'essais thérapeutiques sans pour autant aboutir à un consensus sur l'outil à retenir pour chaque pathologie.

Nous décrivons d'abord toute la problématique de la mesure et du choix des meilleurs outils dans le domaine des MNM de l'enfance. Après une revue de la littérature concluant aux limites des outils actuels, nous proposons de compléter l'étude de validation de la Mesure de Fonction Motrice (MFM), échelle développée à partir de 1998. Enfin, nous inspirant de la méthodologie décrite dans le premier chapitre, nous décrivons toutes les étapes aboutissant à la création d'une classification en grades de sévérité d'atteinte fonctionnelle motrice : le NM-Score.

Ce travail vise à :

- proposer une méthodologie d'analyse des propriétés métrologiques des outils de mesure applicable quel que soit le domaine de l'évaluation en médecine ;
- fournir un état des connaissances sur les caractéristiques des outils d'évaluation de la fonction motrice des patients porteurs de MNM ;
- fournir à la communauté scientifique internationale deux outils d'évaluation de la fonction motrice avec des qualités métrologiques satisfaisantes autorisant leur utilisation en pratique clinique et en recherche.

CHAPITRE 1 : CONTEXTE ET DEMARCHES DE LA RECHERCHE

A L'EVALUATION EN MEDECINE

Les progrès de la médecine et de la technologie représentent une contribution indéniable pour la vie humaine. Ils permettent de réduire le taux de mortalité et de guérir aujourd'hui ce qui, hier encore, tuait, et de mieux soulager la douleur physique. Cette notion de progrès médical, surtout omniprésente pendant les trente glorieuses, a subi un recul concomitant de l'apparition de la qualité dans le champ sanitaire. Ce progrès médical était d'une telle évidence depuis la dernière guerre mondiale grâce à l'avènement des antibiotiques que la notion de qualité n'avait que peu de sens et l'augmentation du coût de la santé paraissait justifiée par l'importance des progrès réalisés. Malheureusement en France, actuellement, le coût de la santé croît plus vite que le PIB et le niveau de vie des ménages [41]. Ceci a conduit les tutelles à s'interroger sur la qualité des soins et en particulier à évaluer le rapport coût/efficacité des soins. Les concepts de qualité et d'évaluation se sont donc imposés dans l'organisation et la gestion de la santé depuis une vingtaine d'années.

A.1 JUSTIFICATIONS DE L'EVALUATION EN MEDECINE

Evaluer c'est apprécier objectivement l'efficacité de telle ou telle stratégie pour permettre un meilleur choix pour le malade. C'est mettre en place les procédures permettant d'en vérifier la bonne réalisation, détecter et quantifier les écarts, imaginer et fournir les éléments pour mettre en place les éléments de corrections [92].

L'évaluation se justifie de façon prioritaire d'un point de vue éthique car, face au développement exponentiel de nouveaux traitements et de nouvelles technologies, nous nous devons d'évaluer leur bien-fondé (bénéfiance) pour le patient et leur absence de nocivité (maléfiance). L'évaluation s'est aussi imposée parallèlement à l'évolution de la relation paternaliste du médecin vis à vis de son patient vers la décision médicale partagée. Le médecin n'est plus tout puissant dans la décision de soin. Le médecin doit considérer la volonté du patient (ou de sa famille dans le cas des enfants) qui demande à être informé et à participer activement aux décisions médicales le concernant.

Le paradigme de l'« Evidence Based Medicine » (EBM) ou la médecine fondée sur les preuves, illustre bien cette évolution de la médecine et l'importance de l'évaluation. Selon ses fondateurs [117] la médecine fondée sur les faits conjugue :

- **L'expertise du clinicien et le patient** : « Par expertise clinique individuelle, nous entendons la compétence et le jugement que chaque clinicien acquiert par l'expérience et la pratique clinique. L'expertise se manifeste de plusieurs manières, mais surtout par des diagnostics efficaces et efficaces ainsi que par un discernement approfondi et de la compassion dans son attitude face au patient, l'épreuve qu'il ou elle vit, ses droits, et ses préférences ».
- **Les meilleures données cliniques externes.** « Par meilleures données cliniques externes, nous entendons les recherches pertinentes sur un plan clinique, souvent issues de la recherche médicale fondamentale, mais surtout des recherches cliniques sur les tests diagnostiques centrés sur le patient (y compris les examens cliniques) les plus exacts et précis, sur la puissance des marqueurs pronostiques, et enfin sur l'efficacité et l'innocuité des schémas thérapeutiques, de réadaptation et de prévention ».

Le processus de l'EBM passe tout naturellement par les étapes suivantes :

1. la formulation d'une question clinique claire et précise à partir d'un problème clinique posé ;
2. la recherche dans la littérature d'articles cliniques pertinents et appropriés sur le problème ;
3. l'évaluation critique de la validité et de l'utilité des résultats trouvés (« niveau de preuve ») ;
4. la mise en application des résultats de l'évaluation dans la pratique clinique pour une prise en charge personnalisée de chaque patient.

«L'évaluation des pratiques professionnelles (EPP) mentionnée à l'article L.4133-1-1 a pour but l'amélioration continue de la qualité des soins et du service rendu aux patients par les professionnels de santé. Elle vise à promouvoir la qualité, la sécurité, l'efficacité et l'efficience des soins et de la prévention et plus généralement la santé publique, dans le respect des règles déontologiques. Elle consiste en l'analyse de la pratique professionnelle en référence à des recommandations et selon une méthode élaborée ou validée par la Haute Autorité de santé et inclut la mise en œuvre et le suivi d'actions d'amélioration des pratiques.»

Décret du 14 avril 2005 article L. 41-33

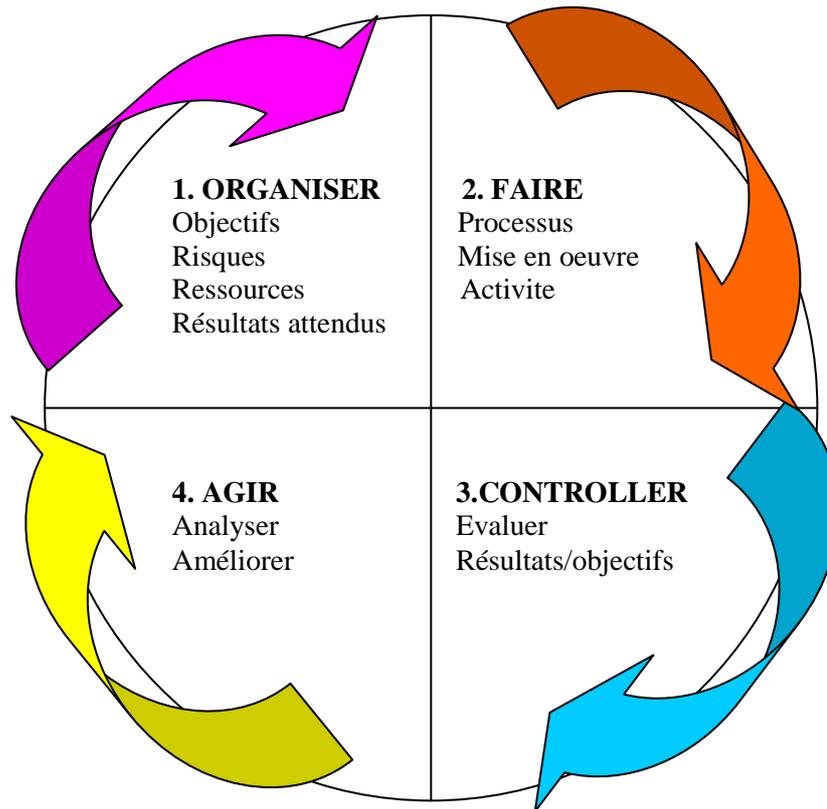


Figure 2 : Roue de Deming. Place de l'évaluation dans la réalisation des programmes d'EPP : Observer, comparer, améliorer, mesurer, pérenniser...

L'EPP doit être une évaluation formative et intégrée. C'est en intégrant la formation et l'évaluation à l'exercice quotidien (Figure 2), et en développant des modes d'exercices ou formation et évaluation font partie du travail quotidien que cette évolution majeure sera conduite avec succès.

Les accréditations, les recommandations thérapeutiques ou les évaluations des pratiques professionnelles font intervenir des normes, des référentiels et une logistique qui a pour objectif l'amélioration de la qualité des soins. Dans le champ médical, ces références pour juger (norme, situation antérieure) sont parfois difficiles à trouver et reposent sur des conférences de consensus, des avis d'experts ou des recommandations de pratique clinique. Cette évolution, si elle tend à être normative à l'extrême, pourrait avoir des effets délétères qui vont à l'encontre de la qualité effective des soins. L'évaluation imparfaite de la qualité des soins, par référence obligatoire à la norme, aboutit à un nivellement par le bas du soin, car la norme est finalement le plus petit dénominateur commun de la qualité [41].

A.2 DIFFERENCIATION DE L'ÉVALUATION DES MOYENS DES PROCEDURES ET DES RESULTATS

Que la stratégie soit belle est une chose, mais n'oubliez pas de regarder le résultat.

Winston Churchill

L'évaluation nécessite au préalable une définition de l'objectif poursuivi. Le concept d'évaluation dans le domaine de la santé a beaucoup évolué ; initialement centré sur le résultat, il concerne désormais quasi-exclusivement les procédures de soin.

Si l'évaluation de la procédure est privilégiée, le risque est grand de voir l'amélioration de la santé des patients passer au second plan [41] (Figure 3).

Figure 3 : Dessin issu de l'article de D Dupage, 2008. Qualité et santé. Première partie : qualité des moyens ou qualité des résultats ?

La place croissante de la prise en compte des procédures dans l'évaluation de la qualité a conduit à la segmentation en qualité interne et qualité externe :

- La qualité externe correspond à l'amélioration du sort des patients, individuellement ou en tant que groupe. C'est la qualité fondée sur des critères extérieurs au soin lui-même. Il s'agit de fournir des services conformes aux attentes des patients. Ce type de démarche passe ainsi par une nécessaire écoute des patients, mais doit permettre également de prendre en compte des besoins implicites, non exprimés par les bénéficiaires.

- La qualité interne correspond à l'amélioration du fonctionnement interne de la structure de soins. L'objet de la qualité interne est de mettre en oeuvre des moyens permettant de décrire au mieux l'organisation, de repérer et de limiter les dysfonctionnements et de soigner au meilleur coût. Les bénéficiaires de la qualité interne sont la direction et les personnels de la structure, ainsi que ses financeurs. Le destinataire final du soin, le patient, est le plus souvent exclu de la démarche.

La médecine étant un art autant qu'une science, supprimer l'évaluation de la qualité externe au prétexte de sa subjectivité est une erreur. Cette erreur est issue d'une démarche scientifique ayant dépassée son objectif. Exclu des processus décisionnels, le patient est également exclu de la mesure de la qualité.

« La qualité des soins ne peut donc progresser à partir de la seule évaluation de la qualité interne. Cette évaluation aboutit à une baisse de la qualité globale et de la diversité des approches. La qualité interne exclusive réduit l'humain à un organisme standardisé en s'intéressant à des populations et nie la singularité de l'individu. Si la procédure est bonne pour la population, elle est censée être applicable à chaque patient » [41].

A.3 LE PRINCIPAL ECHAPPE A L'EVALUATION

La subjectivité a été progressivement bannie de l'évaluation scientifique. Or, la médecine est une science de l'Homme qui est fondamentalement un sujet et non un objet.

« L'objectivisation de la médecine a bien sûr permis de réels progrès en mesurant précisément l'impact de nos interventions sur des phénomènes bien définis. Mais elle a aussi été terriblement sclérosante en excluant du champ de la science médicale l'essentiel de ce qui caractérise l'Homme, à savoir les émotions, les affects, la complexité, ou encore la versatilité et l'oubli. La science expérimentale a abouti à une "objectivation du sujet", à un réductionnisme cartésien de l'individu réduit à une somme d'organes, de mécanismes biologiques où à un "sujet-type" dont le destin sera applicable à ses semblables » [41]. Or, tout n'est pas mesurable tout n'est pas évaluable et objectivable et encore moins transposable d'un sujet à l'autre...

Certains proposent une rupture de paradigme et une relativisation de la place de l'objectivité. Edgar Morin [103] est sévère avec ce réductionnisme. Pour lui, le dogme réductionniste mène à une « intelligence parcellaire, compartimentée, mécaniste, disjonctive, qui brise le complexe du monde en fragments disjoints, fractionne les problèmes, sépare ce qui est relié, unidimensionnalise le multidimensionnel. Qu'il s'agit là, d'une intelligence à la fois myope, presbyte, daltonienne, borgne ; qui finit le plus souvent par être aveugle. Elle détruit dans l'œuf toutes les possibilités de compréhension et de réflexion, éliminant aussi toutes chances d'un jugement correctif ou d'une vue à long terme. Ainsi, plus les problèmes deviennent multidimensionnels, plus il y a incapacité à penser leur multi dimensionnalité ; plus progresse la crise, plus progresse l'incapacité à penser la crise ; plus les problèmes deviennent planétaires, plus ils deviennent impensés. Incapable d'envisager le contexte et le complexe planétaire, l'intelligence réductionniste aveugle rend inconscient et irresponsable. »

B L'ÉVALUATION DANS LE DOMAINE DES MALADIES NEUROMUSCULAIRES DE L'ENFANCE

B.1 L'ÉVALUATION EN MÉDECINE PHYSIQUE ET READAPTATION

Selon l'Organisation Mondiale de la Santé (OMS), « La médecine physique et de réadaptation » (MPR) peut être définie comme une spécialité qui a pour rôle de coordonner et d'assurer la mise en application de toutes les mesures visant à prévenir ou à réduire au minimum inévitable les conséquences fonctionnelles, physiques, psychiques, sociales et économiques des déficiences ou des incapacités. Dans le cas des maladies chroniques, qui sont souvent congénitales, la MPR comporte la mise en œuvre méthodique des actions nécessaires à la réalisation de ces objectifs, depuis le début de l'affection dans l'enfance et tout au long de la vie, en permettant une intégration optimale du patient dans son milieu ambiant et dans la société.

Plus précisément, la « rééducation » vise par ses techniques à aider l'individu à réduire ses déficiences et incapacités. La « réadaptation » est définie comme l'application coordonnée et combinée de mesures dans les domaines médical, social, psychique, technique et pédagogique, qui peuvent aider le patient à trouver la place qui lui convient le mieux dans la société et/ou à conserver sa place en s'adaptant à ses incapacités lorsqu'elles deviennent stabilisées et persistantes. Enfin, la « réinsertion » représente l'ensemble des mesures médicosociales visant à optimiser le retour dans la société en évitant les processus d'exclusion.

L'évaluation est par nature au centre de cette spécialité. Alors que certaines spécialités médicales peuvent facilement utiliser la guérison ou l'espérance de vie comme critère de jugement, le champ des MNM impose une précision et la recherche d'autres critères d'évaluation.

L'évaluation permet la compilation de données cliniques issues de groupes génétiquement homogènes [27], permettant l'étude de l'histoire naturelle de pathologies et une meilleure connaissance de ces pathologies.

L'évaluation nous permet aussi de mesurer l'effet de nos actions thérapeutiques en termes d'amélioration des activités des patients (mobilité, préhension etc..) au sens de l'Organisation Mondiale de la Santé (OMS) (Figure 7 p :29). Au niveau individuel et particulièrement dans le domaine des MNM, l'évaluation permet le suivi de l'évolution des capacités fonctionnelle motrices des patients (Figure 4). Ce suivi régulier permet, de façon objective, d'apprécier l'évolution de la maladie et sert de support à la discussion avec le patient et sa famille. De plus, l'évaluation permet de prévoir, parfois plusieurs mois à l'avance, la survenue d'étapes importantes pour le patient (par exemple, la perte de marche chez le patient atteint de DMD [139]), permettant la mise en place au préalable de toutes les adaptations nécessaires.

L'évaluation, en utilisant des outils qui se veulent standardisés, permet un langage commun favorisant la communication entre professionnels souvent issus de différents horizons (médical, paramédical, social).

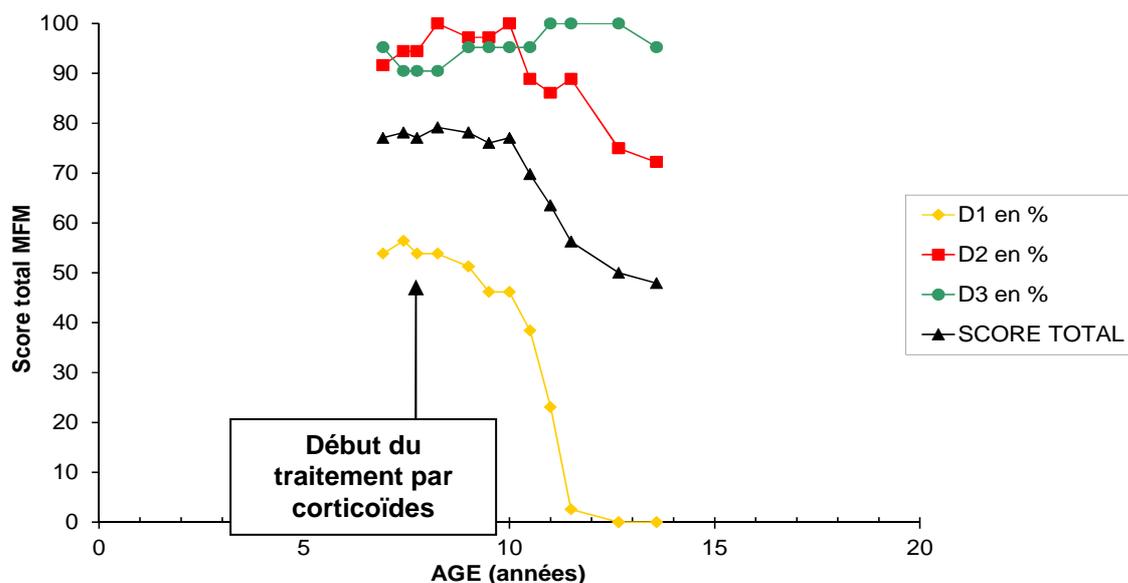


Figure 4 : Courbe d'évolution de la Mesure de Fonction Motrice chez un patient atteint d'une dystrophie musculaire de Duchenne.

La corticothérapie permet une relative stabilisation pendant les premières années des scores de la MFM (D1 : position debout et transferts, D2 : motricité axiale et proximale et D3 : motricité distale).

Source : patient suivi dans le service de l'Escale

B.2 L'ÉVALUATION EN RECHERCHE CLINIQUE

Le développement actuel de la recherche dans le champ des MNM a imposé la nécessité de précisément mesurer l'effet des traitements et d'évaluer les résultats, c'est-à-dire les conséquences de ce traitement sur le patient de manière précise et reproductible.

De nombreuses recherches sont en cours dans le champ des MNM. Il s'agit en particulier de thérapies dites innovantes :

- La thérapie génique, qui est une piste thérapeutique en plein essor. Elle consiste à remplacer la fonction d'un gène défectueux ou manquant par celle d'un gène thérapeutique inséré dans la cellule. Les premières maladies ayant bénéficié de travaux de thérapie génique sont des maladies génétiques rares, en particulier des déficits immunitaires sévères.
- La thérapie cellulaire, qui relève le défi de remplacer des cellules déficientes ou disparues par des cellules saines. Cette technique utilise l'apport de cellules particulières ; les cellules souches dont les cellules IPS (cellules adultes différenciées redevenues pluripotentes).
- La chirurgie du gène, qui consiste à réparer le gène lui-même par des techniques utilisant le saut d'exon, la translecture de codons stop ou des méga nucléases. Ces techniques qui ont moins de dix ans sont déjà à l'essai chez l'homme.
- Enfin, les médicaments qui sont une piste plus traditionnelle. De nouvelles molécules sont en effet sans cesse découvertes et les médicaments déjà existants n'ont pas fini de révéler tous leurs potentiels.

Quelle que soit la technique utilisée, l'efficacité de ces nouveaux traitements doit être prouvée. Ceci passe obligatoirement par des essais thérapeutiques, très coûteux en temps et en argent. L'évaluation de chaque nouvelle thérapie démarre par des essais précliniques, sur des cultures de cellules humaines ou animales (*in vitro*) ou sur un modèle animal (*in vivo*). Si ces essais apportent la preuve que le traitement est susceptible d'être efficace et bien toléré chez l'homme, un essai clinique, encadré en France par la loi n° 88-138 du 20 décembre 1988 (dite loi Huriet-Sérusclat) relative à la protection des personnes qui se prêtent à des recherches biomédicales, est mené auprès de sujets malades volontaires, et parfois auprès de sujets sains.

La construction d'un essai clinique suit une méthodologie rigoureuse avec un choix avisé des critères de jugement, ou « outcomes measures », qui conditionne la pertinence des résultats obtenus. En effet, la mise en évidence de l'effet d'un traitement passe par le choix raisonné mais toujours discutable des meilleurs critères possibles mesurant l'efficacité et la tolérance de ce traitement.

Les critères de jugement sont les critères d'évaluation utilisés dans l'essai thérapeutique pour mesurer l'efficacité du traitement. Leurs choix appropriés lors d'un essai clinique est crucial pour la capacité de l'essai à conclure, à partir de données fiables, à la présence ou à l'absence d'un effet thérapeutique significatif.

Dans l'idéal, on attend de ces critères qu'ils soient pertinents, c'est-à-dire : 1/ qu'ils répondent au problème posé ; 2/ qu'ils soient signifiants en termes cliniques ; 3/ qu'ils fassent l'objet d'un consensus dans la communauté médicale ; 4/ qu'ils soient disponibles chez tous les sujets ; 5/ que leurs recueils ne soit pas pénible pour les patients ; 6/ que leurs mesures soient faciles, précises, reproductibles ; 7/ qu'ils soient sensibles, c'est-à-dire aptes à déceler de petites différences. Classiquement, il est demandé aux critères de jugement d'être objectif, c'est-à-dire relatif à un objet pensé et facilement matérialisable par un modèle physiopathologique. Il faut cependant tenir compte qu'une mesure chiffrée n'est pas implicitement objective et cela même lorsque ses conditions de mesure sont parfaitement codifiées et standardisées. Les critères subjectifs sont eux relatifs au sujet pensant et tentent de décrire l'intimité psychique de nos patients [35]. Le caractère subjectif d'une mesure n'enlève en rien de sa valeur, cette mesure peut être aussi précise et scientifiquement recevable qu'une mesure objective [44][45]. Le plus important réside dans la qualité de la mesure, qu'elle soit subjective ou objective. Enfin, le critère d'évaluation devra dans l'idéal être direct, c'est-à-dire traduisant directement le phénomène médical étudié ou, à défaut, être validé en comparaison avec un gold standard dans la pathologie. D'une façon générale, il faut éviter de multiplier les critères dont le nombre est à l'origine d'une multiplicité de tests rendant l'interprétation très délicate. Le protocole doit donc toujours privilégier un critère qualifié de "principal", également utilisé pour le calcul du nombre de sujets nécessaires à la mise en évidence de la différence espérée entre les traitements. Les autres critères ne sont, dans la hiérarchie, considérés que comme des critères "accessoires", dont il n'est pas question de majorer ultérieurement l'importance, au vu des résultats.

Le choix du critère d'évaluation le plus adapté devra se faire en fonction du plan expérimental de l'étude, et plus précisément en fonction de la population incluse dans l'essai (type de maladie neuromusculaires, âge des patients et évolutivité de la maladie), de l'état fonctionnel des patients au moment du recrutement, de la durée de l'étude et enfin de l'effet possible du traitement [101].

Ces critères d'évaluation sont des mesures d'un phénomène physiologique. Le choix de ces critères lors d'un essai thérapeutique doit donc se faire en deux étapes ;

- la première étape qui est le choix du concept à mesurer (variable latente)
- la seconde étape qui est le choix de l'outil à utiliser, le plus précis et adapté pour obtenir une mesure fidèle de ce concept.

La mesure obtenue peut être modélisée de la façon suivante : **MESURE = VARIABLE LATENTE + Ecart (Variable latente)**

Cet écart dépend du mesuré, de l'outil utilisé et de la personne utilisant l'outil. Le type d'outil agit plutôt sur l'écart entre la mesure et l'état clinique alors que le juge agit plutôt sur la mesure (Figure 5).

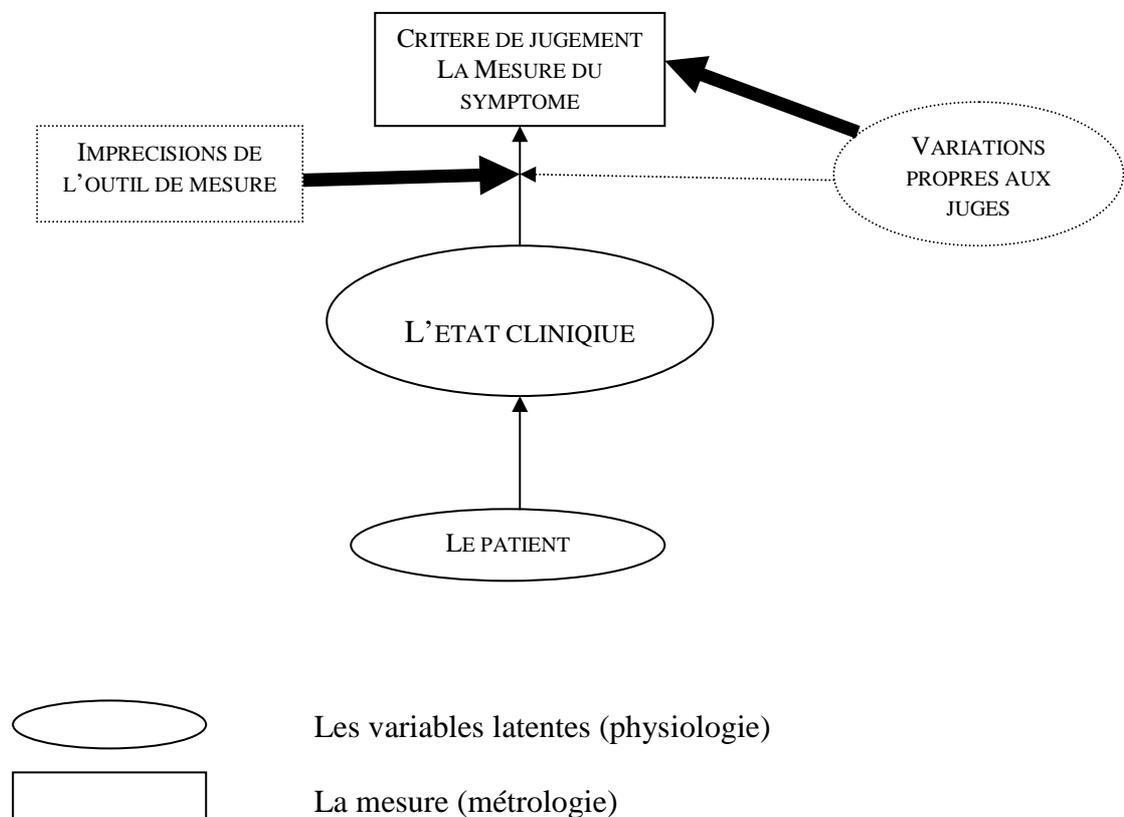


Figure 5 : Modélisation de la mesure

C DEFINIR LE CONCEPT AVANT DE TENTER DE LE MESURER

C.1 QUELS SONT LES CONCEPTS A MESURER DANS LES MALADIES NEUROMUSCULAIRES ?

La définition précise du « concept à mesurer » est un préambule indispensable avant le choix de l'outil permettant d'obtenir la mesure. Ce concept à mesurer est appelé variable latente, ou construit en métrologie. L'étape de définition du concept est indispensable car elle permettra ensuite de vérifier l'adéquation entre l'outil et le concept à mesurer.

Le symptôme commun à l'ensemble des MNM est la diminution de la force musculaire touchant différents muscles de l'organisme à des degrés variables. La cible thérapeutique dans les essais cliniques est de ce fait prioritairement la fibre musculaire. Selon les cas, peuvent aussi s'associer des douleurs, des rétractions articulaires, une diminution d'endurance, une fatigue, et parfois des atteintes viscérales en fonction du type de muscle atteint (insuffisance respiratoire, insuffisance cardiaque, troubles de la déglutition) ayant des conséquences beaucoup plus générales en terme de quantité de vie (espérance de vie) mais aussi de qualité de vie. Dans le champ moteur, les différents concepts à mesurer sont donc potentiellement très nombreux du fait de l'atteinte systémique de ces maladies et surtout de leurs conséquences dans différents domaines (Tableau 1).

La famille des classifications internationales de l'OMS est un outil précieux qui permet de décrire dans un langage uniformisé et normalisé une maladie et ses conséquences sur la santé. Les problèmes de santé (maladie, troubles, lésions et traumatismes) sont classés essentiellement en fonction de la CIM-10 (Classification Internationale des maladies, 10^{ème} révision) [73] qui fournit un cadre étiologique.

Les classifications internationales du handicap (CIH) sont au contraire des classifications des conséquences de la maladie (CIH 1 : version 1980) ou des composantes de la santé (CIH 2 : version 2001).

La première classification du handicap et de la santé établie dans les années 80 sous l'impulsion du professeur Phillip Wood [72] par l'OMS est la classification internationale des déficiences, incapacités et handicap (CIDIH) ou CIH1 (Figure 6). Ce modèle a eu le mérite de s'intéresser à de nouveaux indicateurs de santé distincts des critères de jugement d'efficacité globaux du système de soins comme la survie. Il n'était, en effet, plus possible de continuer à se baser uniquement sur des critères de jugement simples du fait de l'émergence, en partie due aux progrès thérapeutiques, des maladies chroniques produisant des séquelles d'ordre physique, psychique et sociale sur de longues périodes. Grâce à ce modèle, il était possible de décrire un degré de morbidité résiduelle (incapacité) liée à une maladie et les conséquences sociales de cette pathologie (handicap) sans relation de proportionnalité évidente entre l'importance des déficiences, les incapacités et le handicap.

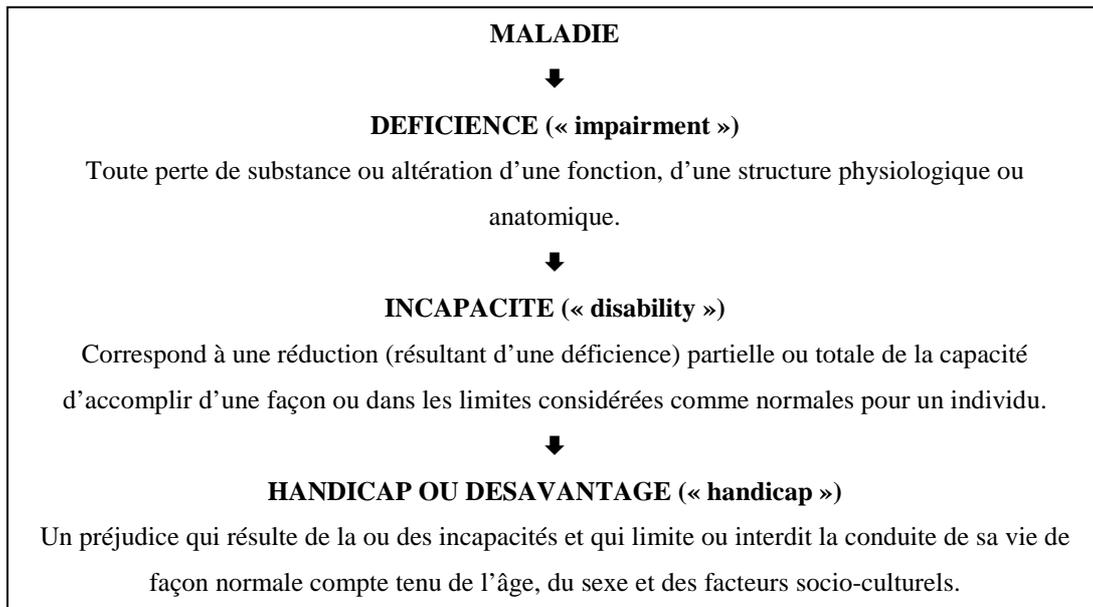


Figure 6 : Classification internationale des déficiences, incapacités et handicap (CIDIH ou CIH1)

En 2001, la classification internationale du fonctionnement du handicap et de la santé (CIF ou CIH 2) [74] (Figure 7) s'est éloignée d'une classification des « conséquences de la maladie » pour devenir une classification des « composantes de la santé ». Les « composantes de la santé » définissent ce qui constitue la santé, alors que les « conséquences » se focalisent sur l'impact de la maladie ou tout problème de santé qui peut en résulter. Ainsi la CIF adopte une position neutre par rapport à l'étiologie.

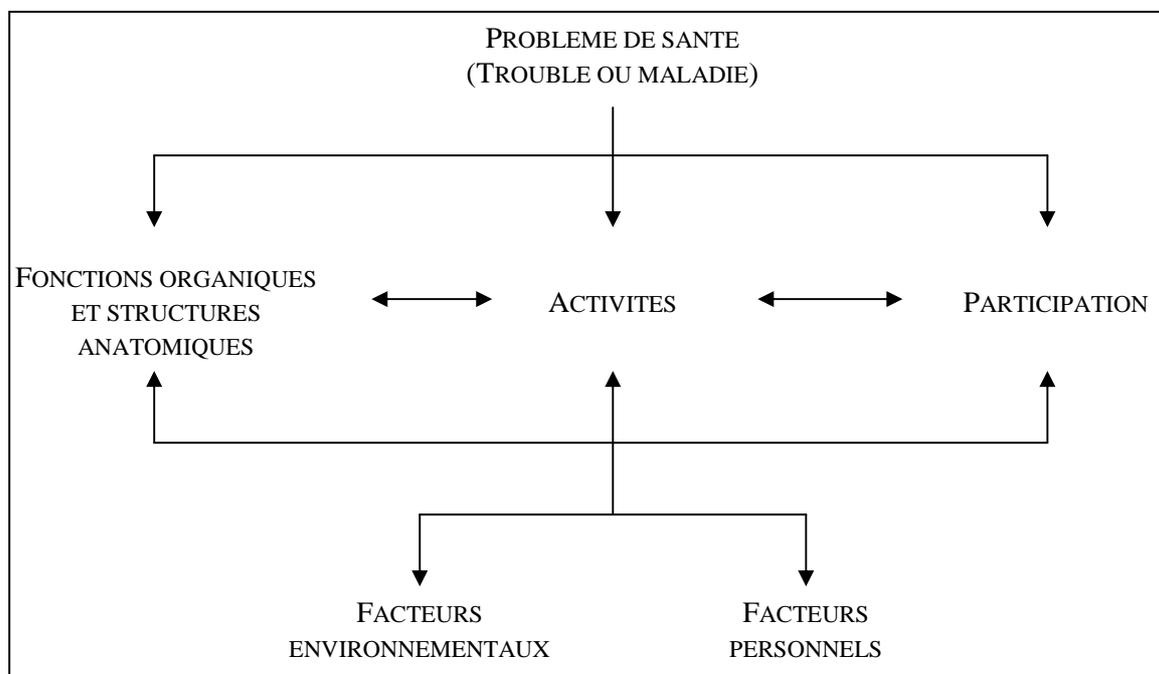


Figure 7 : Classification internationale du fonctionnement du handicap et de la santé (CIF ou CIH2)

Les innovations les plus marquantes de la CIF par rapport au modèle de Wood sont:

- la définition des termes de ce modèle sur des aspects positifs du fonctionnement des personnes, rompant définitivement avec la conception antérieure du modèle de Wood uniquement tourné sur le versant négatif ;
- la présence de deux catégories de déficiences : structures anatomiques et fonctions organiques ;
- la pluralité des interférences dans le schéma contrairement à la linéarité du modèle de Wood ;
- la prise en compte de l'interaction des personnes avec leur environnement avec une liste de facteurs environnementaux intégrée à la classification ;
- d'avoir proposé un modèle spécifique adapté à l'enfant et l'adolescent : CIF EA contenant des informations détaillées et complémentaires tenant compte des spécificités du modèle dans cette population [75].

La CIF permet de décrire des situations relatives au fonctionnement humain et aux restrictions qu'il peut subir ; elle fournit un cadre pour organiser cette information. La CIF ne concerne pas que les personnes handicapées, elle est d'application universelle [20]. Elle décrit les domaines de la santé et les domaines connexes de la santé en prenant comme perspective l'organisme (fonctions organiques et structures anatomiques), la personne en tant qu'individu (activité), et la personne en tant qu'être social (participation) en interaction avec son environnement. Alors que le « fonctionnement » se rapporte aux fonctions organiques, aux activités de la personne et à sa participation au sein de la société, le « handicap », au contraire, sert de terme générique pour désigner les déficiences, les limitations d'activités ou les restrictions de participation. Les domaines décrivant la composante « Activité et Participation » font l'objet d'une liste unique qui couvre toute la gamme des domaines de la vie, de l'apprentissage élémentaire, aux domaines plus complexes comme les relations avec autrui ou l'occupation d'un emploi. La CIF a aussi introduit les notions de capacité et performance en donnant aux facteurs environnementaux et personnels une dimension déterminante. En effet la capacité traduit, le niveau de fonctionnement le plus élevé possible qu'une personne est susceptible d'atteindre à un moment donné.

L'évaluation de la capacité doit être ajustée pour neutraliser les influences variables des différents environnements sur chaque personne, ce qui requiert un environnement normalisé (uniforme ou standardisé). L'évaluation de la performance nécessite au contraire d'être appréciée dans le cadre habituel d'évolution, ce qui inclue les différentes variables environnementales réellement rencontrées. L'écart entre capacité et performance reflète la différence d'impacts entre environnement standardisé (situation de test) et environnement usuel (situation écologique). Elle constitue ainsi un guide utile pour déterminer ce qui peut être modifié dans le cadre de vie de la personne concernée pour améliorer son niveau de réalisation. Les conditions de mesure permettent de différencier capacités et performances.

Les différents concepts à mesurer dans les MNM peuvent être représentés en utilisant la classification en domaines et en sous domaines proposées par la CIF (Tableau 1).

Tableau 1 : Les différents concepts à mesurer en fonction de la CIF

FONCTIONS ORGANIQUES	ACTIVITES		PARTICIPATION
	CAPACITE	PERFORMANCE	
<p>b2 : Fonctions sensorielles et douleurs</p> <p>-Douleur -Troubles sensitifs</p>	<p>d4: Mobilité</p> <p>Marcher et se déplacer -Marcher 10m -Monter 4 marches d'escalier -Se tenir debout -Courir dans un couloir</p> <p>Changer et maintenir la position du corps -Se transférer en position assise</p> <p>Porter déplacer et manipuler des objets -Soulever -Attraper</p>	<p>d4: Mobilité</p> <p>Marcher et se déplacer -Se déplacer dans le quartier, dans un centre commercial -Se déplacer dans la maison</p> <p>Changer et maintenir la position du corps -Se tourner dans son lit -Se lever devant son lavabo</p> <p>Porter déplacer et manipuler des objets -Porter les courses -Porter ses enfants</p>	<p>d6: Vie domestique</p> <p>-Réalisation des taches ménagères</p>
<p>b7 : Fonctions de l'appareil locomoteur et liées au mouvement</p> <p>-Force musculaire -Mobilité articulaire -Trophicité musculaire -Tonus musculaire</p>	<p>d5: Entretien personnel</p> <p>-se laver -s'habiller</p>	<p>d5: Entretien personnel</p> <p>-se laver au domicile -s'habiller au domicile</p>	<p>d8: Grands domaines de la vie</p> <p>-Education scolaire -Accès à l'emploi</p>
<p>b4: Fonctions des systèmes cardio-vasculaire, hématopoïétique, immunitaire et respiratoire</p> <p>-Fonction respiratoire -Fonction cardiaque</p>			<p>d9: Vie communautaire, sociale et civique</p> <p>-Récréation et loisirs -Religion</p>

C.2 DEFINITION DE LA PLACE DE L'EVALUATION FONCTIONNELLE MOTRICE DANS LES MALADIES NEUROMUSCULAIRES

Dans la littérature, deux types de mesure sont principalement retrouvées pour évaluer l'effet d'une thérapeutique dans le champ des MNM :

- les mesures de la déficience ou d'altération des structures anatomiques et fonctions organiques (b7 : l'appareil locomoteur et lié au mouvement selon la CIF), comme la mesure de force musculaire ;
- les mesures dites fonctionnelles ou mesure des limitations d'activités (d4 : mobilité selon la CIF) et restriction de participation (d6 : Vie domestique selon la CIF), comme la réalisation des activités domestiques dans la vie quotidienne découlant des déficiences.

Ainsi les mesures de la force musculaire (testing manuel et testing quantifié), si elles permettent une évaluation de la déficience motrice [24][120], elles sont classiquement décrites comme difficile à mesurer en cas de faiblesse musculaire importante. Pour le testing manuel, la coopération est parfois difficile à obtenir surtout chez l'enfant de moins de 6 ans [25] et il n'est parfois pas assez sensible à certains niveaux évolutifs ou dans certaines pathologies, comme ASI où une stabilité est observée lorsque les possibilités fonctionnelles se dégradent [69]. Quant au testing quantifié, qui utilise un appareillage parfois coûteux, non disponible dans tous les centres, ses limites sont souvent une reproductibilité inter observateur insuffisante, limitant souvent son utilisation à des études mono centriques utilisant le même matériel et le même utilisateur [110]. De plus, il n'y a pas toujours de corrélation franche entre la lésion organique (diminution de la force musculaire, diminution de la capacité vitale) et les conséquences fonctionnelles (diminution des capacités fonctionnelles) [118]. Or, il est crucial de considérer, en particulier dans les essais thérapeutiques, les altérations des fonctions organiques comme la force (aspects mécaniques) avec leur pertinence fonctionnelle, c'est-à-dire leur impact physiologique sur le mouvement et le retentissement de la faiblesse musculaire dans la vie quotidienne [100][118].

Dans les MNM, l'évaluation fonctionnelle motrice, qui s'impose donc comme une mesure du retentissement de la maladie neuromusculaire sur les limitations fonctionnelles et les restrictions de participation d'un patient (cette évaluation étant bien acceptée par le patient et les thérapeutes [26][69][70][100][118]), répond bien à l'ambition scientifique de la CIF et à la nécessité de mesures qualitatives.

Selon l'objectif poursuivi (recherche, évaluation clinique, objectifs thérapeutiques, adéquation au projet de vie, évaluation d'efficience, efficacité de moyens de compensation...), le choix du domaine à mesurer est déterminant pour décider de l'outil pertinent. En effet, plusieurs domaines peuvent être isolés au sein de l'évaluation fonctionnelle. D'abord en fonction de la structure anatomique visée : membres supérieurs, membres inférieurs, tronc ou face. Mais aussi, en fonction des conditions de l'évaluation, on différenciera les capacités motrices en situation de test, des performances motrices du patient dans son environnement.

Le terme de fonction motrice n'avait pas une place claire dans la première classification du handicap (modèle de Wood) [72] et se trouvait au sein des « disabilities » qui comprenaient les « functional limitations » proches des déficiences et les « activities restriction » proche des handicaps [61].

Dans la CIF, au contraire, le terme de fonction motrice est clairement identifié dans le chapitre 7 : « Fonctions de l'appareil locomoteur et liées au mouvement ». En effet, les fonctions organiques désignent les fonctions physiologiques des systèmes organiques. L'utilisation en parallèle du classement des fonctions organiques et des structures anatomiques permet d'associer la fonction de base (comme la fonction relative à la puissance musculaire) à son corrélât structurel (la force musculaire). Cette identification conceptuelle n'a pas simplifié la position de l'évaluation fonctionnelle, qui reste à mon sens plutôt dans le domaine des activités et participation, car il s'agit d'évaluer ce qu'une personne peut faire ou l'exécution d'une tâche et non la fonction d'un organe.

C.3 ET LA QUALITE DE VIE

Dans la CIF, la mesure des composantes de santé peut être définie comme externe ou extérieure, c'est-à-dire ne reposant pas sur l'avis exprimé des patients. Il s'agit d'une mesure dite objective. Une mesure objective est relative à un objet pensé (comme la tension artérielle, la température), elle est facilement matérialisable par un modèle physiopathologique. La mesure objective nous apporte en médecine des informations diagnostiques et pronostiques inégalables, mais elle n'explore pas le ressenti de l'individu [41].

Certaines mesures de qualité de vie (QDV) seront dites, au contraire, subjectives. Le caractère subjectif de ces mesures de QDV ne diminue pas leurs valeurs, ces mesures sont aussi précises et scientifiquement recevables que les mesures dites objectives [44][45]. Lorsque l'on s'intéresse au patient en tant que sujet pensant (heureux, fatigué, triste douloureux...), la mesure est obligatoirement subjective et tente de décrire l'intimité psychique des patients [35]. Le sujet pensant, souffrant d'une maladie et vivant, a la supériorité sur l'organe souffrant.

La QDV relative à la santé constitue donc un critère de jugement alternatif pertinent et disponible pour juger de l'intérêt du traitement pour le patient et pour le système de santé dans le champ des MNM [69][70][90][104][141] [144]. Il s'agit d'un concept large portant sur la santé physique d'un individu, son état psychologique, ses croyances personnelles, ses relations interpersonnelles et son intégration sociale [2]. Il est donc nécessaire de bien caractériser le concept de QDV avant de tenter de le mesurer. Des questionnaires spécifiques aux MNM ont été développés comme le INQoL [137]. Le PedsQL [133][134], questionnaire générique de mesure de la QDV des enfants a démontré, avec son module spécifique « Neuromusculaire », une applicabilité, une fiabilité ainsi qu'une validité satisfaisante dans une population de patients atteint d'une ASI [71] ou d'une DMD [34].

Nous avons analysé les réponses à un questionnaire de QDV générique (le « Vécu Santé Perçu – Adolescent », VSP-A) d'adolescents porteurs d'une MNM. Après avoir précisé ce que l'on entendait par le terme de QDV, notre objectif était de décrire de façon détaillée les réponses de ces adolescents et de situer leurs évaluations par rapport à la population générale. Nous avons ensuite étudié la relation entre cette mesure de QDV et les limitations d'activités de ces patients. Les résultats des mesures de QDV dans le champ des MNM doivent être interprétés avec prudence car les facteurs influençant la mesure sont multiples et plusieurs auteurs montrent l'absence de corrélation entre QDV et déficience motrice [1][3][58] [77].

Les résultats de ces travaux ont été publiés dans la revue Journal of Adolescent Health (Annexe 1) :

Self perception of quality of life by adolescents with neuromuscular diseases

Carole Vuillerot, Isabelle Hodgkinson, Alvine Bissery, Anne Marie Schott-Pethelaz, Jean Iwaz, René Ecochard, Marie Charlotte d'Anjou, Marie Christine Commare and Carole Bérard.

Purpose - Little is known about quality of life of adolescents with neuromuscular diseases or the factors that influence it. We searched whether physical impairment, physical disability, and medical complications were predictors of low quality of life.

Methods - Motor function, health, orthopaedic status, and rehabilitation were assessed in 43 adolescents with neuromuscular diseases (age: 13.8±1.7; sex ratio 2.9/1). Quality of life was measured with the VSP-A (Vécu Santé Perçu par l'Adolescent), a validated health-related quality of life self-perception test. A mixed linear regression related quality of life to impairment, disability, and respiratory status. Comparisons were made with results from an age/sex-matched nondisabled group.

Results - On the average, the VSP-A scores in physically disabled adolescents were: i) similar to those of the nondisabled group as regards vitality, body image, relationships with parents and friends, and physical and psychological well-being; ii) higher as regards school performance (score 68 vs. 52.8) and relationships with teachers (67.4 vs. 43.2); iii) lower as regards leisure activities (43.9 vs. 60.9). Physical disability and physical impairment were not negatively associated with seven out of the nine VSP-A dimensions but physical impairment was negatively associated with leisure activities and vitality ($p < 0.001$ and $p < 0.01$ respectively). Adolescents with ventilatory support did not express lower scores than non-ventilated ones (67.7 ± 11 vs. 62.9 ± 15 , $p = 0.39$).

Conclusions - These surprising results should question our medical, educational, and rehabilitation practices. Already well-managed disabled adolescents should benefit from less compassionate but more daring and dynamic interpersonal contacts.

J Adol Health 2009;46(1):70-6

D CHOISIR UN BON OUTIL D'ÉVALUATION POUR OBTENIR UNE MESURE VALIDE DE LA FONCTION MOTRICE

Après avoir précisément défini le concept à mesurer, il convient de choisir un outil approprié permettant d'obtenir une mesure précise de ce concept. Il s'agit, le plus souvent, d'un instrument d'hétéro-évaluation, c'est-à-dire qu'un « juge » extérieur évalue le patient en utilisant un outil. La façon dont va coter ce juge va participer à l'erreur de mesure (Figure 5 p : 27). Les outils standardisés avec un manuel d'utilisateur ou une formation nécessaire à leur utilisation permettent de minimiser la part de cette erreur de mesure causée par des erreurs de cotation mais il persistera toujours l'erreur due à la façon de coter des juges qui influence directement la mesure. Quelques auto-évaluations existent, il s'agit surtout d'outil mesurant les performances du patient dans son environnement (par exemple, le questionnaire ACTVLIM basé sur la propre perception de la gêne fonctionnelle, ou l'outil ABILHAND qui évalue la difficulté perçue par le patient lors de la réalisation d'activités manuelles et détermine l'habileté manuelle du patient). Pour mesurer l'impact de traitements éventuels (dans le cadre d'essais thérapeutiques) sur les capacités des patients porteurs d'une MNM, il faut disposer d'indicateurs pertinents pour un groupe de patients prédéfini avec une métrologie valide et dont la faisabilité a été vérifiée.

D.1 LES DIFFERENTS TYPES D'OUTILS

D.1.1 LES SCORES FONCTIONNELS

Dans le domaine des MNM, les travaux de Vignos [136] et de Brooke [24][25] ont permis l'établissement des premiers scores fonctionnels ou « functional rating scales ». Ces outils permettent une description imagée relativement simple du patient. Ils permettent une description rapide d'un grade ou d'un stade évolutif plus qu'une réelle mesure quantitative des limitations fonctionnelles motrices.

D.1.2 LES TESTS CHRONOMETRES

Les tests chronométrés sont de plus en plus utilisés dans les essais thérapeutiques, en particulier chez le patient porteur d'une DMD (Ataluren, IGF1, Perindopril, GSK2402968). Ils consistent à mesurer le temps nécessaire pour passer d'une position, (assise ou couchée, à une position debout ou le temps nécessaire pour parcourir une distance donnée (10 m par exemple). Ces derniers ont été utilisés pour prédire l'âge de perte de marche [87]. Il peut s'agir aussi de mesurer la distance maximale parcourue par le patient pendant un temps donné, par exemple le « 12-minute walk test » ou le « 6-minute walk test » [88][89]. Ces tests sont cependant suspectés d'être dépendants de la compliance du patient et difficilement applicable chez l'enfant de moins de 5 ans [53][113].

D.1.3 LES ECHELLES DE MESURE

Les échelles de mesure sont constituées de plusieurs exercices (ou items) rassemblés ou non en domaines. Lors de la passation d'échelle fonctionnelle motrice, le patient est soumis à une série d'exercices ; la qualité de réalisation de chaque exercice (ou item) étant cotée sur une échelle de valeur allant du parfaitement réalisé à totalement échoué.

Cette échelle de valeur, ou échelle de Likert, peut être au minimum binaire (0 : échoué, 1 : réussi) et au maximum décrire très précisément en différents niveaux la qualité du mouvement. Une cotation détaillée sur une échelle de Likert de 4 points permet une cotation plus précise et garantit une meilleure sensibilité au changement que des cotations en 3 points [16]. De même il est conseillé pour améliorer la reproductibilité de ces outils de recevoir une formation à l'utilisation de ces échelles et de disposer d'un manuel précis décrivant la façon de coter chaque item.

D.2 UNE METHODOLOGIE ROBUSTE DE VALIDATION : LA THEORIE DES QUESTIONNAIRES

Lorsque l'on s'intéresse aux mesures, on doit aborder les imprécisions en lien avec cette mesure. Une méthodologie robuste de construction de l'outil de mesure, associée à une étude de validation complète, sont des pré-requis indispensables pour s'assurer des bonnes propriétés métrologiques d'un outil et de son utilisation possible lors d'un essai thérapeutique [46].

D.2.1 LA THEORIE CLASSIQUE DES TESTS VS LA THEORIE DE REPONSES AUX ITEMS

La théorie classique des tests (TCT) repose sur l'erreur de mesure qui se définit comme toute variance non pertinente au but d'un test qui produit de l'inconsistance dans les résultats [18].

Les principaux postulats de base de la théorie classique des tests sont les suivants :

- tout score observé (X) est composé de deux parties, soit le score vrai (V) et de l'erreur (E)

$$X = V + E;$$

- E est l'erreur de mesure non liée aux buts du test et essentiellement imprévisible

- les scores vrais et les scores d'erreur correspondants ne sont pas en corrélation;

- les erreurs aléatoires ont tendance à s'équilibrer (moyenne tend vers zéro), alors que la moyenne des scores observés d'une mesure répétée tend vers le score vrai.

Selon Bernier et Pietrulewicz [18], l'erreur de mesure est composée de trois sources d'erreurs possibles, indépendantes les unes des autres :

- l'erreur engendrée par le test lui-même (l'échantillonnage des items, l'ambiguïté des énoncés, l'imprécision des instructions relatives à la façon de répondre, tout ce qui peut inciter le répondant à répondre au hasard, etc.) ;

- l'erreur liée aux conditions de passation du test (la standardisation ou non des procédures, l'expérience des examinateurs, la disposition de l'examineur lors de la passation) ;

- l'erreur liée aux répondants au test (la motivation, l'habitude de passer des tests, l'anxiété, l'apprentissage différentiel, etc.).

Notons cependant que l'écart entre la mesure et le score « vrai » est quasi inévitable lorsque le score concerne un construit tel que la fatigue, la motricité... Pour que cet écart soit nul, il faudrait que chaque sujet réponde de manière identique à chaque item. La personne à qui l'item est administré et la nature de l'item lui-même influencent la probabilité d'une réponse à cet item en particulier.

Une limitation importante de la TCT est que la façon de coter du juge et la difficulté de l'item ne peuvent pas être estimés séparément. La TCT ne donne qu'une estimation de fiabilité correspondant à une erreur de mesure standard alors que la précision de mesure est connue pour varier selon façon de coter du juge [63].

Selon la TCT, la fidélité (définition p :38) est une condition nécessaire, mais non suffisante de la validité. Il est nécessaire de connaître la proportion de variance d'erreur avant de s'intéresser à la variance vraie. Ceci implique qu'un coefficient de validité sera toujours inférieur au coefficient de fidélité. Pour un très faible coefficient de fidélité, il n'est d'aucun intérêt de s'attarder à la validité, car déjà une très grande proportion des résultats est reliée à l'erreur et n'évalue pas le domaine d'intérêt. À l'inverse, il est possible d'affirmer qu'un test valide est nécessairement fidèle.

Selon Hambleton et Jones [65], la principale limite du modèle classique est que les scores obtenus sont dépendants du test, c'est-à-dire que le score vrai d'un répondant est dépendant des questions ou tâches de l'outil par lesquelles il est testé. Hambleton et van der Linden [64] énoncent de plus que les statistiques d'un item ou d'un test sont dépendantes de l'échantillon des sujets testés.

Contrairement à la TCT, la théorie de réponse aux items (TRI) s'intéresse aux items plutôt qu'à un test dans son ensemble [30][66][135]. La TRI vise à obtenir des mesures autant que possible indépendantes de l'instrument utilisé et de l'échantillon testé.

Elle vise, d'une part, l'estimation la plus pertinente et précise possible de l'habileté des individus à partir de leurs réponses aux items et, d'autre part, l'évaluation des qualités psychométriques des items. Elle prend appui sur un modèle mathématique que l'on peut représenter par une fonction reliant l'habileté de l'individu à la probabilité de réussir un item. Cette fonction est appelée courbe caractéristique d'item [132]. Trois principaux modèles unidimensionnels prévalent dans les écrits scientifiques, et chacun d'entre eux relie l'habileté de la personne à un ou plusieurs paramètres tels l'indice de difficulté de l'item, l'indice de discrimination de l'item et l'indice de pseudo-chance [135].

Il sera ici question du modèle Rasch [109], il s'agit d'un modèle à un paramètre qui étudie la relation entre l'habileté de la personne, la difficulté de l'item et la probabilité d'une solution correcte. Il postule que les items ne se distinguent les uns des autres que de façon négligeable sur le plan de la pseudochance et du pouvoir discriminant [5]. Les analyses selon le modèle de Rasch produisent des informations détaillées concernant les échelles de cotation, les items, les personnes et d'autres facteurs telle la sévérité des juges [135]. Elles convertissent les données ordinales en mesures d'intervalles [143]. Elles permettent d'obtenir différentes mesures :

- soit les mesures d'une personne, c'est-à-dire la position de chaque personne évaluée sur le continuum, du minimum jusqu'au maximum du construit sous-jacent ;
- les mesures d'items qui indiquent quelle part du construit sous-jacent l'item représente ;
- et les mesures de l'évaluateur qui indiquent le niveau de sévérité de celui-ci.

Elles permettent également d'identifier les éléments impropres : soit qu'un item n'est pas un indicateur valide du construit, soit qu'un individu n'a pas été mesuré de façon valide, soit qu'un évaluateur n'utilise pas l'échelle de façon valide.

L'estimation de la validité d'un instrument de mesure est un processus sans fin, car il varie selon le contexte et le but de l'utilisation. L'analyse des items permet, dès les premières étapes de construction du test, de garantir l'obtention de coefficients de fidélité et de validité élevés par la sélection, la substitution et la révision des items. De nouvelles considérations théoriques ou cliniques peuvent justifier des procédures de validation différentes. Les coefficients obtenus mesurent un degré de validité, et non une notion de tout ou rien. L'interprétation doit donc se faire en considérant l'ensemble des caractéristiques du processus de validation qui est en constante relation avec le cadre théorique [5].

D.2.2 LES QUALITES METROLOGIQUES D'UN INSTRUMENT DE MESURE [7][46][112]

La pertinence ou la qualité d'un instrument de mesure utilisé par un professionnel de la santé est évaluée à partir de trois principaux aspects : sa fidélité ou fiabilité (traduction imparfaite du terme anglais « reliability »), sa validité et sa sensibilité au changement [46][112]. La fidélité s'intéresse à la capacité d'un instrument de mesure à produire des résultats précis et constants dans des situations comparables, tandis que la validité réfère principalement à sa capacité à évaluer la caractéristique d'intérêt. Un outil est discriminant si il est capable de différencier deux individus par la mesure.

D.2.2.1 La Fidélité (« reliability »)

La fidélité est la certitude qu'un test mesure des scores vrais, alors que le coefficient de fidélité donne le degré de fidélité [18]. La fidélité peut être mesurée par la capacité de l'échelle à donner des résultats comparables dans des situations comparables [46].

On dit qu'un test est fidèle dans la mesure où il y a le moins d'erreurs possible ou lorsque la variance de l'erreur est la plus petite possible. L'étude de la fidélité concerne donc l'étude des sources de variabilité associées à un instrument de mesure utilisé dans un contexte défini et auprès de sujets précis.

On distingue quatre principales façons d'estimer la fidélité :

(a) la consistance ou cohérence interne (« Internal consistency ») qui s'intéresse à l'homogénéité des items. Le coefficient alpha de Cronbach [33] évalue la cohérence interne d'un ensemble d'items correspondant à un domaine clinique, c'est à dire la force des inter-corrélations entre items d'un même domaine. Les items de chaque domaine devant former un tout cohérent. Plus les items sont liés entre eux plus le coefficient alpha est proche de 1. En pratique, la cohérence interne doit être assez grande (coefficient alpha > 0,70-0,80), sans que le coefficient ne soit trop proche de 1.

Un coefficient de 1 peut, en effet, signifier que plusieurs items sont redondants. Notons cependant que cette mesure n'est pas invariante : le coefficient alpha varie d'une population à l'autre. Un alpha faible dans une population homogène sera plus fort dans une population aux caractéristiques hétérogènes.

(b) la fidélité « test-retest » qui évalue la stabilité des résultats dans le temps d'une mesure. L'intervalle entre les deux mesures, qui est généralement court, doit être suffisant pour éviter les phénomènes d'apprentissage du test utilisé. La situation clinique à mesurer doit être stable entre les 2 mesures. En général, l'indice de test-retest est légèrement inférieur à l'indice de cohérence interne, un coefficient > 0.75 indique une bonne stabilité test-retest. La fidélité test-retest peut aussi se rapprocher de la notion de reproductibilité pour une auto-évaluation.

(c) la fidélité intra-juge (« Intra-rater reliability ») indique la cohérence des mesures. Elle est définie par la stabilité du résultat au cours des mesures répétées, l'état du sujet restant stable pendant ce laps de temps. Elle est à rattacher de la notion de précision de la mesure. Elle peut aussi se rapprocher de la notion de reproductibilité pour une hétéro-évaluation.

(d) la fidélité inter-juge (« Inter-rater reliability ») qui estime le degré de concordance entre des observateurs indépendants mesurant le même phénomène dans le cadre obligatoire d'une hétéro-évaluation [5][18][40]. Dans ce cas l'erreur de mesure résulte de l'imprécision du questionnaire et des variations propres aux juges eux-mêmes.

Pour évaluer de manière précise ces fidélités, plusieurs méthodes peuvent être utilisées. Le coefficient de corrélation intra-classe (ou ICC) est le quotient de la variance inter-sujet de la grandeur étudiée divisée par la variance totale (celle inter-sujet + celle de l'erreur résiduelle). Il a été montré, sous des conditions assez générales, qu'il est équivalent à un kappa pondéré [50]. Or, un kappa pondéré considère qu'il y a concordance aussi sur des éléments en dehors de la diagonale dans le tableau croisé, et donc un kappa pondéré est plus élevé qu'un kappa non pondéré sur les mêmes données. Si l'ICC est considéré comme s'appliquant uniquement sur des échelles quantitatives ou pouvant être considérées comme, le kappa, au contraire, est alors plutôt utilisé pour des échelles ordinales. Dans tous les cas, kappas ou ICC, ces grandeurs dépendent de la variabilité de l'échantillon et cela est surtout vrai pour l'ICC. La valeur d'un ICC calculé sur un échantillon de sujets varie de zéro (voire inférieur à Zéro) à un. La fidélité est d'autant plus grande que l'ICC est proche de 1.

La technique de Bland et Altman [22] est complémentaire du calcul des ICC. Les principales étapes de cette méthode sont :

- la représentation graphique du nuage de points figurant les sujets. On porte pour chaque patient, la différence d de ses deux cotations en ordonnée et leur moyenne en abscisse ;
- une étude visuelle du nuage de points pouvant indiquer par exemple que le désaccord inter juges est d'autant plus grand que les cotations sont élevées ;
- le calcul des limites contenant 95 % des différences d ;
- la détection des patients dont les points sont en dehors de ces limites et recherche éventuelle de la cause de ces grands désaccords.

Bien qu'étant essentielle, en estimant le pourcentage d'erreur et conséquemment le pourcentage du score vrai, la fidélité n'est pas suffisante pour prétendre à la validité d'un instrument de mesure [40].

D.2.2.2 La validité

La validité s'intéresse à savoir si un instrument, utilisé dans un objectif et un contexte particulier et auprès de sujets particuliers, mesure ce qu'il prétend mesurer et s'il varie avec ce qu'il mesure.

On reconnaît quatre types de validité :

(a) La validité d'apparence (« face validity ») représente le jugement subjectif en fonction de l'utilisateur prenant en compte les aspects visibles de l'échelle de façon superficielle : la longueur du questionnaire, le libellé des items, les modalités de réponse...

(b) La validité de contenu (« content validity ») s'intéresse à la représentativité de l'ensemble des énoncés qui constituent le concept à mesurer. Encore appelée spécificité, elle concerne la pertinence du contenu de l'outil établie par un jugement d'experts. Elle juge si les questions sélectionnées représentent bien toutes les facettes du concept à mesurer. La sélection des items retenus pour composer l'outil doit avoir été réalisée par une équipe composée d'experts médicaux et parfois de malades.

(c) la validité de critère (« criterion validity ») estime l'intensité du lien statistique (degré de corrélation) entre un instrument de mesure et une autre mesure indépendante considérée comme la référence. On dissocie la validité concomitante ou concordante (« concurrent validity ») où les deux mesures sont réalisées dans le même temps et la validité prédictive (« predictive validity ») où il s'agit de déterminer si un événement peut être anticipé ou prédit par un changement du résultat de la mesure ; il existe dans ce cas un délai entre les deux mesures de plusieurs semaines ou plusieurs mois.

(d) la validité de construit (« construct validity ») qui étudie le degré auquel un test mesure le construit qu'il prétend mesurer [5][18][40]. Elle s'affirme au fur et à mesure que des expériences successives confirment les hypothèses théoriques émises lors de la conception de l'échelle. Elle correspond au degré de précision avec lequel les concepts à l'étude sont représentés par des énoncés spécifiques. La validité contre critère peut donc être incluse dans la validité de construit car elle mesure le degré auquel on peut affirmer que le construit opérationnalisé est corrélé au concept qu'il est censé représenter.

Du point de vue de la théorie classique des tests, la validité représente la proportion de la variance vraie qui est pertinente aux buts de l'utilisation d'un test.

D.2.2.3 Sa sensibilité au changement (Responsiveness) [12]

La sensibilité au changement est une propriété distincte du pouvoir discriminant. Un instrument est dit sensible au changement s'il est capable de mesurer avec précision les variations en plus (amélioration) ou en moins (aggravation) du phénomène mesuré.

Il doit permettre un classement suffisamment fin des individus et être capable de repérer des variations cliniquement perceptibles [38][39]. Ces variations peuvent être spontanées, du fait de l'évolution naturelle de la maladie ou provoquées par une action thérapeutique quelconque. Si la sensibilité au changement d'un outil est insuffisante, on conclurait à tort à l'inefficacité de l'action entreprise.

Il s'agit de la capacité de l'outil d'évaluation à détecter un changement dont l'importance est significative et observable d'un point de vue clinique. Cette qualité métrologique permet de répondre à la question suivante : De combien d'unités doit varier le résultat obtenu pour que je puisse affirmer que la condition de mon patient a réellement changé ? Cette valeur se nomme la plus petite différence cliniquement significative (« minimal clinically important difference », MCID) [37][62][111].

Pour être évaluée la sensibilité au changement nécessite donc un suivi longitudinal suffisant des patients. La sensibilité au changement est dépendante de la reproductibilité. Pour les anglo-saxons, la « sensitivity to change » est la capacité à détecter un changement quelle que soit sa valeur alors que le « responsiveness » est la capacité à détecter un changement cliniquement pertinent.

La taille de l'effet (« effect size ») est l'importance du changement moyen dû à l'intervention ou à la progression d'une maladie comparée à la variabilité de la mesure chez des sujets stables au temps initial (changement moyen/écart type au temps initial).

La réponse moyenne standardisée (« standardized response means ») équivaut au ratio entre le changement moyen sur l'écart type du changement.

Il faut être prudent dans l'interprétation des résultats obtenus à l'aide d'un outil d'évaluation pour lequel aucune « plus petite différence cliniquement significative » n'est fournie. En effet, il est alors difficile pour le clinicien de déterminer de combien d'unités doit varier le score d'un patient à l'outil pour pouvoir affirmer que le patient a réellement évolué.

D.2.2.4 L'utilité ou applicabilité en pratique

L'utilité ou l'applicabilité réfère, quant à elle, à des facteurs pratiques tels le coût, la facilité d'utilisation, le temps qu'il requiert, etc. [86].

L'échelle doit être un outil pragmatique et facilement applicable. L'outil doit être acceptable dans la pratique en fonction de la population d'étude, en particulier chez l'enfant il ne doit pas être trop long. Les critères sont le temps de réponse au questionnaire, le nombre de refus de participer, la compliance (nombres d'items renseignés sur la totalité des items).

Andersen [7] propose des guidelines sous la forme de critères d'évaluation simplifiés pour analyser les différentes propriétés métrologiques des divers outils de mesure dans le champ du handicap. Ces différents critères sont cotés en grades ; le grade A équivalent à « excellent published track record » alors que le grade C équivalent à « insuffisance scientifique des résultats publiés ». Le grade B correspond à un niveau intermédiaire (Tableau 2)

CARACTERISTIQUES	DEFINITION	CRITERES D'EVALUATION UTILISES
CONCEPTION DE L'OUTIL	Les différents domaines identifiés sont-ils mesurés par l'outil ? Ces concepts sont-ils issus de la CIF ?	Les différents domaines identifiés sont-ils : A- totalement couverts par la mesure B- partiellement couverts (par exemple, faiblesse dans quelques domaines ou manquement de représentation dans un domaine dit central C- insuffisamment couverts
NORMES, VALEURS DE REFERENCES	Des valeurs de référence en population générale sont-elles disponibles par genre, groupe d'âge ? Y a-t-il des données comparatives avec des PH ?	Des données publiées sont disponibles pour : A- La population générale et les PH B- La population générale ou les PH C- Aucune norme n'est publiée pour cet outil
MODELE DE LA MESURE	L'outil capture-t-il les détails du concept, sa globalité ou les différences réelles entre les personnes ? Y a-t-il des effets plafond ou plancher ?	20% des sujets ont des scores situés aux extrêmes de l'outil A- Oui B- Quelques problèmes C- Effet plancher ou plafond important de l'outil
VALIDITE DE CONTENU ET D'APPARENCE DE L'OUTIL	Les items de l'outil représentent-ils correctement le concept mesuré et sont-ils adaptés aux PH ?	Evaluation qualitative A- Les PH ont participé au développement de l'outil et l'acceptabilité a été publiée B- L'étude de validité d'apparence montre de légers biais C- Les biais sont évidents
APPLICABILITE	Est-ce que le temps de passation et le contenu sont acceptables pour les sujets interrogés ? Les PH ont-elles été impliquées dans le développement, ou ont-elles fait des remarques sur l'applicabilité de l'outil ?	La mesure est A- Brève (< 15 minutes) et acceptable pour les PH B- Plus longue (mais appropriée) ou avec quelques problèmes d'acceptabilité C- De longueur et d'acceptabilité problématiques pour les PH
PASSATION DE L'OUTIL	Avec quelle facilité l'outil est-il administré, scoré et interprété ?	A- Score calculé à la main, de résultats métriques appropriés et facilement interprétable par les chercheurs, cliniciens et PH B- Calculs des scores de façon informatisée et interprétation plus obscure des résultats C- Calculs coûteux, complexes et/ou interprétation difficile
	L'outil est-il disponible en différentes versions ? Y a-t-il des adaptations pour les PH lourdement ? Existe-t-il un manuel d'utilisateur ?	A- Différentes versions sont disponibles et évalués ; des adaptations sont disponibles pour PH ; Il existe un manuel d'utilisateur B- Quelques adaptations ont été faites pour les PH C- Aucune adaptation ni manuel d'utilisateur n'est disponible

CARACTERISTIQUES	DEFINITION	CRITERES D'EVALUATION UTILISES		
VALIDITE	L'outil mesure-t-il ce qu'il prétend mesurer ? L'outil permet-il de distinguer des PH avec des niveaux de fonctionnement différents en termes d'activité ou de participation ? (validité convergente, discriminante)	Corrélations Convergentes Multitrait Multimethod Matrix A- $\geq 0,6$ B- $>0,3$ à $<0,6$ C- $\leq 0,3$	Validité discriminante Différences entre groupes conformes à la hiérarchie des degrés de sévérité observés	Analyses factorielles A- Confirmée B- Qques problèmes C- Faible
REPRODUCTIBILITE	L'outil fournit-il des réponses cohérentes ? La fiabilité test-retest ou inter-observateur est-elle satisfaisante, de même que la cohérence interne de l'échelle ?	ICC ou Kappa Test Retest A- $\geq 0,75$ B- $0,4$ à $0,75$ C- $<0,4$	Alpha de Cronbach Cohérence interne A- $\geq 0,8$ B- $<0,8$ et $>0,7$ C- $<0,7$	ICC Inter observateur A- $\geq 0,75$ B- $>0,4$ à $<0,75$ C- $<0,4$
SENSIBILITE AU CHANGEMENT	Est-ce que l'outil est sensible aux changements importants ?	Critères de jugement de la sensibilité au changement A- Statistiquement significative, confirmée par une impression clinique B- Evidence modérée C- Faible ou basée uniquement sur des critères statistiques		

Tableau 2 : Critères d'évaluation des propriétés métrologiques d'outils de mesure en santé selon Andersen

Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 2000; 81(12 suppl 2): 15-20 PH : Personnes handicapé

CHAPITRE 2 : DEVELOPPEMENT ET TRAVAUX

A REVUE DE LA LITTERATURE DES DIFFERENTS INSTRUMENTS

A.1 OBJECTIFS

L'objectif de cette revue de la littérature était d'identifier et de décrire tous les outils évaluant la fonction motrice utilisables chez les patients porteurs d'une MNM. En effet, dans ce champ des MNM, il n'est pas possible de trouver « le » seul outil ou « la » seule échelle qui pourraient être utilisés pour toutes les études et toutes les cohortes, indépendamment de la maladie, de l'âge et de nombreuses autres variables.

A.2 METHODES

A.2.1 DE LA RECHERCHE BIBLIOGRAPHIQUE

Différentes sources disponibles sur Internet ont été interrogées :

- la base de données bibliographique PUBMED qui indexe les articles du domaine biomédicale (www.ncbi.nlm.nih.gov/pubmed/) ;
- les bases Cochrane, EMConsulte, Google Scholar et Science Direct qui ont permis d'élargir la recherche aux résumés de congrès et aux travaux de thèse ;
- le registre « Registry of Outcomes » (ROM) (<http://www.researchrom.com/search>), géré par le réseau Treat-NMD, qui vise à répertorier les critères de jugement disponibles pour les équipes menant à bien des études cliniques dans le domaine des MNM.

Les mots-clés retenus pour les recherches ont été :

- pour les échelles : « Outcome assessment (health care) », « severity of illness index », « questionnaires », « scale » « disability evaluation », « neurologic examination » ;
- pour la pathologie : « motor skills », « motor skills disorders », « locomotion », « neuromuscular diseases », « neuromuscular manifestations ».

Critères d'inclusion :

Ont été retenus les outils :

- fournissant une mesure de la fonction motrice. Certains outils pouvant mesurer plusieurs concepts en lien avec l'atteinte musculaire dont la fonction motrice ;
- applicable aux patients porteurs d'une MNM ;
- avec des données publiées dans la littérature internationale.

Critères d'exclusion :

N'ont pas été retenus les outils :

- ne s'appliquant pas au moins à la DMD ou à l'ASI, MNM de l'enfance les plus fréquentes et pour lesquelles des essais thérapeutiques sont en marche ;

- développés qu'en français ou non publié dans la littérature internationale malgré leurs valeurs probables, comme le score moteur développé par le professeur Annie Barois pour les ASI [11].

Cette analyse ne correspond pas à une recherche exhaustive, mais plutôt à une analyse approfondie pour chaque outil identifié comme outil de mesure de la fonction motrice applicable aux MNM selon la littérature internationale.

A.2.2 DE L'ANALYSE DES QUALITES METROLOGIQUES

Le travail a consisté à analyser, pour chacun des outils inclus, les propriétés métrologiques disponibles dans les données de la littérature. La qualité des études de validation a été étudiée pour chaque outil selon une méthodologie précise inspirée des travaux d'Elena M. Andersen [7] et de Jacques Fermanian [46] sur les qualités métrologiques minimales pour un outil d'évaluation en MPR. Il ne s'agit pas de donner un classement des outils, du meilleur au moins bon, mais plutôt de fournir pour chacun ses forces et ses faiblesses pour orienter le lecteur vers le choix de l'outil approprié en fonction du design précis de l'étude et, plus précisément, en fonction de la population incluse dans l'essai (type de maladie neuromusculaires, âge des patients et évolutivité de la maladie), de l'état fonctionnel des patients au moment du recrutement, de la durée de l'étude et enfin de l'effet possible du traitement [101].

A.3 RESULTATS DE LA REVUE DE LA LITTERATURE

Dix-neuf outils répondant aux critères de sélection initiaux ont été retenus. Ces outils sont différenciés selon leur type (score fonctionnel, test chronométrés ou échelle de mesure) et selon ce qu'il mesure (Tableau 3). Les échelles de mesures sont différenciées selon qu'il s'agisse de questionnaires, mesurant plutôt des performances, ou de passation d'items en situation de test, mesurant plutôt des capacités motrices.

De manière à favoriser les comparaisons et le travail de synthèse, une fiche résumant les différentes données disponibles dans la littérature a été rédigée pour chaque outil (pp 48 à 67).

Lorsque l'outil est disponible en ligne, un lien est indiqué dans DESCRIPTION permettant d'aboutir aux documents PDF.

A.3.1 TABLEAU RECAPITULATIF

Le tableau 3 résume les différents outils décrits dans la littérature classés en fonction du concept mesuré dans la CIF. Plusieurs outils, dits « composites », évaluant des grandeurs différentes (MDFRS, ACTIVLIM, WeeFIM) sont classés dans plusieurs domaines de la CIF.

FONCTIONS ORGANIQUES	ACTIVITES		PARTICIPATION
	CAPACITE	PERFORMANCE	
b2 : Fonctions sensorielles et douleurs - Douleur par excès de nociception neuropathiques -Troubles sensitifs	d4 : Mobilité -Marcher et se déplacer 6MWT -Changer et maintenir la position du corps -Porter déplacer et manipuler des objets UL MODULE CHOP INTEND EXPHFMF MHFMS MFM 32 & 20 TIMP NSAA	d4 : Mobilité Marcher et se déplacer VIGNOS WeeFIM EK Scale MDFRS Changer et maintenir la position du corps VIGNOS Porter déplacer et manipuler des objets ABILHAND ACTIVLIM BROOKE MDFRS	d6 : Vie domestique -Réalisation des tâches ménagères -Résolution de problème WeeFIM EK Scale MDFRS
b7 : Fonctions de l'appareil locomoteur et liées au mouvement -Force musculaire -Mobilité articulaire -Trophicité musculaire -Tonus musculaire MDFRS	HFMS HMAS MDFRS GMFM		d8 : Grands domaines de la vie -Education scolaire -Accès à l'emploi
b4 : Fonctions des systèmes cardio-vasculaire, hématopoiétique, immunitaires et respiratoire -Fonction respiratoire MDFRS -Fonction cardiaque	d5 : Entretien personnel	d5 : Entretien personnel ACTIVLIM WeeFIM Ek Scale MDFRS	d9 : Vie communautaire, sociale et civique -Récréation et loisirs -Religion

Tableau 3 : Les différents outils d'évaluation de la fonction motrice selon la CIF

A.3.2 DESCRIPTION DETAILLEE DE CHAQUE OUTIL

SCORE DE BROOKE

TYPE D'OUTIL : SCORE FONCTIONNEL

CE QUI EST MESURE : CAPACITES MOTRICITE PROXIMALE ET DISTALE (MEMBRES SUPERIEURS)

POPULATION / AGE	ENFANTS
PATHOLOGIES	DYSTROPHIES MUSCULAIRES PROGRESSIVES
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	<ol style="list-style-type: none"> 1. Starting with arms at the sides, the subject can abduct both arms in a full circle until they touch above the head 2. Can raise both arms above head only by flexing the elbow (i.e. shortening the circumference of the movement) or using accessory muscles 3. Cannot raise hands above head but can raise an 8oz (250 ml) glass of water to mouth (may use one or both hands) 4. Can raise hands to mouth (can raise each hand separately) but cannot raise an 250 ml glass of water to mouth 5. Cannot raise hand to mouth but can use hands to hold pen or pick up coins from table 6. Cannot raise hands to mouth and has no useful function of hands
FORMATION	NON
NOMBRED'ITEMS	6
DUREE DE LA PASSATION	
COTATION	1
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	
METHODOLOGIE DE VALIDATION	
VALIDITE D'APPARENCE ET DE CONTENU	
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	
FIDELITE INTER OBSERVATEUR	ICC = 0.96-0.99 (95% CI)
FIDELITE INTRA OBSERVATEUR	ICC = 0.96-0.99 (95% CI)
SENSIBILITE AU CHANGEMENT	Peu sensible au changement [87][124]
ESSAIS THERAPEUTIQUES	DELOS : Idebenone dans la DMD (en cours)
COMMENTAIRES	
REFERENCES	[24][25]

Tableau 4 : Score de Brooke

SCORE DE VIGNOS

TYPE D'OUTIL : SCORE FONCTIONNEL

CE QUI EST MESURE : PERFORMANCES MARCHE STATION DEBOUT ET MONTEE DES ESCALIERS (MEMBRES INFERIEURS)

POPULATION / AGE	ENFANTS
PATHOLOGIES	DYSTROPHIES MUSCULAIRES PROGRESSIVES
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	<ol style="list-style-type: none"> 1. Walks and climbs stairs without assistance 2. Walks and climbs stairs with aid of railing 3. Walks and climbs stairs slowly with aid of railing (over 12 seconds for 4 standard stairs) 4. Walks unassisted and rises from chair but cannot climb stairs 5. Walk unassisted but cannot arise from chair or climb stairs 6. Walks only with assistance or walks independently with long leg braces 7. Walks in long leg braces but requires assistance for balance 8. Stands in long leg braces but unable to walk even with assistance 9. Is in wheelchair 10. Confined to bed
FORMATION	NON
NOMBRE D'ITEMS	10
DUREE DE LA PASSATION	
COTATION	1 A 10
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE CONSTRUCTION	Version modifiée d'une classification publiée par Swinyard CA et al. <i>Gradients of functional ability of importance in rehabilitation of patients with progressive muscular and neuromuscular disease, Arch Phys Med 1957;38:574</i>
METHODOLOGIE DE VALIDATION	
VALIDITE D'APPARENCE ET DE CONTENU	
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	
FIDELITE INTER OBSERVATEUR	
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	DELOS : Idebenone dans la DMD (en cours)
COMMENTAIRES	
REFERENCES	[136]

Tableau 5: Score de Vignos

6-MINUTE WALK TEST 6MWT

TYPE D'OUTIL : TESTS CHRONOMETRES

CE QUI EST MESURE : CAPACITES MOTRICES DEFINIES COMME LA DISTANCE PARCOURUE EN 6 MINUTES

POPULATION / AGE	ENFANTS
PATHOLOGIES	DYSTROPHIE MUSCULAIRE DE DUCHENNE AMYOTROPHIE SPINALE INFANTILE
STATUT/FONCTIONNEL	PATIENTS MARCHANTS
DESCRIPTION	
FORMATION	NON
NOMBRED'ITEMS	1
DUREE DE LA PASSATION	15-20 MINUTES AVEC INSTRUCTIONS
COTATION	
VERSIONS DISPONIBLES VALIDEES	
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Version modifiée adaptée aux enfants: Couloir intérieur d'environ 8 pieds de large avec une surface dure et un marquage au sol sur 25m, Encouragements tout au long du test, Chutes et pauses prises en compte dans le résultat
METHODOLOGIE DE VALIDATION	DMD : 1/Validité Test- retest (1 semaine) 2/Corrélation avec âge, données anthropométriques, cadence, la longueur du pas 3/Etude longitudinale pour analyser la sensibilité au changement : chez 18 patients dont 13 sous corticostéroïdes (délai moyen entre 2 évaluations entre 39 et 87 semaines) en comparaison avec 22 enfants sains 4/ Etude de sensibilité au changement de plusieurs outils fonctionnels sur 1 an (NSAA, 6MWT) ASI : 1/Corrélation avec autres mesures : HFMSE, force musculaire, 10m marche/course (temps)
VALIDITE CONTRE CRITERE	DMD : Corrélation avec poids, taille: $r > 0.89$: Longueur du pas et cadences sont des variables prédictives du 6MWT ; $r = 0.52$ $p < 0,001$ (NSAA) ASI : Corrélation avec HFMSE score ($r = 0.83$, $p < 0.0001$), 10-m marche/course ($r = -0.87$, $p < 0.0001$), and force musculaire flexion genou ($r = 0.62$, $p = 0.01$)
AUTRE VALIDITE DE CONSTRUIT	
FIDELITE INTER OBSERVATEUR	$r = 0,91$; ICC=0,91
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	67% des patients DMD ont une diminution de leur 6MWT (moyenne -15,9%) contre une augmentation de 2,1% du 6MWT chez les sujets sains ($p < 0.027$). Changement moyen sur 1 an chez DMD: -25.8m (SD 74.3)
ESSAIS THERAPEUTIQUES	PTC 124 : Ataluren chez les patients DMD IGF1 chez les patients DMD Perindopril chez les DMD < 7 ans DMD114117 : GSK2402968 chez les DMD < 5 ans marchants
COMMENTAIRES	Plus de chutes dans le groupe DMD que sujets sains
REFERENCES	[10][88][89][97][102]

Tableau 6: 6-minute walk test

ABILHAND FOR NMD

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE (QUESTIONNAIRE)

CE QUI EST MESURE : PERFORMANCES DANS LES ACTIVITES MANUELLES

POPULATION / AGE	ENFANTS ET ADULTES
PATHOLOGIES	MNM
STATUT/FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Presentation: http://www.rehab-scales.org/activlim-presentation.html Cotations : http://www.rehab-scales.org/abilhand-kids-downloads.html
FORMATION	NON
NOMBRED'ITEMS	22 (4 SPECIFIQUE ENFANTS) 4 (SPECIFIQUE ADULTES) ET 14 (COMMUN)
DUREE DE LA PASSATION	
COTATION	0-1-2 (IMPOSSIBLE-DIFFICILE-FACILE)
VERSIONS DISPONIBLES VALIDEES	ANGLAIS, FRANÇAIS, HOLLANDAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Première version avec 53 items pour enfants et 44 items pour adultes (avec 35 items en commun) proposée par un groupe d'experts à partir d'Abilhand et d'Abilhand kids : 124 patients adultes et les parents de 124 patients enfants ont été interrogés La sélection des 22 items de la version pour MNM a été faite grâce à des analyses successives sur cette population en utilisant le modèle Rasch(avec comme hypothèse :échelle ordinaire de cotation, unidimensionalité, monotonie, indépendance locale, absence de redondance entre items)
METHODOLOGIE DE VALIDATION	Modèle Rasch
VALIDITE CONTRE CRITERE	ABILHAND NMD mesures corrélée fortement avec ACTIVLIM mesures ($r=0.76$) mais très peu avec la force musculaire (grip)($r=0.36$ main droite et $r=0.40$ main gauche).
AUTRE VALIDITE DE CONSTRUIT	Mesure unidimensionnelle fournissant une mesure linéaire de la fonction motrice des membres supérieurs. La difficulté des items est continue et progressive. L'échelle possède une bonne précision ($r=0.95$) et les valeurs des paramètres de difficulté des 22 items présentent une bonne correspondance avec les valeurs de la variable latente des patients de l'échantillon, ainsi qu'un bon continuum et une absence de recouvrement.
COHERENCE INTERNE	
FIDELITÉ INTER OBSERVATEUR	
FIDÈLITÉ INTRA OBSERVATEUR	
SENSIBILITÉ AU CHANGEMENT	
ESSAISTHÉRAPEUTIQUES	
COMMENTAIRES	Auto-évaluation(adultes) et hétéroévaluation (enfants)
REFERENCES	[131]

Tableau 7: Abilhand for neuromuscular diseases

ACTIVLIM

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE (QUESTIONNAIRE)

CE QUI EST MESURE : PERFORMANCES DANS LES ACTIVITES DE LA VIE QUOTIDIENNE

POPULATION / AGE	ENFANTS ET ADULTES
PATHOLOGIES	MNM
STATUT/FONCTIONNEL	MARCHANTS ET NON MARCHANTS
DESCRIPTION	Presentation : http://www.rehab-scales.org/activlim-presentation.html Cotation : http://www.rehab-scales.org/activlim-downloads.html
FORMATION	NON
NOMBRED'ITEMS	22 (4 SPECIFIQUE ENFANTS) 4 (SPECIFIQUE ADULTES) ET 14 (COMMUN)
DUREE DE LA PASSATION	
COTATION	0-1-2 (IMPOSSIBLE-DIFFICIEL-FACILE)
VERSIONS DISPONIBLES VALIDEES	FRANÇAIS, HOLLANDAIS, ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Version préliminaire contenant 126 items issus de différentes échelles existantes (ABILHAND, SF 36, SIP, ALSFRS, ADL self report, PQLI, EK scale...) + Avis d'experts + avis de patients Sélection des items (22) : utilisation du modèle Rasch avec comme hypothèses (échelle ordinale de cotation, unidimensionnalité, monotonie, indépendance locale, absence de redondance entre items) Population : 124 enfants et 245 adultes
METHODOLOGIE DE VALIDATION	Utilisation de modèle Rasch pour étude de la validité de contenu et de construit Utilisation modèle TCT pour les propriétés classiques (contre critère, fiabilité inter observateur et sensibilité au changement)
VALIDITE CONTRE CRITERE	Corrélation avec MIF (p=0,85), Vignos (p=-0,83) et Brooke (p=-0,63)
AUTRE VALIDITE DE CONSTRUIT	Pas de différences significatives dans les scores entre âge, langue, genre. Différences significatives (F=15,92 p<0,001) entre types de MNM Items classés par difficulté des plus difficiles aux moins difficiles (logit range 3,57 to -3.33).
COHERENCE INTERNE	Caractère unidimensionnel de l'échelle : pas de différence significative entre scores attendus et observés
FIDELITE INTER OBSERVATEUR	Fidélité Test –retest : ICC = 0,93 à 0,98
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	Etude réalisée sur 132 patients évalués 2 fois à 21+/- 4 mois d'intervalle avec Activlim : Approche par groupe et approche individuelle : 1/ les patients se disant aggravés fonctionnellement ont des scores moyens qui diminuent de façon significative 2/ par diagnostic(DMD > MD et NP)
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	Autoévaluation (adultes) et hétéro-évaluation (enfants)
REFERENCES	[129][130]

Tableau 8: Activlim

CHOP INTEND

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION / AGE	ENFANTS (jeunes enfants et nourrissons)
PATHOLOGIES	ASI DE TYPE 1 ET TOUTES PATHOLOGIES NEUROMUSCULAIRES AVEC FAIBLESSE MUSCULAIRE MAJEURE
STATUT/FONCTIONNEL	PATIENTS NON MARCHANTS
DESCRIPTION	Feuille de cotation : http://www.columbiasma.org/docs/cme-2010/CHOP%20INTEND%20for%20SMA%20Type%20I%20-%20Score%20Sheet.pdf
FORMATION	NON
NOMBRED'ITEMS	16
DUREE DE LA PASSATION	
COTATION	0-1-2-3-4
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	2 étapes : 1/Première version CHOP TOSS à partir de différentes échelles 2/ Test expérimental (26 enfants SMA I) avec 3 échelles CHOP TOSS, TIMP, Peabody : Coefficient de Cronbach utilisé pour déterminer les items pertinents et éliminer les items redondants Coefficient de Cronbach utilisé pour réduire le nombre d'items
METHODOLOGIE DE VALIDATION	Etude de la validité inter (10 enfants avec MNM cotation par vidéo du fait de la fatigabilité des enfants & 8 enfants sains) et intra observateur (9 enfants SMA I)
VALIDITE CONTRE CRITERE	Corrélations faites avec le nombre de copies SMN2 (3 copies $r=-0,83$), assistance respiratoire ($r = -0.74$, $P < .0001$, $n = 26$)
AUTRES VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	ICC (10 enfants avec MNM)=0,98 ICC (8 enfants sains)=0,93
FIDELITE INTRA OBSERVATEUR	ICC (9 enfants avec ASI I) = 0.96
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[55][56]

Tableau 9 : CHOP INTEND

EGENKLASSIFIKATIONSCALE

EK- SCALE

TYPE D'OUTIL : ECHELLE DE MESURE COMPOSITE (QUESTIONNAIRE)

CE QUI EST MESURE : PERFORMANCES MOTRICES GLOBALES + PERFORMANCES ACTIVITES VIE QUOTIDIENNE + SYMPTOMES RESPIRATOIRES.

POPULATION / AGE	ENFANTS ET ADULTES
PATHOLOGIES	DMD & ASI
STATUT FONCTIONNEL	PATIENTS NON MARCHANTS
DESCRIPTION	
FORMATION	OUI
NOMBRE D'ITEMS	10 (7 ITEMS POUR LE TRONC ET LA MOBILITE DES MEMBRES SUPERIEURS ET 3 ITEMS POUR LES SYMPTOMES RESPIRATOIRES)
DUREE DE LA PASSATION	
COTATION	0-1-2-3
VERSIONS DISPONIBLES VALIDEES	DANOIS, ANGLAIS, ESPAGNOL
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Items proposés par 2 experts danois en MNM. Traduction de l'échelle en anglais et modifications apportées par des experts internationaux à l'occasion de plusieurs réunions. The scale was translated into English and in its original
METHODOLOGIE DE VALIDATION	Etude de validation de contenu et contre critère: 56 patients DMD et 38 patients ASI Etude longitudinale (5 ans) chez 19 patients DMD et 13 patients ASI non marchants
VALIDITE CONTRE CRITERE	DMD : Spearman Corrélation : Force musculaire (-0,66 p<0,01), Capacité vitale (-0,42 p<0,05)
AUTRE VALIDITE DE CONSTRUIT	Analyse factorielle
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	ICC= 0,98 (95% CI) kappa = 0,67-0,94
FIDELITE INTRA OBSERVATEUR	ICC= 0,98 (95% CI)
SENSIBILITE AU CHANGEMENT	19 patients DMD : Changement significatif dans le score EK (-4 points, p<0.001) ; 13 patients ASI : Pas de changement significatif (-1 point, p=0.09)
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	10 minutes annoncées pour la passation mais pas de moyenne dans les publications
REFERENCES	[123][124]

Tableau 10: EgenKlassifikation scale

EXPANDED HAMMERSMITH FUNCTIONAL MOTOR SCALE HFMSE

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE
CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION / AGE	ENFANTS
PATHOLOGIES	ASI III
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Feuille de cotation : http://www.columbiasma.org/docs/cme2010/Hammersmith%20Functional%20Motor%20Scale%20Expanded%20for%20SMA%20Type%20II%20and%20III%20-%20Score%20Sheet.pdf
FORMATION	NON
NOMBRE D'ITEMS	33
DUREE DE LA PASSATION	
COTATION	0-1-2
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Association 20 items HFMS (cotés 0-1-2) + 13 items sélectionnés dans la GMFM (cotés 0-1-2-3)
METHODOLOGIE DE VALIDATION	Etude de la validité contre critère et fiabilité inter observateur sur 38 patients ASI (21 type 2 et 17 type 3)
VALIDITE CONTRE CRITERE	Coefficient Spearman=0,97 (GMFM 75) Coefficient Spearman =0,90 (Echelle de fonction motrice non validée)
AUTRE VALIDITE DE CONSTRUIT	Peu d'effet plafond contrairement à l'HFMS dans l'ASI de type 3
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	
FIDELITE INTRA OBSERVATEUR	ICC=0,99 (95% CI)
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[105]

Tableau 11 : Expanded Hammersmith functional motor scale

FUNCTIONAL INDEPENDENCE MEASURE FOR CHILDREN WEEFIM

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE COMPOSITE (QUESTIONNAIRE)

CE QUI EST MESURE : 3 DOMAINES : PERFORMANCES MOTRICES, PERFORMANCES DANS LES ACTIVITES DE LA VIE QUOTIDIENNE (AVQ), ET COGNITION

POPULATION / AGE	ADULTES ET ENFANTS
PATHOLOGIES	ASI
STATUT FONCTIONNEL	MARCHANTS ET NON MARCHANTS
DESCRIPTION	http://www.udsmr.org/Documents/WeeFIM/WeeFIM_II_System.pdf
FORMATION	NON
NOMBRE D'ITEMS	18 ITEMS
DUREE DE LA PASSATION	
COTATION	1-2-3-4-5-6-7
VERSIONS DISPONIBLES VALIDEES	ANGLAIS, FRANÇAIS, CHINOIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Adaptation directe de la MIF adulte
METHODOLOGIE DE VALIDATION	39 patients atteints d'ASI ont été évalués par la MIF. Etude de la différence des scores moyens entre chaque groupe (ASI de type 1, 2 et 3) dans chaque domaine (mobilité, activités de la vie quotidienne et cognition)
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	Différence significative entre les types I et II et les types IIIa et IIIb concernant le score global par contre pas de différence entre II et IIIa
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[31][32][52]

Tableau 12: Functional independence measure for children

GROSS MOTOR FUNCTION MEASURE

GMFM

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES EN 5 DOMAINES (A) COUCHE (B) POSITION ASSISE (C) 4 PATTES (D) DEBOUT (E) MARCHE, COURSE ET SAUT

POPULATION / AGE	ENFANTS (5-18 ANS)
PATHOLOGIES	ASI I,II,III
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Feuille de cotation : http://motorgrowth.canchild.ca/en/GMFM/gmfmscoresheets.asp
FORMATION	OUI
NOMBRE D'ITEMS	88
DUREE DE LA PASSATION	
COTATION	0-1-2-3
VERSIONS DISPONIBLES VALIDEES	ANGLAIS, FRANCAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Echelle créée et validée pour les enfants présentant une Paralyse Cérébrale Score ajusté pour la population ASI qui élimine 10 items nécessitant une position en procubitus ventral impossible en raison d'un flexum de hanche + scoliose dans cette population
METHODOLOGIE DE VALIDATION	Etude de la fiabilité inter et intra observateur Etude de la validité contre critère
VALIDITE CONTRE CRITERE	Corrélation avec QMT ($r=0.63$ (D)(E)- 0.86 (A), $p>0,0001$)
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	Kappa 0,72
FIDELITE INTRA OBSERVATEUR	ICC= $0,96-0,98$ (95% CI)
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	45-60 min de passation annoncé dans les articles princeps de développement de l'échelle mais pas de moyenne chez les patients porteurs d'ASI
REFERENCES	[69][70][104] [114]

Tableau 13 : Gross motor function measure

HAMMERSMITH FUNCTIONAL MOTOR SCALE

HFMS

TYPE D'OUTIL : ECHELLE DE MESURE

CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION (AGE)	ENFANT (>30 MOIS-15ANS)
PATHOLOGIES	ASI II & III
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Feuille de cotation : http://www.treat-nmd.eu/downloads/file/oms/ominfo/ROM7.1.OMP.0001.pdf Manuel d'utilisation : http://www.treat-nmd.eu/downloads/file/oms/ominfo/ROM7.2.OMP.0002.pdf
FORMATION	OUI (APPARUE DANS UN 2EME TEMPS)
NOMBRE D'ITEMS	20 (CLASSES PAR ORDRE DE DIFFICULTE)
DUREE DE LA PASSATION	
COTATION	0-1-2 (SAUF 2 ITEMS COTES 0-2)
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Choix des items : Expérience clinique des auteurs. Première étude sur des enfants sains >29 mois + 58 patients SMA Choix des critères d'âges de passation : >90% des enfants normaux de plus de 30 mois ont un score maximal: Ordre de difficultés des items : réussite aux items des 58 enfants avec SMA type 2 et 3
METHODOLOGIE DE VALIDATION	13 et 35 enfants (patients ou enfants sains ?) évalués par 2 cotateurs : validité inter observateur 90 patients ASI type II et III non marchants évalués à 0 (90), 3 mois (66), 6 mois (24) : validité intra observateur
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	Coefficient de corrélation (méthode non explicitée) moyenne = 0,95 Accord inter-observateur > 99% (34/35)
FIDELITE INTRA OBSERVATEUR	68% n'ont aucune variation des scores entre 0 et 3 mois. 37.5% n'ont aucune variation des scores entre 0 et 6 mois et 37.5% ont une variation de moins d'1 point.
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	Phenylbutyrate dans l'ASI [99] TROPHOS : Olesoxime (TR019622) dans l'ASI (en cours)
COMMENTAIRES	La méthodologie de validation inter et intra observateur est très discutable. 10-15 minutes annoncé par les auteurs mais absence de moyenne dans les publications Manuel d'utilisateur publié en 2009 alors que l'échelle date de 2003
REFERENCES	[91][98][99]

Tableau 14 : Hammersmith functional motor scale

HAMMERSMITH MOTOR ABILITY SCALE

HMAS

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)
CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION / AGE	ENFANTS
PATHOLOGIES	DMD
STATUT FONCTIONNEL	Patients marchants
DESCRIPTION	
FORMATION	Non
NOMBRE D'ITEMS	20
DUREE DE LA PASSATION	
COTATION	0-1-2
VERSIONS DISPONIBLES VALIDEES	Anglais
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	
METHODOLOGIE DE VALIDATION	Etude prospective sur 3 ans
VALIDITE CONTRE CRITERE	Force musculaire (MRC) : $r=0,83$; $n=61$
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	$r=0,98$, $n=40$
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	

REFERENCES [120]

Tableau 15 : Hammersmith motor ability scale

MESURE DE FONCTION MOTRICE

MFM-32

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

CE QUI EST MESURE : CAPACITES MOTRICES EN 3 DOMAINES D1 POSITION DEBOUT ET TRANSFERTS ; D2 MOTRICITE AXIALE ET PROXIMALE ET D3 MOTRICITE DISTALE

POPULATION / AGE	ENFANTS ET ADULTES DE 6 A 60 ANS
PATHOLOGIES	MNM
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Feuille de cotation et manuel : http://www.mfm-nmd.org/le-manuel-utilisateur.aspx Matériel nécessaire pour passation : chronomètre, CD collé sur une feuille, 10 pièces de monnaie, une balle de tennis, un crayon ; une feuille A4 (70gr), une feuille cartonnée
FORMATION	OUI
NOMBRE D'ITEMS	32
DUREE DE LA PASSATION	36 MINUTES
COTATION	0-1-2-3
VERSIONS DISPONIBLES VALIDEES	FRANÇAIS, ANGLAIS, NEERLANDAISE, ESPAGNOLE, PORTUGAISE, ALLEMANDE, ITALIENNE, ARABE
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Création d'une première échelle à 75 items à partir de la GMFM+expérience des investigateurs+étude pilote Réduction du nombre d'items : 51 items puis 32 items en enlevant les items concernant l'évaluation de la face ayant une faible reproductibilité ainsi que d'autres items très corrélés entre eux ou mal tolérés par certains sujets
METHODOLOGIE DE VALIDATION	1/Etude de validation sur 303 patients: cohérence interne, fidélité inter et intra observateur, validité contre critère (clinical global impression, Vignos, Brooke, MIF) 2/Etude de validation sur 100 patients adultes (++) Steinert et FSH)
VALIDITE CONTRE CRITERE	1/Corrélation Spearman pour le score total: CGI médecin (0,88), CGI kiné (0,91), Vignos (-0,91), Brooke (-0,85), MIF (0,91) 2/Corrélation Spearman pour le score total : différents testing rachis(0,63), face(0,05), Membre supérieur (0,78) et inférieur (0,8), Vignos (-0,8), Brooke (-0,85)
AUTRES VALIDITE DE CONSTRUIT	Analyse en composantes principales : 1/Identification des 3 dimensions expliquant 75% de la variance 2/ Testing des membres et scores fonctionnels =64% de l'information totale
COHERENCE INTERNE	Cohérence interne =coefficient de Cronbach entre 0,89 et 0,98 pour les dimensions et 0,99 pour le score total
FIDELITE INTER OBSERVATEUR	ICC=0,96 (95%CI)
FIDELITE INTRA OBSERVATEUR	ICC=0,99 (95%CI)
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	ASIRI : Rilusol dans l'ASI (non publié); Trophos : Olesoxime dans l'ASI (en cours) ; PMP 22 (DMD)
COMMENTAIRES	
REFERENCES	[13][14][15][139]

Tableau 16: mesure de fonction motrice-32

MESURE DE FONCTION MOTRICE

MFM-20

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

**CE QUI EST MESURE : CAPACITES MOTRICES EN 3 DOMAINES D1 POSITION DEBOUT ET
TRANFERTS ; D2 MOTRICITE AXIALE ET PROXIMALE ET D3 MOTRICITE DISTALE**

POPULATION / AGE	ENFANTS 2-7 ANS
PATHOLOGIES	MALADIES NEUROMUSCULAIRES
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	Feuille de cotation et manuel : http://www.mfm-nmd.org/le-manuel-utilisateur.aspx Matériel nécessaire pour passation : chronomètre, CD collé sur une feuille, 10 pièces de monnaie, une balle de tennis, un crayon ; une feuille A4 (70gr), une feuille cartonnée
FORMATION	OUI
NOMBRE D'ITEMS	20
DUREE DE LA PASSATION	26 MINUTES (+/- 8,5 MINUTES) [12-50 MINUTES]
COTATION	0-1-2-3
VERSIONS DISPONIBLES VALIDEES	FRANÇAIS, ANGLAIS, NEERLANDAISE, ESPAGNOLE, PORTUGAISE, ALLEMANDE, ITALIENNE, ARABE
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Réduction du nombre d'item à 20 après élimination des 12 items difficiles (plus de 20% des enfants n'ont pas le score maximum) ou impossibles à réaliser chez 190 enfants sains entre 2 et 6 ans
METHODOLOGIE DE VALIDATION	Etude de validation (88 patients): cohérence interne, fidélité, validité contre critère (Vignos,, Brooke, clinical global impression, échelle visuelle analogique)
VALIDITE CONTRE CRITERE	Corrélation Spearman pour le score total: CGI kiné et EVA (0.74-0.79), D1 vs Vignos (0,86), D2 vs Brooke (0,69)
AUTRE VALIDITE DE CONSTRUIT	Analyse factorielle : seul 3 items (5, 21, 23) ne se retrouvent pas dans le domaine initial de fonction motrice Différences significatives entre le score total moyen entre groupes (diagnostics $p < 0,001$, sévérité évaluée par le médecin (CGI) $p < 0,0001$)
COHERENCE INTERNE	Cohérence interne =coefficient de Cronbach> 0.9 pour D1 et D2 et =0.69 pour D3
FIDELITE INTER OBSERVATEUR	34 patients : kappa entre 0.37 et 0.94 sauf item 14 (0.22) ICC=0,99 (95% CI) Total score et ICC=0,91-0.99 (95% CI) D1, D2, D3
FIDELITE INTRA OBSERVATEUR	17 patients : kappa entre 0.56 et 1.0 ICC=0,99 (95% CI) Total score et ICC=0,91-0.99 (95% CI) D1, D2, D3
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[36]

Tableau 17: Mesure de fonction motrice-20

MODIFIED HAMMERSMITH FUNCTIONAL MOTOR SCALE MHFMS

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION (AGE)	ENFANT (1-18 ANS)
PATHOLOGIES	AMYOTROPHIE SPINALE INFANTILE II & III
STATUT FONCTIONNEL	PATIENTS NON MARCHANTS
DESCRIPTION	
FORMATION	NON
NOMBRE D'ITEMS	20
DUREE DE LA PASSATION	15-30 MIN
COTATION	0-1-2
VERSIONS VALIDEES	DISPONIBLES ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Modifications de l'HFMS pour qu'elle soit utilisable en multicentrique de façon longitudinale chez les petits enfants de -de 30 mois grâce à l'avis d'experts pour la validité de contenu.
METHODOLOGIE DE VALIDATION	13 ASI type 2 SMA (age 2.2–9.7ans) évalués 2 fois (2 jours de suite) par le même évaluateur = validité intra observateur 14 vidéos évaluées par 4 cotateurs = validité inter observateur 22 ASI type II (age 9-30 mois) évalués 2 fois à 2,9 mois d'intervalle = validité intra observateur 25 enfants sains (age 9-30 mois évalués une seule fois = calcul de l'âge limite (inférieur) de réussite de l'échelle
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	ICC=0,953 (95%CI)
FIDELITE INTRA OBSERVATEUR	ICC=0,986 (95%CI) ICC = 0.96, SEM 1.86
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	SMA CARNI-VAL : L carnitine + ac valproate dans l'ASI [76] Open label study of valproic acid : Valproate dans l'ASI [126]
COMMENTAIRES	Attention autre nom utilisé = extended version
REFERENCES	[79][80]

Tableau 18: Modified Hammersmith functional motor scale

MUSCULAR DYSTROPHY FUNCTIONAL RATING SCALE

MDFRS

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE COMPOSITE (TEST ET QUESTIONNAIRE)

CE QUI EST MESURE : MOBILITE, PERFORMANCES DANS LES ACTIVITES DE LA VIE QUOTIDIENNE, DEFICIENCES (FORCE MUSCULAIRE, CONTRACTURES, SCOLIOSE, SYMPTOMES RESPIRATOIRE) ET FONCTION MOTRICE DES MEMBRES SUPERIEURS

POPULATION / AGE	ENFANTS ET ADULTES
PATHOLOGIES	DYSTROPHIES MUSCULAIRES
STATUT FONCTIONNEL	PATIENTS MARCHANTS ET NON MARCHANTS
DESCRIPTION	
FORMATION	NON
NOMBRE D'ITEMS	33
DUREE DE LA PASSATION	
COTATION	1-2-3-4
VERSIONS DISPONIBLES VALIDEES	ANGLAIS, PORTUGUAIS, CHINOIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	53 items retenus à partir d'entretiens semi dirigés avec 25 patients, d'une revue de la littérature et d'avis d'experts Phase expérimentale sur 85 patients permettant d'aboutir à une version à 33 items répartis en 4 domaines : MOBILITE, PERFORMANCES DANS LES ACTIVITES DE LA VIE QUOTIDIENNE (AVQ), DEFICIENCES ET FONCTION MOTRICE DES MEMBRES SUPERIEURS
METHODOLOGIE DE VALIDATION	Etude de validation sur 121 patients : fiabilité, validité de construit, validité contre critère et cohérence interne de l'échelle
VALIDITE CONTRE CRITERE	Coefficient Spearman entre 0.65 et 0.91 ($p < 0,001$) : indice de Barthel, Brooke et Vignos, force musculaire, fonction pulmonaire
AUTRE VALIDITE DE CONSTRUIT	Analyse factorielle confirme les 4 domaines de l'échelle Peu d'effet plancher et plafond
COHERENCE INTERNE	Cohérence interne : Coefficient de Cronbach entre 0,84 et 0,97
FIDELITE INTER OBSERVATEUR	ICC = 0.99
FIDELITE INTRA OBSERVATEUR	Test -retest ICC=0.99
SENSIBILITE AU CHANGEMENT	Pour groupe DMD (31 patients) SRM (0.51 à 0.97) pour mobilité, déficience et fonction des membres supérieurs et SRM=0.37 pour activité de la vie quotidienne Pour groupe DMB, FSHD, LGMD (12 patients) SRM (0.39 à 0.84)
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	5 à 20 min de passation selon le statut fonctionnel mais pas de moyenne dans la population
REFERENCES	[85]

Tableau 19: Muscular dystrophy functional rating scale

NORTH STAR AMBULATORY ASSESSMENT

NSAA

TYPE D'OUTIL : ECHELLE DE MESURE (TEST)

CE QUI EST MESURE : CAPACITES MOTRICES GLOBALES

POPULATION / AGE	ENFANTS (2-15 ANS)
PATHOLOGIES	DMD
STATUT FONCTIONNEL	PATIENTS MARCHANTS
DESCRIPTION	Instructions et feuille de cotation : http://www.muscular-dystrophy.org/assets/0000/6388/NorthStar.pdf
FORMATION	OUI
NOMBRE D'ITEMS	17
DUREE DE LA PASSATION	
COTATION	0-1-2
VERSIONS DISPONIBLES VALIDEES	ANGLAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Groupe de travail constitué de kinésithérapeutes anglais : Adaptation de l'HFAS + quelques items supplémentaires pour patients DMD sous corticoïdes:
METHODOLOGIE DE VALIDATION	Etude de la fidélité inter et intra-observateur : 13 cotateurs ont participé à deux sessions de formation à l'utilisation de la NSAA. La fidélité inter-observateur a été estimée après chacune des sessions à partir de cotations réalisées sur vidéos. Une nouvelle cotation demandée 1 mois à chacun des cotateurs a permis de calculer une fidélité intra observateur -Etude de la sensibilité au changement (106 patients)
VALIDITE CONTRE CRITERE	Corrélation Spearman : $r=0.505$ $p<0.01$ (temps marche/course 10 m), $r=0.701$ $p<0.01$ (temps pour se relever du sol) Corrélation de Spearman : $r=0.52$ $p<0.001$ (6MWT)
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	ICC=0,995(95% CI)
FIDELITE INTRA OBSERVATEUR	ICC=0,95(95% CI)
SENSIBILITE AU CHANGEMENT	Le changement moyen des scores NSAA sur 1 an : -2.2 (+/- 3.7) points
ESSAIS THERAPEUTIQUES	DMD114117 : GSK2402968 chez les DMD >5ans marchants
COMMENTAIRES	Possibilité de chronométrer certains tests (10 m marche, course, se relever du sol) 10 min maximum 15 minutes annoncé comme temps de passation mais pas de moyenne dans les publications Existence du NSAA for SMA [119]: modification du NSAA pour être applicable dans l'ASI
REFERENCES	[42][94][95][97][119]

Tableau 20 : North star ambulatory assessment

TEST OF INFANT MOTOR PERFORMANCE TIMP

TYPE D'OUTILS DE MESURE : ECHELLE DE MESURE DE FONCTION MOTRICE (OBSERVATION ET TESTS)

CE QUI EST MESURE : CAPACITES FONCTIONNELLES MOTRICES GLOBALES

POPULATION (AGE)	ENFANT
PATHOLOGIES	ASI I (FAIBLESSE MUSCULAIRE SEVERE)
STATUT FONCTIONNEL	PATIENTS NON MARCHANTS
DESCRIPTION	
FORMATION	OUI
NOMBRE D'ITEMS	42 (13 + 29)
DUREE DE LA PASSATION	
COTATION	0-1(13 ITEMS COTES A PARTIR DE L'OBSERVATION) 0-1-2-3-4-5-6 (29 ITEMS COTES EN SITUATION DE TESTS)
VERSIONS DISPONIBLES VALIDEES	ANGLAIS, FRANÇAIS, PORTUGAIS
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Echelle mise au point pour une population d'enfants à partir de 32SA pour suivi jusqu'à 4 mois (jusqu'au retournement) afin de dépister d'éventuels atteintes motrices.
METHODOLOGIE DE VALIDATION	Etude de validation chez bébés tout venant entre 34SA et 4 mois Etude de validation dans une population d'enfants présentant une ASI de type I Validité inter observateur : 5 vidéos et 18 cotateurs Validité intra observateur : 11 enfants avec SMA type I évalués par 8 cotateurs
VALIDITE CONTRE CRITERE	
AUTRE VALIDITE DE CONSTRUIT	Entre 32SA et 4 mois : Pas d'effet plancher ni plafond, sensible aux changement de niveau fonctionnel ainsi que discriminant pour différencier les nouveaux nés à risque des autres. Non étudié chez MNM
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	Après entraînement à partir de vidéos: Kappa = 0.61 (95% CI 0.58-0.64)
FIDELITE INTRA OBSERVATEUR	Kappa = 0.71 (95% CI 0.68-0.75)
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[48]

Tableau 21: Test of infant motor performance

UPPER LIMB MODULE

UL MODULE

TYPE D'OUTILS :ÉCHELLE DE MESURE (TEST)	
CE QUI EST MESURE : CAPACITES MOTRICES DES MEMBRES SUPERIEURS	
POPULATION / AGE	ENFANT (A PARTIR DE 30 MOIS)
PATHOLOGIES	ASI 2 ET 3
STATUT FONCTIONNEL	NON MARCHANT
DESCRIPTION	Matériel nécessaire pour passation :une cannette boisson 33 cl, une pièce de monnaie, un stylo, une feuille pré imprimée, poids avec 500gr de sable, un verre en plastique, une boite type Tupperware, poids avec 1x 50 g et 2x 100gr dans sac plastique, lampe électrique avec bouton on/off et feuille de papier avec 2 cercles dessinés
FORMATION	NON (MAIS MANUEL D'UTILISATEUR ET FEUILLE DE COTATION)
NOMBRE D'ITEMS	9
DUREE DE LA PASSATION	5-10 MINUTES
COTATION	0-1-2
VERSIONS DISPONIBLES VALIDEES	ANGLAISE
QUALITES METROLOGIQUES	
METHODOLOGIE DE CONSTRUCTION	Sélection de 20 items par un panel de professionnels à partir de plusieurs échelles disponibles dans la littérature (Griffiths developmental scales, Bayley, TIMP, AIMS, Melbourne, GMFM, Mov ABC, JHFT et MFM, HFMS, HFMSSE et moHFMS, vignos, EK-scale) et de leur expérience auprès de patients ASI Réduction de 20 à 15 items : Items testés sur 25 enfants ASI 2 et 3 (28mois-15ans) Réduction de 15 items à 9 items et amélioration de la description des items et de la façon de les coter : groupe de travail constitué de 5 kinésithérapeutes investis dans TREAT NMD
METHODOLOGIE DE VALIDATION	Etude de validation 40 enfants sains d'age préscolaire (30-48mois) : afin de déterminer l'âge limite de passation de l'échelle 45 ASI non ambulant (40 type 2 et 5 type 3) : validation contre critère 12 ASI non ambulant : Validité inter observateur
VALIDITE CONTRE CRITERE	Score plus variables dans la population d'enfants malades que sains Non liés à l'âge (r=-0,19) Coefficient de Pearson avec HFMS = 0,75 (p<0,0001)
AUTRE VALIDITE DE CONSTRUIT	
COHERENCE INTERNE	
FIDELITE INTER OBSERVATEUR	ICC(12 enfants) = 0,97
FIDELITE INTRA OBSERVATEUR	
SENSIBILITE AU CHANGEMENT	
ESSAIS THERAPEUTIQUES	
COMMENTAIRES	
REFERENCES	[96]

Tableau 22 : UpperLimb module

A.4 CONCLUSIONS DE LA REVUE DE LA LITTÉRATURE

Cette revue de la littérature a permis d'identifier et de décrire précisément 19 outils d'évaluation fonctionnelle motrice utilisés chez les patients porteurs d'une MNM.

Une première revue de la littérature avait été réalisée par Sandrine Ginvarc'h en 1998 [61], au préalable de la construction et validation de la MFM, et avait conclu à la nécessité de développer un nouvel outil d'évaluation de la fonction motrice.

Depuis le début de cette revue de la littérature (2007), beaucoup de progrès ont été réalisés dans la validation de ces outils en particulier grâce aux réseaux comme Treat-NMD. Les différentes équipes impliquées dans le développement de ces outils ont pu se rencontrer et ont pris conscience de l'importance de la rigueur de validation, préambule indispensable à l'utilisation des outils dans les essais thérapeutiques [19][23][100].

Les scores fonctionnels (« Functional Rating scale ») sont les premiers à avoir été développés dans le domaine des MNM ; il s'agit du score de Vignos pour la classification fonctionnelle au niveau des membres inférieurs et du score de Brooke pour les membres supérieurs. Les études de validation les concernant sont absentes des articles princeps. Les fonctions motrices étudiées dans ces 2 scores sont limitées ; aucun score n'est disponible pour donner une évaluation rapide, reproductible de la fonction motrice globale des patients porteurs d'une MNM dans les principaux secteurs de la fonction motrice.

Ces scores fonctionnels manquent de précision et de sensibilité au changement ; le changement de fonction motrice nécessaire pour passer d'un niveau à un autre étant trop important [87][124]. Ces outils sont surtout intéressants en recherche clinique pour une description de la population d'étude ou pour servir de référence dans des études de validation d'outils de mesure, mais sont insuffisamment précis pour être utilisés seuls dans un essai thérapeutique pour évaluer les éventuels changements de la fonction motrice des patients.

Le score de Vignos [136] fournit une évaluation des performances motrices des membres inférieurs par la description des possibilités de déambulation [82][83]. Le reproche fait le plus couramment à cette classification dans la DMD est que la plupart des patients ne passe pas par tous les niveaux en particulier les niveaux 6, 7, 8 car l'utilisation d'attelles longues de marche est actuellement rare dans cette pathologie [84]. Les fidélités inter et intra-observateur sont par ailleurs satisfaisantes [51].

Le score de Brooke [24][25], dont la construction a été inspirée des travaux de Vignos, est une évaluation des capacités fonctionnelles motrices des membres supérieurs[82][83]. Il est constitué de 6 épreuves de passation rapide, ne nécessitant qu'un verre d'eau, un crayon et une pièce de monnaie.

Lue YJ et al [84] rapportent dans un article récent l'insuffisance de ces scores fonctionnels pour juger de l'évolution de dystrophinopathies lentement progressives (FSH, Dystrophie musculaire de Becker ou dystrophie musculaires des ceintures) du fait d'une validité discriminante insuffisante et d'un effet plancher trop large.

La grande majorité des outils retenus dans cette revue de la littérature sont des échelles de mesure quantitative des capacités motrices en situation de test (CHOP INTEND, HFMSE, GMFM, HFMS, HMAS, NSAA, MFM 32, MFM 20, MHFMS, TIMP et UL Module).

A partir de la passation d'un certain nombre d'items, un score global (+/- par domaine) correspondant à la fonction motrice globale (par domaine) du patient peut être calculé. Seul l'UL module ne s'intéresse qu'à la fonction motrice des membres supérieurs. Avec la MFM, il s'agit des seules échelles avec des items s'intéressant à la motricité distale. Aucune échelle fonctionnelle motrice ne contient des items concernant la motricité fonctionnelle de la face alors que l'atteinte faciale est présente dans plusieurs pathologies (Dystrophie myotonique de Steinert (DMS), DFSH). Benaim et al [13] au sujet de la MFM, rapportent l'absence de corrélation significative entre les scores MFM et le testing musculaire de la face dans une population d'adultes atteint de DMS et de DFSH. Les auteurs recommandent donc l'utilisation du testing musculaire en plus des échelles fonctionnelles pour monitorer l'évolution de la motricité de la face.

Certaines échelles mesurent plutôt les performances des patients, en particulier l'indépendance fonctionnelle et l'autonomie, dans les activités de la vie quotidienne (ACTIVLIM, EK scale, ABILHAND et MDFRS).

Les échelles dites composites mesurent plusieurs entités différentes : des capacités comme la fonction motrice, des performances comme l'autonomie dans la vie quotidienne, mais aussi parfois des déficiences (les capacités respiratoires, le langage...). Il s'agit de la WeeFIM, de l'EK-Scale et de la MDFRS.

La population cible de ces échelles est parfois réduite à une seule pathologie : la DMD (HMAS et NSAA) ou l'ASI (HFMS, HFMSE, MoHFMS, CHOP INTEND et TIMP), ou à un statut fonctionnel particulier : marchant (NSAA, HMAS et HFMSE) ou non marchant (TIMP, MoHFMS, EK-Scale et CHOP INTEND). Les échelles utilisables uniquement chez le sujet marchant n'autorisent pas le suivi longitudinal des patients quand ceux-ci perdent la marche au cours de l'étude. La multiplication des échelles de mesure impose aux professionnels une maîtrise de plusieurs outils en fonction de la population cible ce qui augmente le temps de passation et le risque d'erreur de cotation [16]

Le temps de passation est une donnée manquante pour beaucoup d'outils ; des temps de passation sont annoncés dans le « Registry of outcome measures » mais aucune moyenne sur la population n'est donnée dans les publications princeps. Pourtant, il s'agit d'un élément important à prendre en compte surtout lors d'un essai thérapeutique chez des patients fatigables.

La propriété métrologique de loin la plus étudiée dans ces échelles est la fidélité et en particulier la fidélité intra observateur évaluée le plus souvent par l'ICC (avec ou sans intervalle de confiance), toujours supérieur à 0,90. Le kappa, plus sévère que l'ICC est très rarement utilisé (MFM 20, TIMP et GMFM). Il est alors plutôt de l'ordre de 0,70. L'étude de la cohérence interne des échelles, est rarement retrouvée sauf pour la MFM et ACTIVLIM.

La façon dont les concepteurs ont développé l'outil est importante à connaître car une mauvaise méthodologie de développement peut remettre en cause la validité d'un outil. On citera par exemple l'HFMSE développée en associant des items de l'HFMS et de la GMFM alors que la cotation des items de l'HFMS se fait en 3 points et la GMFM en 4 points. Une autre méthode discutable consiste, par exemple à sélectionner, des items dans des échelles existantes pour constituer un nouvel outil et utiliser par la suite ces mêmes échelles pour la validation contre critère de cet outil (ABILHAND, HFMSE et UL Module).

Les autres propriétés métrologiques des échelles sont très insuffisamment étudiées, en particulier il y a peu d'étude de sensibilité au changement [16]. Il paraît pourtant capitale de s'assurer que ces échelles ont une sensibilité au changement suffisante pour ne pas conclure à tort à une absence d'efficacité du traitement dans un essai thérapeutique. Enfin, la validité de construit constitue certainement la validité la plus compliquée à analyser et elle s'affirme avec le temps au fur et à mesure de la confirmation des hypothèses préliminaires. La validité de critère est l'analyse la plus fréquemment retrouvée et est alors volontiers analysée en calculant les coefficients de corrélation avec le plus souvent la force musculaire, la CGI (clinical global impression) ou d'autres échelles de mesure de fonction motrice.

Enfin, des modèles issus de la théorie de réponse aux items (TRI) comme l'analyse selon le modèle de Rasch commencent à être appliqués à ce type d'échelle (ACTIVLIM et ABILHAND), mais de façon encore insuffisante.

Concernant les tests chronométrés, ils sont souvent utilisés en pratique clinique mais finalement, le seul outil retrouvé dans la littérature est le 6MWT. Il existe des guidelines concernant son utilisation mais chez l'adulte essentiellement [10]. Les enfants peuvent, eux, être moins enclins à suivre les mêmes instructions pendant 6 minutes (par exemple, ils peuvent courir, sauter, ou continuer à marcher alors qu'on leur a demandé de s'arrêter) [113]. Le manque de motivation pour réaliser l'exercice et la mauvaise compréhension des consignes peuvent également affecter les capacités des enfants durant la passation du 6MWT [53]. Les difficultés cognitives et troubles du comportement fréquemment retrouvés chez les patients DMD constituent une limite notable à l'utilisation de ce test. Enfin, les garçons atteints de DMD ont tendance à tomber fréquemment surtout à proximité de la perte de marche [108] ce qui est aussi responsable d'une diminution de la précision de la mesure et parfois d'une sécurité insuffisante pour les enfants. Enfin, un des inconvénients majeurs de ces tests chronométrés reste qu'ils ne s'adressent qu'aux patients marchants ou capables de réaliser les activités motrices demandées. Ils ne permettent donc pas de suivre le patient durant toute l'évolution de sa maladie. Scott et al. rapportent une mauvaise reproductibilité test-retest du fait des erreurs possibles de mesure du temps, même si les positions de départ et d'arrivée sont parfaitement définies [118]. Mac Donald et al [88][89] proposent une version modifiée du 6MWT spécifiquement pour les enfants présentant une DMD.

En résumé, les grandes lignes de cette revue de la littérature sont :

- **La MFM a été développée selon une méthodologie clairement explicitée avec une première étude de validation sur 303 patients ayant permis de conclure à de bonnes propriétés métrologiques (cohérence interne, validité de construit, validité contre critère et fiabilité inter et intra observateur). Elle est la seule échelle de mesure de la fonction motrice globale utilisable chez l'enfant et l'adulte avec tous types de MNM quel que soit le statut fonctionnel du patient (marchant ou non marchant) :**
- **Il est nécessaire de s'assurer de la sensibilité au changement de ces outils avant de conclure à leur utilisation possible dans les essais thérapeutiques ;**
- **Il faut développer l'application des modèles type théorie de réponse aux items pour l'analyse de la validité de construit et de la cohérence interne de ces outils ;**
- **Il manque une classification fonctionnelle s'intéressant à la fonction motrice globale (c'est-à-dire dans les 3 domaines de la fonction motrice) du patient porteur d'une MNM.**

B ETUDES DE VALIDATION DE LA MFM : LA SENSIBILITE AU CHANGEMENT

Afin de compléter l'étude de validation de la MFM, et en particulier pour s'assurer de la capacité de cet outil à mesurer des changements cliniquement perceptibles de fonction motrice, plusieurs études de sensibilité au changement ont été conduites dans différentes populations de patients.

B.1 LA SENSIBILITE AU CHANGEMENT SUR LA POPULATION GLOBALE

Dans cette étude, l'objectif principal était d'étudier la sensibilité au changement de la MFM à partir d'un effectif important (> 150 patients). Nous avons ainsi inclus des patients, enfants et adultes, issus de la population générale des patients porteurs d'une MNM. Les MNM sont pour beaucoup des maladies évolutives, même si cette évolutivité est hautement variable selon la maladie, voir même entre patients au sein d'une même maladie. L'objectif était donc de montrer qu'à 1 an d'intervalle entre 2 mesures il existait bien une diminution significative des scores MFM moyens sur la population totale et dans les pathologies les plus évolutives. De plus, nous voulions étudier la corrélation entre l'évolution des scores MFM et l'appréciation subjective des patients ou de leur médecin concernant l'évolution fonctionnelle (stabilisation, aggravation, amélioration). La méthodologie utilisée, de type « Anchor-based approach » [37][62][111] est complémentaire des méthodes habituelles pour mesurer la sensibilité au changement en prenant en compte les différences cliniquement perceptibles (« minimally clinically-meaningful differences ») et non seulement les différences statistiquement significatives (« statistically-significant differences »).

Les résultats de cette étude ont été soumis dans la revue Archives of Physical Medicine and Rehabilitation et est en révision (Annexe 2) :

Responsiveness of the Motor Function Measure in neuromuscular diseases

Carole Vuillerot, Christine Payan, Françoise Girardot, Jacques Fermanian, Jean Iwaz, René Ecochard, Carole Bérard, and the MFM study group.

Objectives - Study the responsiveness (sensitivity to change) of the Motor Function Measure in detecting change in neuromuscular disease patients with the intent of using this Measure in future clinical trials.

Methods - Design: Prospective cohort observational study

Setting. Inpatient and outpatient facilities for follow-up and treatment of neuromuscular diseases.

Participants: Patients (152) with various neuromuscular diseases aged 6-60 years old.

Main Outcome measure(s): We used the Motor function Measure Total score and its four subscores on two measurements grossly one year apart. The physicians and the patients (or proxy) were asked to provide their perceived change in functional status since the first Motor Function Measure. These changes were expressed in three outcomes: deterioration, stability, or improvement

Results - The overall mean±SD annual total score change was -2.4 ± 5.5 points ($p<0.0001$), Duchenne patients presenting the most significant change (-5.8 ± 6.3 , $p<0.0001$). The change in patients reporting deterioration (34%) was significantly larger than that of those reporting stability (47%) or improvement (10%) (-4.4 ± 6.4 vs. -2.0 ± 5.6 and $+0.9\pm 4.4$ points, respectively; $p<0.01$). The annual total score changes were significantly greater in physician-rated deteriorated (49%) vs. stable patients (51%) with mean differences in scores -5.3 ± 7.6 and -1.2 ± 5.3 and, respectively ($p<0.001$).

Conclusions - The Motor Function Measure showed a good responsiveness, especially in Duchenne patients, and agreements with patients' and physician's perceived change. Confirming this responsiveness requires larger age-groups of patients with Duchenne and other neuromuscular diseases as well as disease-specific inter-examination delays.

En révision in Archives of Physical Medicine and Rehabilitation

B.2 LA SENSIBILITE AU CHANGEMENT PAR PATHOLOGIE

B.2.1 DANS LA DYSTROPHIE MUSCULAIRE DE DUCHENNE

Dans cette étude, la sensibilité au changement de la MFM a été étudiée dans la DMD, MNM la plus fréquente. La dégradation progressive de la fonction motrice dans cette maladie est bien connue et a été décrite par de nombreux auteurs en utilisant différentes mesures (force musculaire, fonction motrice, capacité vitale) [4][81][87]. Plusieurs essais thérapeutiques sont en préparation ou en cours justifiant l'intérêt porté depuis plusieurs années aux « outcomes mesures » dans cette pathologie, en particulier ceux mesurant la fonction motrice qui peuvent mettre en évidence une éventuelle efficacité thérapeutique [100][118].

L'objectif principal de l'étude était d'analyser le changement moyen des scores MFM (Score total, D1, D2 et D3) à court terme (3 mois) et à moyen terme (1 an) dans une population de patients DMD non traitée pour compléter l'histoire naturelle de la maladie. Nous avons aussi tenté de préciser les scores MFM moyens (Score total et D1) 1 an avant la perte de marche. Enfin, nous avons comparé l'évolution de la fonction motrice de 12 patients traités par corticoïdes à long terme à 12 patients contrôle.

Les résultats de cette étude ont été publiés dans la revue Dev Med Child Neurol (Annexe 3) :

Monitoring changes and predicting loss of ambulation in Duchenne Muscle Dystrophy with the Motor Function Measure

Carole Vuillerot, Françoise Girardot, Christine Payan, Jacques Fermanian, Jean Iwaz, Capucine de Lattre, Carole Bérard.

Aim - To assess changes in motor function in males with Duchenne muscular dystrophy using the Motor Function Measure.

Methods - A three-month and a one-year study included only physiotherapy-treated patients: 13 of mean age 11y 7mo (SD 1y 10mo) and 41 of mean age 14y 1mo (SD 5y 5mo), respectively. A third study compared 12 steroid-treated to 12 age- and motor-function-matched untreated patients (mean age of treated ones 10y 2 mo, SD 2y 2mo).

Results -. Over three months, the MFM D1 subscore (standing and transfers) decreased significantly (-4.7%; $p < 0.01$). Over one year, all MFM subscores decreased significantly: -4.9% for D1 ($p < 0.01$), -7.7% for D2 (axial and proximal motor capacity) ($p < 0.01$); -4.3% for D3 (distal motor capacity) ($p = 0.03$), and -5.8% for the Total score ($p < 0.01$). A threshold value for loss of ambulation and a predictive value one year before that loss were estimated (around Total score 70 and D1 subscore 40). Compared to the controls, steroid-treated patients had more stable Total scores (-0.59 vs. -5.87; $p = 0.02$) and D2 subscores (+0.98 vs. -8.50; $p < 0.01$).

Interpretations - These results foster the use of the MFM in everyday patient management to prepare for loss of ambulation and in clinical trials to follow-up patients receiving various treatments.

Dev Med Child Neurol. 2010 Jan;52(1):60-5. Epub 2009 Apr 22

B.2.2 DANS L'AMYOTROPHIE SPINALE INFANTILE

Dans cette étude, nous avons étudié l'applicabilité et la sensibilité au changement de la MFM dans l'ASI. Compte tenu de la différence importante de phénotypes dans l'ASI en fonction du type (I, II ou III), un outil comme la MFM, applicable quel que soit le statut fonctionnel des patients, est particulièrement intéressant.

Devant la multiplication actuelle des essais thérapeutiques dans l'ASI, il nous a paru indispensable de publier une étude d'histoire naturelle de la MFM dans cette pathologie. Les agences du médicament manquent, en effet, de publications pour décider des meilleurs outils et les industriels ont, quant à eux, tendance à privilégier les outils déjà utilisés dans des essais antérieurs et manquent de données adaptées à leur projet. Des données de MFM chez 112 patients ont ainsi été collectées dans 19 centres et ces données ont été étudiées dans chaque type d'ASI. La sensibilité au changement a été analysée dans chaque type d'ASI en fonction des domaines de la MFM (D1 pour ASI III et D2 pour ASI II). Un nombre de sujet nécessaire a pu être calculé à partir de nos résultats en fonction du sous type d'ASI, de l'effet attendu du traitement et la puissance de l'étude.

Les résultats de cette étude ont été soumis dans la revue *Dev Med Child Neurol* (Annexe 4) :

Use of the Motor Function Measure to Assess Motor Function in Patients with Spinal Muscular Atrophy

Carole Vuillerot, Christine Payan, Jean Iwaz, René Ecochard, Françoise Girardot, Carole Bérard, and the MFM Spinal Muscular Atrophy Study Group.

Some treatments of spinal muscular atrophy can slow the disease progression but few responsive and valid tools can measure their effects on the motor function. The Motor Function Measure seems appropriate. The relationships between this test scores and age were studied according to the disease type. In 112 patients (5.7 to 59 years), there was a moderate inverse relationship between age and the Total score in disease Types II and 3. Patients followed-up less than 6 months showed little score variations; the others showed slow deteriorations (-0.9 and -0.6 points/year for Types II and III, respectively). A substantial responsiveness was found with the proximal/axial subscore in Type II patients (Standardized Response Mean =1.29) and the standing/transfers subscore in Type III patients aged 10-15 years (Standardized Response Mean =0.94). The Motor Function Measure may thus be used to monitor disease progression in ambulant or non-ambulant patients participating to clinical trials.

Soumis à Dev Med Child Neurol

C LA CREATION D'UNE CLASSIFICATION EN GRADES DE SEVERITE DANS LE DOMAINE DES MNM DE L'ENFANCE : LA CLASSIFICATION NM-SCORE

C.1 JUSTIFICATIONS SCIENTIFIQUES DU PROJET

Proposer une échelle de fonction motrice à un patient suppose la disponibilité d'un thérapeute formé à la passation et un temps de passation en moyenne d'une demi-heure pour le patient quelle que soit l'échelle utilisée. De plus, en pratique clinique courante ainsi que dans certains projets d'étude, une évaluation aussi précise de la fonction motrice n'est pas nécessaire, une appréciation en niveaux ou grades de sévérité de l'atteinte fonctionnelle serait suffisante pour constituer des groupes de patients homogènes en termes de fonction motrice. Des scores fonctionnels existent dans la littérature, mais ils ne prennent pas en compte les différentes composantes de la fonction motrice (c'est-à-dire la position debout et les transferts, la motricité axiale/proximale et la motricité distale) ou ne sont pas suffisamment adaptés à la population. Ainsi, les scores de Brooke [24][25] et de Vignos [136], 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, les résultats de la revue de la littérature confirment que peu de données sont disponibles concernant la validation de ces outils.

Nous proposons une nouvelle classification ayant comme originalité une approche différente de la MFM. Il s'agit d'une approche plus qualitative de la fonction motrice, autorisant une classification des patients par analogie entre l'état fonctionnel du patient et la description clinique littérale correspondant à chaque grade de sévérité. Le patient est, pour chaque domaine de fonction motrice, assimilé à un grade de sévérité de la classification grâce à l'interrogatoire et l'observation clinique. Cette classification peut être assimilée aux grading utilisés en cancérologie pour typer les atteintes tumorales en fonctions de différents critères.

Nous souhaitons créer un outil dont l'objectif principal est une classification rapide et reproductible des patients porteurs d'une maladie neuromusculaire en 5 grades de sévérité. L'originalité de cette classification par rapport à une version réduite de MFM réside dans la description clinique des grades de sévérité par des items pertinents de manière qualitative. Il ne s'agit pas comme dans la MFM de faire passer des épreuves aux patients et d'en conclure un score mais plutôt de se fonder sur l'interrogatoire et l'observation des patients pour les classer dans un grade de sévérité selon leur performance dans leur milieu de vie. Cette nouvelle classification sera utile en pratique clinique comme en recherche clinique.

En s'inspirant des travaux de Palisano et al. [106] dans le domaine de la paralysie cérébrale, nous proposons de développer une classification en grades de sévérité d'atteinte fonctionnelle motrice pour chacun des 3 domaines de la MFM (D1 : debout et transferts, D2 : motricité axiale et proximale, D3 : motricité distale).

Comme imagée sur la figure 8, la problématique principale est le passage d'une échelle quantitative comme la MFM (en %) à une échelle qualitative de description d'un état fonctionnel (en niveau ou grade de sévérité). Il s'agit d'imaginer cliniquement un patient en fonction de son grade de sévérité fonctionnel et de réussir à décrire ce patient de manière la plus précise possible et intelligible de tous.

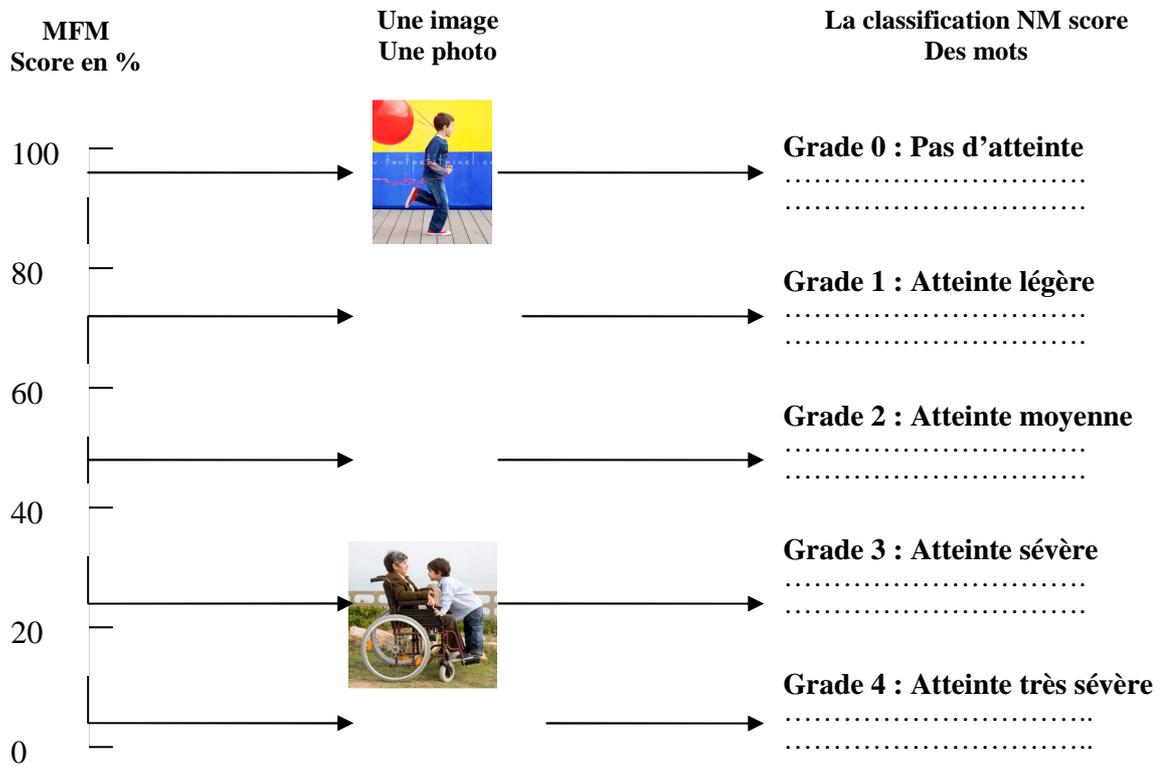


Figure 8 : Passage d'une échelle quantitative (MFM) à une description en grades de sévérité d'atteinte fonctionnelle motrice (NM-Score)

Une autre problématique est de bien distinguer les processus physiopathologiques et les mesures d'état, de manière à créer des schémas purs de contaminations entre les descriptions des domaines. Sur la figure 9, il s'agira de minimiser l'influence sur la mesure des différents domaines D1 D2 et D3 entre eux, symbolisée sur la figure par les traits en pointillés. De même, l'examineur ou le patient lui-même à l'autre extrémité du schéma ont une influence sur la mesure qu'il faudra étudier et prendre en compte.

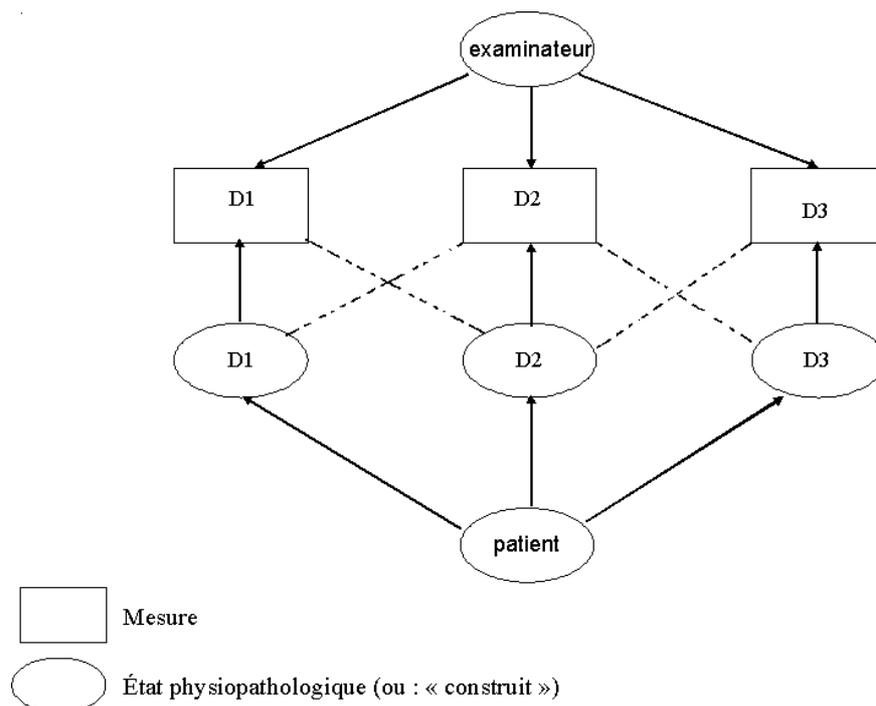


Figure 9 : Lien entre mesure et construit du NM-Score

C.2 ETUDE PRELIMINAIRE A PARTIR DES DONNEES DE L'ETUDE DE VALIDATION INITIALE DE LA MFM

Dans l'étude de validation initiale de la MFM portant sur 303 patients [14], parallèlement à la MFM ont été recueillies des données correspondant à l'impression clinique globale du patient (« clinical global impression » : CGI). Le médecin était interrogé sur la sévérité de l'atteinte du patient qu'il devait qualifier de légère = 1, de moyenne = 2, de sévère = 3 ou de très sévère = 4.

Les représentations graphiques des scores MFM en fonction de l'âge dans chacun des 4 groupes de sévérité (Figure 10) montre que, quel que soit l'âge des sujets, les scores MFM des patients se regroupent autour d'une valeur d'autant plus élevée que le niveau fonctionnel du patient est jugé bon.

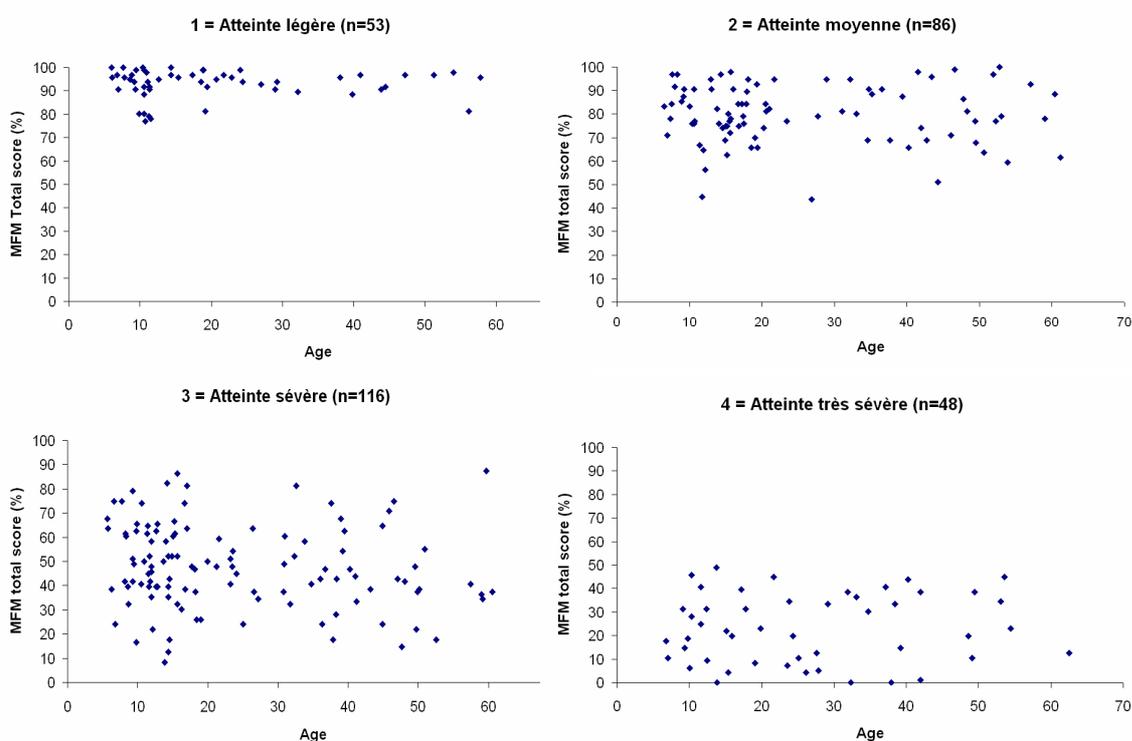


Figure10 : Répartition des patients en fonctions des scores MFM dans chaque groupe de sévérité (atteinte légère, moyenne, sévère et très sévère)

La dispersion des valeurs est d'autant plus réduite que l'atteinte est dite légère ou très sévère, aux extrêmes de la classification. La moyenne des scores MFM totaux est d'autant plus élevée que la CGI augmente (1 = 93,0 +/- 6,1 ; 2 = 79,9 +/-12,4 ; 3 = 47,7+/-17,3 ; 4 = 23,6 +/-14,6), avec un écart type élevé surtout pour les atteintes dites sévères dont la dispersion des valeurs autour de la moyenne est la plus grande (Figure11). Ces 4 groupes ne sont pas homogènes en terme de score total de la MFM (test de Kruskal Wallis, critère de regroupement = CGI, $p < 0,001$).

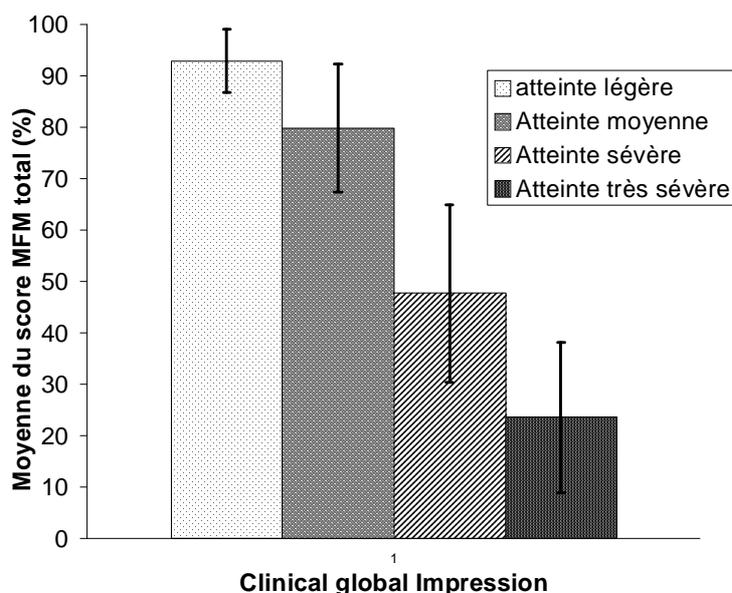


Figure 11 : Score moyen MFM (+ déviation standard) pour chaque groupe de sévérité (atteinte légère, moyenne, sévère et très sévère)

La conclusion de cette étude préliminaire sur les 303 patients de l'étude de validation de la MFM est qu'il semble possible, simplement par la qualification en léger, moyen, sévère et très sévère, de différencier plusieurs groupes de sévérité d'atteinte fonctionnelle motrice chez les patients porteurs d'une MNM. Les grades extrêmes de cette classification semblent naturellement les plus discriminants. Une meilleure description de chacun de ces groupes de sévérité permettrait une précision plus importante du classement, surtout pour les groupes intermédiaires où la variabilité des patients est la plus importante.

C.3 MATERIELS ET METHODES

L'objectif de ce travail est de créer une classification en grade de sévérité fonctionnelle motrice dans les 3 domaines de la MFM (D1, D2, D3). Cette classification, appelée NM-Score, doit être facilement réalisable en consultation, valide, fidèle et reproductible.

C.3.1 TABLEAU SYNOPTIQUE DE L'ETUDE

Dans le tableau ci-dessous (Tableau 23), apparaissent les différentes étapes de l'étude avec pour chacune un calendrier précis, la méthodologie utilisée ainsi que le nombre de patients et professionnels impliqués.

Phase	1	2	3	4
	Construction	Etude de validité de contenu	Etude de validité inter observateur	Etude de validité contre critère
Période	03/08-02/09	03/09-07/09	09/09-04/10	05/10-05/11
Classification NM-Score	Version provisoire 1 (VP1)	Version provisoire 2 (VP2)	Version finale 1 (VF1)	Version Finale 2 (VF2)
Méthodologie	Workshops	Delphi	Etude	Etude
Professionnels (n)	8	28	23	26
Patients (n)	48	169	71	158

Tableau 23 : Tableau synoptique de l'étude NM-Score

C.3.2 DEVELOPPEMENT DE LA CLASSIFICATION

C.3.2.1 Construction d'une première classification

L'objectif de cette première phase était de proposer une première classification (VP1).

Un groupe d'étude régional constitué de huit professionnels, 4 médecins, 3 kinésithérapeutes et d'1 ergothérapeute, a été créé. Quatre professionnels avaient une activité uniquement pédiatrique, 2 professionnels une activité adulte et 2 professionnels avaient une activité mixte, pédiatrique et adulte.

Les résultats de la revue de la littérature réalisée ont été diffusés au groupe d'étude comme base de discussion. Quatre réunions de travail de 2 heures chacune ont été organisées entre mars 2008 et février 2009. Lors de ces réunions des vidéos ont été visionnées et des dossiers médicaux analysés. Chaque professionnel exprimait son point de vue à partir de son expérience personnelle et en se basant sur la revue de la littérature.

Plusieurs étapes ont été nécessaires :

- 1. L'accord au sein du groupe sur le cahier des charges précis de la classification**
- 2. Le choix et la définition des critères dits majeurs pour classer les patients dans chacun des niveaux pour chaque domaine de la classification (D1, D2, D3)**
- 3. Une étude expérimentale réalisée par les professionnels du groupe d'étude sur les patients vus en consultation pour identifier des critères dits mineurs permettant d'améliorer la description des patients dans chaque domaine**

C.3.2.2 Etude de validation de contenu : Méthode Delphi

Afin d'analyser la validité de contenu de la classification proposée par le groupe de travail (NM-Score VP1) une méthode de consensus a été utilisée de manière à avoir l'avis et les commentaires de plusieurs experts francophones dans le domaine des MNM et de l'évaluation. L'objectif de cette phase était de vérifier la représentativité de l'ensemble des énoncés qui constituent le concept à mesurer et de juger si les questions sélectionnées représentaient bien toutes les facettes du concept à mesurer. Il s'agissait d'analyser la pertinence du contenu et en particulier des différents critères retenus de la classification NM-Score (VP1) pour aboutir à une nouvelle version (VP2) qui entrera dans l'étape suivante de la validation.

❖ Présentation des méthodes de synthèse de l'information

Dans le domaine de la recherche médicale, comme dans le cadre de la recherche sur l'organisation des soins, le recours aux méthodes de synthèse de l'information s'impose fréquemment. Lorsque l'information publiée est insuffisante, seule une synthèse qualitative, reposant sur la recherche d'un consensus, peut être envisagée. Certaines méthodes de consensus permettent de produire des estimations quantitatives (et éventuellement qualitatives) à partir de cette approche qualitative ; il s'agit principalement de la « procédure Delphi » [17][47][68][125]. Les méthodes de consensus produisant des estimations quantitatives relèvent d'une approche systématique et structurée, incluant des procédures itératives (questionnaires et/ou réunions).

Elles s'adressent à des problèmes soigneusement définis, pouvant être traités de façon rapide et économique.

Elles intègrent schématiquement les données empiriques disponibles ainsi que le jugement et l'expérience des participants. Ces derniers sont habituellement des personnes représentatives de leur profession, reconnues comme expertes et/ou capables de mettre en œuvre les recommandations ; dans certains cas, il est utile d'inclure des représentants des patients. Le but de ces méthodes est de mesurer (but évaluatif) et de développer (but interventionnel) le consensus.

En pratique, la mesure du consensus consiste à déterminer dans quelle mesure des experts et/ou des représentants du public sont en accord sur un sujet donné. Ces méthodes respectent toutes l'anonymat des avis émis par les participants, ce qui permet d'éviter les dominations éventuelles (effets « leader d'opinion ») et de limiter les conflits d'intérêt. Le développement d'un consensus consiste à transformer, lorsque cela est possible et de manière « démocratique », l'opinion des membres d'un groupe de travail en consensus, en résolvant les éventuels désaccords observés. La mesure du consensus inclut d'une part l'évaluation du degré d'accord des participants avec les propositions considérées, évalué typiquement à partir du recueil de l'opinion de chaque participant au moyen d'une échelle numérique ou catégorique ; d'autre part l'évaluation de l'accord entre les participants, évalué typiquement par des indicateurs statistiques de tendance centrale (moyenne ou médiane) et de dispersion calculés au niveau du groupe. La « RAND Corporation » aux Etats-Unis propose d'utiliser une échelle ordonnée de 1 à 9, où 1 représente l'accord le plus faible et 9 le plus élevé. Selon elle, la validité contextuelle (« face validity ») est reconnue lorsque le score médian de l'accord est supérieur ou égal à 7, en l'absence de désaccord ; un désaccord est défini comme plus de 30 % des scores individuels repartis entre 1 et 3 et entre 6 et 9 [47].

Concernant la méthode Delphi, l'origine du nom de la procédure remonte à l'antiquité grecque, évoquant le mode d'interprétation et de prévision de l'oracle de Delphes. Elle est utilisée dans le domaine de la santé depuis la fin des années 70, initialement dans le cadre de la recherche en soins infirmiers [17][47][68][125]. C'est la méthode que nous avons utilisée. Il s'agit d'une procédure itérative et interactive comportant 2 à 4 tours. Les participants constituant le groupe d'experts ont été choisis car ils étaient « représentatifs » des connaissances et/ou des perceptions actuelles, relativement impartiaux mais intéressés et impliqués dans la problématique abordée. La taille optimale du groupe est d'au minimum 15 participants (jusqu'à plus de 60 dans de nombreuses publications). Aucun professionnel du groupe de coordination de la méthode ne doit appartenir au groupe de participants. Les avantages de la procédure Delphi sont le coût financier limité, l'absence de limitation géographique, l'anonymat (identités inconnues du groupe) ou le quasi-anonymat des participants (identités connues du groupe, mais pas leurs réponses), et la bonne acceptabilité des résultats (sous réserve du caractère « représentatif » du groupe). Les limites éventuelles sont l'existence d'un possible biais de sélection (dans le cas d'une participation insuffisante de certains membres du groupe) et l'absence d'opportunité pour les participants de débattre ensemble.

❖ Les analyses statistiques

Avant le début de la méthode Delphi, le consensus pour chaque item relatif à la classification NM-Score a été défini comme :

1/ il existe un haut degré d'accord défini par

Score médian de l'accord entre experts ≥ 7

Au moins 80% des experts ont un score ≥ 7

2/ en l'absence de désaccord : le désaccord est défini comme plus de 30% des scores individuels des experts répartis entre 1 et 3 et entre 6 et 9

Un tableau informatique a été utilisée pour analyser les résultats et identifier les items n'ayant pas abouti à un consensus au sein du groupe. Ces items ont alors été proposés au tour suivant.

C.3.3 ETUDE DE VALIDATION INTER OBSERVATEUR

Cette phase de validation avait pour but :

- de mesurer le degré d'accord entre les cotations entre professionnels pour mesurer la fiabilité inter-observateur de la classification ;
- d'apporter des modifications éventuelles à la classification en se basant sur les explications données par les examinateurs en cas de désaccord lors d'une cotation ;
- d'aboutir à une nouvelle version (VF1).

❖ Recrutement des patients et déroulement de l'étude

Ont été inclus 71 patients âgés de 6 à 60 ans, porteurs d'une MNM, connus des deux examinateurs depuis au moins 3 mois et dont leur dernière consultation et/ou leur dernier bilan dataient de moins de 3 mois lors du jour de cotation.

Chaque centre investigateur a reçu un classeur comprenant des lettres d'information pour les patients, le protocole détaillé de l'étude et plusieurs formulaires de classifications.

Tous les thérapeutes ayant accepté de participer à cette étude de validation devaient, dans un premier temps, se familiariser avec la classification NM-Score avant de débiter les cotations. 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 à partir de l'interrogatoire et/ou de l'observation en situation écologique et/ou du dossier du patient. Un même duo de co-examineur ne pouvait coter que 5 patients au maximum.

Après analyse des cotations, une demande de confrontation était envoyée aux thérapeutes concernés en cas de désaccord de cotation afin de déterminer l'origine du désaccord.

❖ Analyses statistiques

La fidélité inter-juge a été analysée en utilisant le coefficient kappa, qui permet de déterminer si l'accord observé est différent d'un accord dû au hasard. Le test indiquait si le coefficient était significativement différent de 0 (accord dû au hasard seulement). La fidélité inter-juge a été évaluée pour chaque domaine de la classification NM-Score. Afin de prendre en compte

l'importance des discordances (plus d'écart de 2 grades que de 1, par exemple), un kappa pondéré, dont les résultats dépendent du système de pondération choisie, été utilisé.

Le coefficient kappa ne prend pas en compte la diversité des évaluateurs, ici les deux « juges » sont les deux « évaluations », ni la maladie, l'âge ou le support de l'évaluation. Une stratification pourrait être utilisée, mais au risque d'une perte de puissance car les effectifs des strates étaient petits.

C.3.4 ETUDE DE LA VALIDITE CONTRE CRITERE

L'objectif de cette étude était la validation contre critère de la classification, c'est-à-dire par comparaison à des mesures de fonction motrice retenues comme gold standard (score MFM, scores de Brooke et de Vignos).

❖ Déroulement de l'étude

Aucune consultation ni examen supplémentaire n'ont été imposés au patient par rapport à son suivi habituel. Chaque centre investigateur a reçu des cahiers d'observation « MFM/NM-Score » dans lesquels la classification NM-Score a été ajoutée au cahier d'observation habituelle de la MFM disponible sur le site Internet de la MFM (<http://www.mfm-nmd.org/accueil.aspx>). 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 selon la classification NM-Score par le thérapeute a été réalisée dans les services le jour de la passation de la MFM prévue pour le patient. Les données étaient rassemblées sur le CRF. Elles comportaient en plus de la MFM et du NM-Score, les scores fonctionnels de Brooke et de Vignos. Toutes les données étaient ensuite rentrées dans la base de données informatisée de la MFM (<https://194.167.35.139/4DCGI/index.html>). Cette base recense, après obtention du consentement du patient, toutes les données MFM passées par les patients dans les centres participant à la base.

Population concernée :

Les patients inclus dans l'étude étaient des patients :

- porteurs d'une MNM ;
- âgés de 6 à 60 ans ;
- ayant donné leur accord afin que les données les concernant puissent être utilisées en recherche clinique.

La validation de la classification devra au moins porter sur 150 patients recrutés de façon multicentrique de mai 2010 à mai 2011.

❖ Analyses statistiques

L'effectif de 150 sujets a été choisi pour cette étude de validation. Cet effectif satisfaisait les contraintes de faisabilité et de puissance de cette étude. La justification statistique de cet effectif ne repose pas sur les risques alpha et bêta, car notre étude était principalement une étude de corrélation. La puissance *a priori* ainsi que l'ampleur des intervalles de confiance n'étaient pas prévisibles de façon formelle, mais l'expérience montre que sur les données du type concerné par cette étude, l'effectif de 150 est tout à fait en mesure de donner satisfaction.

Pour analyser la validité contre critère de la classification NM-Score, nous avons choisi pour chaque domaine (D1, D2, D3) des échelles de référence appelée « fuzzy gold standard » étant donné le caractère imparfait des gold standards dans ce domaine particulier.

Cette phase a comporté une analyse de concordance entre l'échelle élaborée, 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, mais regroupée en 5 classes de 20, pour la comparaison. La concordance entre le grade de sévérité et les scores de Vignos et de Brooke a aussi été étudiée (validité nomologique). Nous avons utilisé des mesures de concordance, telles que la corrélation estimée par le coefficient de Spearman.

C.3.5 VALIDITE DE CONSTRUIT

Cette phase de validation étudie le degré auquel un test mesure le construit qu'il prétend mesurer [5][18].

❖ *Analyses statistiques*

La validité de construit de la classification NM-Score a été étudiée en analysant la structure interne de l'échelle. La démarche est celle d'une analyse factorielle confirmatoire, avec un modèle théorique et une hypothèse à tester.

Le modèle comprenait 3 variables latentes, une pour chaque domaine de fonction motrice (D1, D2 et D3). Pour chaque variable latente, il y a un modèle de mesure (Figure 12).

Par exemple, pour le domaine D1, le modèle de mesure comprend la variable latente "Domaine 1 Station debout et transfert" et 5 indicateurs de cet état avec chacun un facteur d'échelle :

- score au NM-Score D1, facteur d'échelle $\lambda_{1.1}$ fixé (ou contraint) à 1
- score MFM D1 (en valeur absolu ou pourcentage), facteur d'échelle $\lambda_{2.1}$
- grade de l'échelle de Vignos, facteur d'échelle $\lambda_{3.1}$ et en plus deux indicateurs
- score au NM-Score D2, facteur d'échelle $\lambda_{4.1}$
- score au NM-Score D3, facteur d'échelle $\lambda_{5.1}$

Les facteurs d'échelle de ces 2 derniers indicateurs sont ceux à tester.

Les types de modèles de mesure pour les 2 autres variables latentes et les hypothèses correspondantes sont similaires.

Les covariances entre les variables latentes étaient aussi incluses dans le modèle (ψ_{12} , ψ_{13} et ψ_{23}) (Figure12). La variable donnant le diagnostic n'est pas représentée sur le schéma pour la lisibilité.

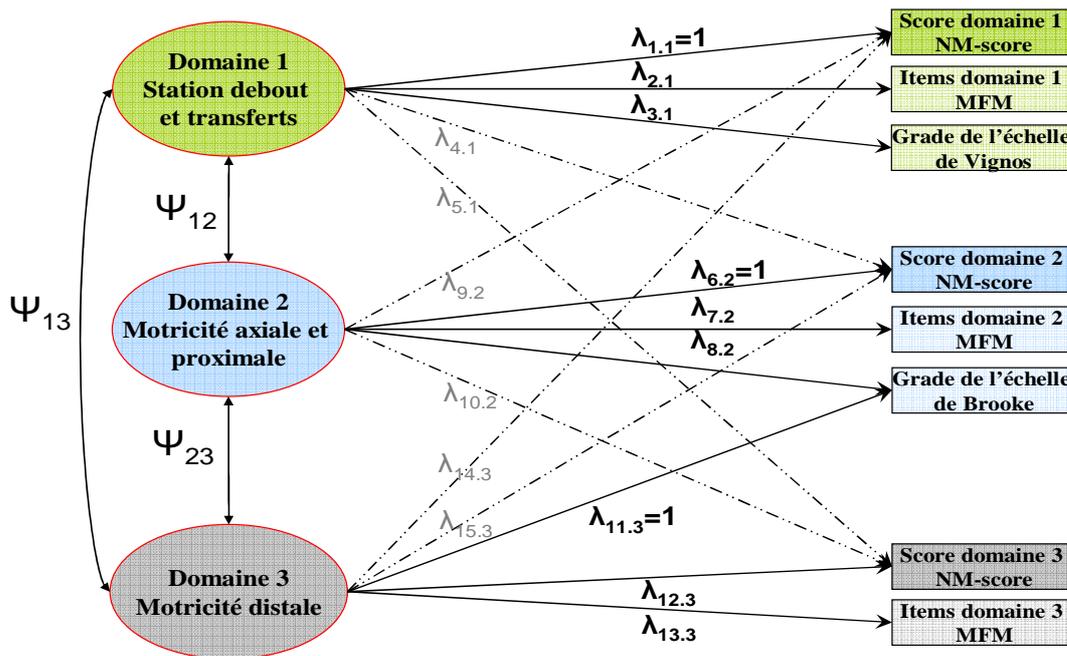


Figure 12 : Représentation graphique de la méthodologie utilisée pour l'analyse de la validité des construits du NM-Score

Les différentes hypothèses à tester pour le modèle complet étaient les facteurs d'échelle $\lambda_{4,1}$ $\lambda_{5,1}$ $\lambda_{9,2}$ $\lambda_{10,2}$ $\lambda_{14,3}$ $\lambda_{15,3}$ (en non gras et correspondant aux flèches en pointillées sur la figure) qui devaient être petits (en valeurs absolues) par rapport aux autres lambdas de leurs domaines respectifs. Mais ces lambdas seront bien différents de 0 de par la pathogénie des MNM atteignant plusieurs domaines de la fonction motrice.

Cette approche par un modèle à variable latente permet d'évaluer la validité convergente [122] des construits D1, D2 et D3. Chaque indicateur doit refléter le construit pour lequel il a été conçu et non les autres construits. Ainsi pour le construit D1, les facteurs d'échelle des indicateurs NM-D2 et NM-D3 doivent être plus petits (en valeur absolue) que le facteur d'échelle de l'indicateur NM-D1 conçu pour refléter le construit D1 (et donc le domaine D1).

NB : un facteur d'échelle n'est égal à la corrélation entre l'indicateur correspondant et la variable latente que si l'indicateur et la variable latente sont standardisés (c'est-à-dire de moyenne égale à 0 et de variance égale à 1), ce qui n'est pas le cas ici.

Cependant, les indicateurs sont ordinaux, donc ont une même variance résiduelle. Donc la corrélation entre un indicateur et sa variable latente est une fonction croissante du facteur d'échelle, donc la corrélation et le facteur d'échelle varient dans le même sens. De plus, la corrélation et le facteur d'échelle sont de même signe.

En bilan, quand le facteur d'échelle est grand, la corrélation est grande ; quand le facteur d'échelle est positif, la corrélation est positive; quand le facteur d'échelle est négatif, la corrélation est négative. Les comparaisons entre facteurs d'échelles d'un même variable latente sont possibles. Le facteur d'échelle informe donc bien sur la corrélation. Par contre, les facteurs d'échelle peuvent être inférieurs à -1 et supérieurs à 1.

C.4 RESULTATS

C.4.1 DEVELOPPEMENT DE LA CLASSIFICATION ET VALIDATION INTER-OBSERVATEUR

C.4.1.1 L'accord au sein du groupe sur le cahier des charges précis de la classification

- Une classification dans chacun des 3 domaines de fonction motrice identifiés dans la MFM :
 - D1 Marche, debout et transferts
 - D2 Motricité axiale et proximale
 - D3 Motricité distale
 - La classification s'adresse à l'enfant et l'adulte entre 6-60 ans pour ne pas avoir à tenir compte des effets du vieillissement ni du développement psychomoteur de l'enfant.
 - Une classification ordinale en 5 grades :
 - Pas d'atteinte = 0**
 - Atteinte légère = 1**
 - Atteinte moyenne = 2**
 - Atteinte sévère = 3**
 - Atteinte très sévère = 4**
- sans prétendre égale la distance d'un grade à l'autre, ni que la distribution des MNM se fasse de façon égale dans les cinq grades. La description des 5 grades est très large. La distinction entre les différents niveaux de fonction motrice est basée sur les limites fonctionnelles, la nécessité d'une aide technique comprenant les aides techniques de déplacement (telles que les cadres et cannes) mais aussi les fauteuils roulants, et dans une moindre mesure la qualité du mouvement. La description de chaque grade doit être simple et compréhensible de tous : parents, patients, kiné et médecins. La définition des termes utilisés doit être sans équivoque. Un effort a été fait pour prendre en compte les possibilités fonctionnelles du patient plus que ses limites.
- Une classification basée sur les performances motrices du sujet. Il s'agit de déterminer quel niveau représente le mieux les performances actuelles du sujet et ses limites dans la fonction motrice. On accorde une importance particulière aux performances habituelles du sujet à la maison, à l'école et autres lieux de vie habituels. Il est par conséquent important de classer les performances habituelles (et non les performances les meilleures).
 - Un outil suffisamment discriminant. La différence entre les 5 grades doit être suffisante pour être cliniquement perçue et correspondre à des situations fonctionnelles différentes.
 - Une classification simple, faisable en moins de 10 minutes en consultation. Cette classification n'a pas l'intention de décrire très précisément la fonction d'un sujet donné mais plutôt de décrire de façon la plus imagée possible le grade fonctionnel du sujet.
 - Des consignes de classification claires pour permettre une reproductibilité satisfaisante.
 - Un style grammatical simple, des termes facilement compréhensibles (+ renvoi à des définitions si nécessaire).
 - Description de chaque grade de sévérité à partir de critères dit « majeurs » représentant le mieux le domaine étudié. Des critères dits « mineurs » comme l'utilisation d'aide technique, d'appareillage ou d'aide humaine sont associée aux critères majeurs pour parfaire la description.

Les critères majeurs doivent être utilisés en premier lieu pour classer les patients. Si il existe une discordance entre les critères majeurs et mineurs, le sujet sera classé prioritairement en fonction du critère majeur.

De façon générale, les patients qui possèdent toutes les fonctions décrites par les critères majeurs pour un grade donné seront probablement classés à ce grade ou au grade supérieur ; au contraire, les patients qui n'ont pas les fonctions décrites pour tel grade seront probablement classés au grade inférieur.

C.4.1.2 Le choix et la définition des critères dits majeurs pour classer les patients dans chacun des grades pour chaque domaine de la classification (D1, D2, D3)

Pour le domaine D1 : Capacités de marche, course et saut habituelles. Par exemple pour D1 en fonction des grades sont proposée ces descriptions du critère majeur

- 0 Capacités de marche, course et saut identiques à celles d'une personne du même âge**
- 1 Atteinte légère : diminution des capacités de marche mais marche 500m et plus (centre commercial).**
- 2 Atteinte moyenne : marche 50m (rue)**
- 3 Atteinte sévère : marche 5m (pièce) ou station debout possible pour certains transferts**
- 4 Atteinte très sévère : ne peut faire aucun pas ni tenir debout quelques soient les conditions.**

Pour le domaine D2 : Capacités habituelles de tenue assise, de motricité proximale et de tenue de tête

Pour le domaine D3 : Capacités habituelles de manipulation des objets

C.4.1.3 Etude expérimentale

Cette étude a été réalisée par les professionnels du groupe d'étude sur les patients vus en consultation pour identifier des critères dits mineurs permettant d'améliorer la description des patients dans chaque domaine

Les patients, selon les cas, ont bénéficié d'un classement dans un ou la totalité des domaines. Le professionnel devait classer le patient en utilisant les critères majeurs identifiés et ensuite répondre aux questions concernant les capacités du patient à remplir différents critères dits mineurs. Ces critères mineurs, par définition, sont qualitatifs, dépendent de l'environnement et des facteurs personnels. Ils ont comme utilité d'illustrer la description.

Pour le domaine D1, par exemple, les critères mineurs sont repris dans le tableau 24 et pour chaque critère mineur le pourcentage de patients remplissant le critère est donné. Un critère mineur est retenu pour améliorer la description d'un grade lorsque qu'au moins 75% (0,75) des patients classés dans ce grade répondaient à ce critère.

Domaine 1	0	1	2	3	4
Nb de patients testés	3	17	10	8	10
Proportion de patients remplissant les critères mineurs :					
Fatigabilité	0	0,75	1	1	0
Diminution de l'endurance	0	0,70	1	1	0
Chutes	0	0,6	0,2	0	0
Anomalies qualitatives de la marche	0	1	1	1	0
Marche sur terrain irrégulier	1	1	0,1	0	0

Domaine 1	0	1	2	3	4
Marche en pente	1	0,5	0,1	0	0
Course	1	0,5	0	0	0
Montée et descente des escaliers +/- rampe	1	1	0,6	0	0
Activités sportives	1	0,9	0,1	0	0
Port d'appareillages	0	0,25	0,3	0,75	0,5
Utilisation d'aide technique pour les déplacements extérieurs	0	0	0,2	1	1
Se relève seul d'un siège ordinaire	1	1	0,9	0,4	0
Se relève seul du sol	1	0,7	0,9	0,1	0
Assure les transferts seul	1	1	1	1	0,1
Appuie sur les membres inf lors des transferts		1	1	1	1

Tableau 24 : Phase de développement du NM-Score : Résultats de l'étude expérimentale pour le NM-Score D1

C.4.1.4 Méthode Delphi et Fidélité inter-observateur

La méthodologie de développement ainsi que les résultats de l'étude de validation inter-observateur de la classification NM-Score ont été explicités dans un article soumis à Acta Neuropathologica (Annexe 5).

Development and reliability of NM-Score: a system to classify motor function in patients with neuromuscular diseases

Carole Vuillerot, Pascal Rippert, Sylvain Roche, Jean Iwaz, René Ecochard, Carole Bérard, and the NM Score Study Group.

To address the need for a standardized system to classify the motor function of patients with neuromuscular diseases, we developed a five-level classification for standing position and transfers (NM-Score D1), for axial and proximal motor function (NM-Score D2) and for distal motor function (NM-Score D3). Delphi survey consensus methods was used to examine content validity and revise the classification until consensus among 29 experts (physical therapists, occupational therapists and medical doctors with expertise in neuromuscular diseases) was achieved. A total of 71 patients were involved to examine the reproducibility of NM-Score. Respectively for NM-Score D1, D2 and D3, interrater reliability coefficients (κ) reached 0.77, 0.69 and 0.642. Analysis of the disagreements between raters allowed modifications in the classification and users instructions. In order to complete the psychometric analysis of the classification a validation study against criteria is ongoing. The next step is the validation of a NM-Score English version in order to use the classification worldwide for its applications in clinical practice and research.

Soumis à Acta Neuropathologica

C.4.2 VALIDATION CONTRE CRITERE

C.4.2.1 Description de la population

Un total de 158 patients a été inclus dans cette phase de validation. Il s'agit essentiellement de patients marchants (124) et qui, pour 42% d'entre eux, possèdent un fauteuil roulant manuel et/ou électrique. La surreprésentation masculine de l'échantillon (102/56) est en lien avec le nombre important de patients atteints de DMD (27) et de Becker (10) touchant exclusivement le sexe masculin. Les autres diagnostics représentés sont : ASI (22), myopathies et DMC (26), neuropathies périphériques (21), DMS (17), DFSH (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 (+/- 3,7). Cent quarante-huit patients ont été évalués par le score de Vignos, 143 patients par le score de Brooke et 158 par la MFM. La distribution des patients selon les grades de la classification NM-Score (0, 1, 2, 3, 4) est présentée pour chaque domaine dans le tableau 25. Comme le montre la faible représentation des grades 3 et 4 du NM-Score D3 (de 10 patients), peu de patients ont en une atteinte distale très sévère. Un résultat similaire est obtenu pour la MFM où 6 patients seulement ont un score MFM-D3 inférieur à 40%. Ceci s'explique par la surreprésentation des maladies à faiblesse proximale prédominante (DMD, ASI, DMC) dans notre population.

NM-Score	Grade	Nb de patients	Fréquence (%)
NM-Score D1	0	8	5,1
	1	63	39,9
	2	41	25,9
	3	17	10,8
	4	29	18,3
Total		158	100,0
NM-Score D2	0	48	30,4
	1	73	46,2
	2	18	11,4
	3	11	7,0
	4	8	5,0
Total		158	100,0
NM-Score D3	0	45	28,5
	1	58	36,7
	2	45	28,5
	3	7	4,4
	4	3	1,9
Total		158	100,0

Tableau 25 : Distribution des 158 patients dans chaque grade de l'échelle NM-Score pour les 3 domaines D1, D2 et D3. En gras sont indiqués les grades peu représentés dans l'échantillon

C.4.2.2 NM-Score versus gold standards

Les distributions du grade de Vignos par grades NM-Score D1 sont représentées dans le tableau 26.

NM-Score D1	Nb de patients	Score de VIGNOS					
		Moy(DS)	min	p25	p50	p75	max
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

Tableau 26 : Distribution du score de Vignos par grades NM-Score D1

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 planché. Au contraire la, dispersion du score de Vignos pour le grade 3 de la classification NM-Score D1 (moyenne, 5,75, minimum : 2, maximum : 9) est forte témoignant d'une grande variabilité de l'atteinte des patients classés dans ce grade.

Les distributions du score de Brooke par grades NM-Score D2 et D3 sont représentées dans le tableau 27.

NM-Score D2	Nb de patients	Score de BROOKE					
		Moy(DS)	min	p25	p50	p75	max
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	Nb de patients	Moy(DS)	min	p25	p50	p75	max
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

Tableau 27 : Distribution du score de Brooke par grades NM score D2 et D3

Les distributions des scores MFM D1, D2 et D3 par grades NM-Score D1, D2 et D3 correspondant sont représentées sous la forme de box plot sur les figures ci-dessous (Figure 13)

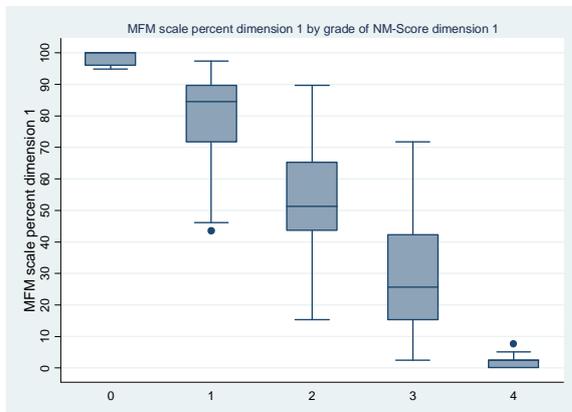


Figure 13a : Box plots des scores MFM D1 en fonction du grade de sévérité NM-Score D1

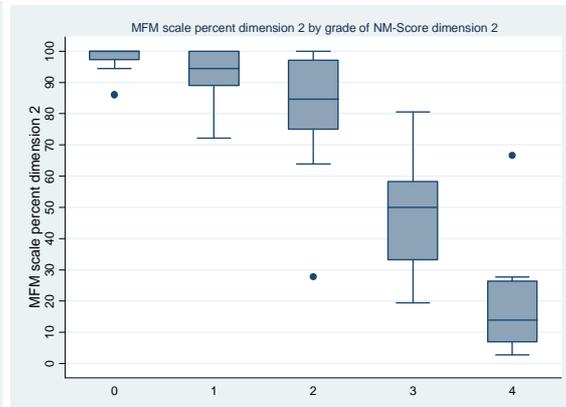


Figure 13b : Box plot des scores MFM D2 en fonction du grade de sévérité NM-Score D2

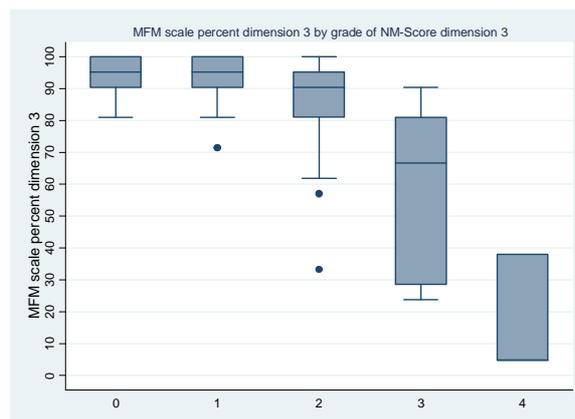


Figure 13c: Box plots des scores MFM D3 en fonction du grade de sévérité NM-Score D3

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 boxplots. 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 boxplots.

C.4.2.3 Validité contre critère : étude de corrélations

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 (95%) (Tableau 28).

	Coefficient Spearman	Intervalle de confiance 95%	
VIGNOS vs NM-Score D1	0,86	0,81	0,89
MFM D1 vs NM-Score D1	-0,88	-0,91	-0,83

	Coefficient Spearman	Intervalle de confiance 95%	
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

Tableau 28 : Coefficients de Spearman

Les meilleures corrélations sont retrouvées pour le NM-Score D1 et NM-Score D2 avec une force de corrélation plus importante avec l'échelle MFM qu'avec les scores fonctionnels.

C.4.3 VALIDATION DES CONSTRUIITS

Ce modèle permet de prendre en compte les différents diagnostics. Afin de prendre en compte la part de variabilité des indicateurs expliquée par les pathologies, ces dernières ont été introduites en effets fixes des modèles de régression à variable latente.

Pour le domaine D1, les résultats obtenus (Tableau 29a) confirment les hypothèses formulées. A patient égal, le NM-Score D1 et le score MFM D1 discriminent les patients quasiment de la même façon. Pour le score de Vignos, la corrélation reste satisfaisante (0,71). Par contre, les facteurs d'échelles pour les NM-Score D2 et D3 qui sont plus petits témoignent d'une faible corrélation entre le NM-Score D1 et les NM-Score D2 et D3. Autrement dit, lorsque l'on connaît le NM-Score D1, on ne peut prédire que de façon très incomplète les NM-Score D2 et D3. Cette corrélation qui n'est pas nulle signifie qu'il y a tout de même un lien entre les différents NM-Score ; lien lié à la pathogénie des MNM avec une atteinte des différents domaines mais aussi probablement lié à l'imprécision de la mesure. Si l'échelle MFM D1 et le score de Vignos sont considérés comme de bonnes mesures du construit D1, la validité convergente du construit D1 n'est pas mis en doute. Ceci est en faveur de l'indicateur NM-Score D1 de l'échelle NM-Score qui représente donc bien ce pour quoi il a été conçu.

Indicateurs du construit D1	Facteurs d'échelle
NM-D1	1 (fixé)
MFM-D1	-0,96
Vignos	0,71
NM-Score D2	0,29
NM-ScoreD3	0,21

Tableau 29a : Facteurs d'échelle pour les indicateurs de construits D1

Pour le domaine D2 (Tableau 29b), l'écart reste supérieur mais il se réduit entre le facteur d'échelle lié au score MFM D2 (-0,83) ou au score de Brooke (0,79) et les facteurs d'échelle liés au NM-Score D1 (0,70) et D3 (0,50). Si l'échelle MFM D2 et le score de Brooke sont considérés comme de bonnes mesures du construit D2, la validité convergente du construit D2 peut être mise en doute. L'indicateur NM-Score D2 de l'échelle NM-Score représente partiellement ce pour quoi il a été conçu.

Indicateurs du construit D2	Facteurs d'échelle
NM-D2	1 (fixé)
MFM-D2	-0,83
Brooke	0,79
NM-Score D1	0,70
NM-Score D3	0,50

Tableau 29b : Facteurs d'échelle pour les indicateurs de construits D2

Pour le domaine D3 (Tableau 29c), les hypothèses ne sont pas confirmées avec un écart plus important entre le facteur d'échelle lié au score MFM D3 (-0,69) ou au score de Brooke (0,99) et les facteurs d'échelle liés aux NM-Score D1 (1,33) et D2 (1,51). De plus, la validité convergente du construit D3 est clairement mise en doute. Ceci n'est pas en faveur de l'indicateur NM-Score D3 de l'échelle NM-Score qui représente mal ce pour quoi il a été conçu.

Les résultats ne dépendent pas de la présence de l'échelle de Brooke parmi les indicateurs des deux construits D2 et D3, ils restent globalement les mêmes lorsque l'analyse est effectuée en l'excluant.

Indicateurs du construit D3	Facteurs d'échelle
NM-D3	1 (fixé)
MFM-D3	-0,69
Brooke	0,99
NM-Score D1	1,33
NM-Score D2	1,51

Tableau 29c : Facteurs d'échelle pour les indicateurs de construits D3

Pour les domaines D2 et D3, les résultats ne sont pas satisfaisants et on ne peut pas distinguer entre imperfection de la mesure et lien physiologique entre l'atteinte des différents domaines. En effet, cela revient à dire que lorsque l'on connaît, pour un patient donné et dans une pathologie donnée, le NM score D3, on est capable de prédire quel va être le NM-Score D1 et D2.

Pour avoir une idée du lien qu'il existe entre l'atteinte des différents domaines chez les patients, des tableaux croisés des scores des différents domaines de l'échelle NM-Score ont été réalisés. D'une manière générale, les diagonales (en italique) dans les tableaux suivants sont « assez bien remplies » et les effectifs de patients diminuent lorsque l'on « s'éloigne des diagonales » : ceci signifie qu'un patient ayant une atteinte d'un certain grade dans un domaine aura tendance à avoir une atteinte de grade assez proche dans les autres domaines.

NM-Score	D2=0	D2=1	D2=2	D2=3	D2=4	Total
D1 = 0	4	4	0	0	0	8
D1 = 1	34	29	0	0	0	53
D1 = 2	8	28	5	0	0	41
D1 = 3	2	9	6	0	0	17
D1 = 4	0	3	7	11	8	29
Total	48	73	18	11	8	158

Tableau 30a : Tableau croisé des NM-Score D1/D2

NM-Score	D3=0	D3=1	D3=2	D3=3	D3=4	Total
D1 = 0	7	1	0	0	0	8
D1 = 1	27	31	5	0	0	53
D1 = 2	8	18	14	1	0	41
D1 = 3	2	6	8	1	0	17
D1 = 4	1	2	18	5	3	29
Total	45	58	45	7	3	158

Tableau 30b : Tableau croisé des NM-Score D1/D3

NM-Score	D3=0	D3=1	D3=2	D3=3	D3=4	Total
D2 = 0	27	19	1	1	0	8
D2 = 1	18	34	20	1	0	53
D2 = 2	0	4	14	0	0	41
D2 = 3	0	1	8	2	0	17
D2 = 4	0	0	2	3	3	29
Total	45	58	45	7	3	158

Tableau 30c : Tableau croisé des NM-Score D2/D3

Plus précisément, un patient aura tendance à avoir une atteinte plus sévère dans le domaine D1 que dans le domaine D2 ou le domaine D3 (cf. effectifs sous les diagonales en noir et en gras dans les tableaux 30a et 30b).

Un patient aura tendance à avoir une atteinte moins sévère dans le domaine D2 que dans le domaine D3 (cf. effectifs au-dessus de la diagonale en surligné et en gras dans le tableau 30c).

Ces deux derniers points permettent d'expliquer en partie les facteurs d'échelles élevés pour les indicateurs NM-Score D1 et D2 du construit D3.

Ces tableaux confirment aussi la répartition non homogène des patients au sein des grades de sévérité. En effet, il y a très peu de patients ayant des atteintes sévère ou très sévère en D2 et D3, quelle que soit l'atteinte en D1.

CHAPITRE 3 : DISCUSSION ET PERSPECTIVES DE RECHERCHE

A DISCUSSION

A.1 MFM

Les trois études de la sensibilité au changement de la MFM réalisées à partir de populations différentes confirment la complémentarité par rapport aux mesures de force de l'évaluation fonctionnelle motrice [100][118].

La fonction motrice mesurée par la MFM sur une population générale de patients adultes et enfants porteurs d'une MNM diminue de façon significative au cours du temps, confirmant l'impression clinique d'évolutivité de ces pathologies. Dans cette population, la diminution la plus significative des scores MFM sur 1 an concerne les patients avec une DMD ($-5,8 \pm 6,3\%$ vs $-4,4 \pm 6,4\%$ pour la population totale). Ce résultat s'explique par l'histoire naturelle de cette pathologie, certainement l'une des plus progressives MNM de l'enfance. Les autres pathologies des patients de l'échantillon ayant des effectifs et une évolutivité moindre, la durée de suivi optimal pour observer des changements significatifs devrait probablement être supérieure à 2 ans.

Nous avons utilisé dans cette étude une approche de type « anchored based » pour relier les changements observés en termes de score MFM à une impression clinique de changement rapportée par les professionnels ou les patients. Cette approche permet de s'intéresser à la notion appelée « Responsiveness » par les anglosaxons qui renvoie à la plus petite différence cliniquement significative (« minimal clinically important difference », MCID) [37][62][111]. Le changement moyen des scores MFM totaux est significativement plus bas chez les patients pour lesquels eux-mêmes ou leurs médecins rapportent une stabilité de leur état fonctionnel, que chez les patients pour lesquels eux-mêmes ou leurs médecins rapportent une dégradation de leur état fonctionnel. En d'autres termes, les changements observés sur les scores MFM (équivalents à une dégradation) sont également exprimés par les patients ou leurs médecins lorsqu'on les interroge sur l'évolution de l'état fonctionnel. Selon nos observations, une diminution de 5 points du score MFM total est perçue par le patient et leurs médecins qui rapportent une dégradation fonctionnelle. Concernant les patients pour lesquels eux-mêmes ou leurs médecins rapportent une stabilité de l'état fonctionnel après 1 an, les scores MFM moyens ont néanmoins tendance à diminuer. Ceci évoque une sensibilité au changement plus importante de la MFM pour mesurer la dégradation fonctionnelle, ce qui est bien l'objectif de l'échelle.

Dans la population de patients porteurs de DMD, il existe une diminution continue exponentielle des scores MFM en fonction de l'âge, en accord avec les résultats publiés concernant l'évolution de la force musculaire chez ces patients [4][81].

La MFM peut être administrée à des patients ambulants et non ambulants, ce qui est particulièrement intéressant dans la DMD où les sujets perdent la marche, souvent à l'adolescence, au cours de l'évolution de leur maladie. S'intéressant à l'évolution des scores MFM sur 3 mois, seul le score D1 diminue de manière significative alors que sur 1 an, tous les scores MFM ont diminué de manière significative.

Chez les patients ambulants, D1 (position debout et transferts) est le domaine le plus informatif avec une diminution moyenne de 17,2% par an avant la perte de la marche. Après la perte de la marche, D2 (mobilité axiale et proximale) qui montre une diminution moyenne de 9,4% par an devient le plus intéressant. Chez les patients de plus de 14 ans, la diminution moyenne de D3 (motricité distale) était de 10,8% par an.

Le domaine D1 étant intimement lié à la capacité de marcher et étant sensible aux changements à court terme dans les DMD, il nous a permis de proposer un calcul prédictif 1 an avant la perte de marche. Alors que l'âge à la perte de la marche varie grandement (entre 6 et 14 ans dans notre groupe), il semble possible de prédire qu'un patient DMD perdra sa capacité de marcher dans une année si le score MFM D1 est proche de 40 % et/ou si le score total MFM est proche 70%. Cette donnée a des conséquences cliniques importantes permettant aux professionnels d'anticiper avec le patient et sa famille cette étape importante de perte de marche.

Les présents résultats sont également prometteurs en ce qui concerne l'utilisation de la MFM dans les essais cliniques pour faire la preuve de l'effet thérapeutique sur l'évolution de la fonction motrice. Dans notre échantillon, comparativement aux témoins non traités, les patients atteints de DMD traités avec des corticoïdes ont présenté une stabilisation du score total MFM et du score MFM D2 sur 1 an.

La classification des ASI a été modifiée à plusieurs reprises afin de refléter au mieux l'histoire naturelle de la maladie [116][146], sans y parvenir complètement. Toute tentative de classification est artificielle et ne reflète pas la réalité physiopathologique en particulier la variabilité interindividuelle importante au sein de cette pathologie [116][146]. La classification des ASI doit être utilisée avec prudence surtout concernant l'annonce du pronostic. Les experts s'accordent sur le fait que les soins les plus appropriés pour les patients atteints d'ASI doivent être adaptés en fonction de leur état fonctionnel actuel plutôt qu'en fonction du type d'ASI, I, II ou III [140]. Dans la population étudiée lors de l'étude de la sensibilité au changement de la MFM dans l'ASI, la répartition des patients en fonction de l'âge et de leur score total MFM tend à confirmer que l'ASI correspond à une seule et même maladie composée d'un spectre continu de sévérité plutôt que de différentes maladies [146]. Il existe des chevauchements entre les patients appartenant à différents types d'ASI. Ainsi on peut décrire des patients considérés comme des « bons type II » (patients moins atteints que d'autres patients de type II au même âge) ayant la même fonction motrice que des patients considérés comme des « mauvais type III » (patients plus atteints que d'autres patients de type III au même âge).

La MFM permettant l'évaluation fonctionnelle quel que soit le statut fonctionnel, elle constitue l'outil unique utilisable chez tous les patients quel que soit le type d'ASI. Comme dans la DMD, le fait que certains patients (les ASI de type III en particulier) soient amenés parfois à perdre la marche au cours de l'évolution encourage l'utilisation de la MFM pour le suivi longitudinal de ces patients. Peu de critères d'évaluation autre que la mortalité [55][56] sont disponibles pour évaluer et suivre l'évolution de maladie avec une faiblesse musculaire importante car les mesures de force sont des mesures fiables dans les types de II et III, mais pas dans le type I.

Confirmant des données de la littérature [11][115][140][145], nous avons pu montrer dans notre étude que l'ASI est une maladie lentement progressive, avec un déclin de la fonction motrice.

Une diminution significative des scores MFM D2 chez les patients ASI de type II et des scores D1 chez les patients ASI de type III a pu être observée sur une période d'1 an. Cette dégradation fonctionnelle, contrairement à la DMD, n'est pas continue mais est constituée d'une alternance de phases stables et de phase d'aggravation. Dans notre échantillon par exemple, ce sont les patients atteints d'une ASI de type III âgés de 10 à 15 ans qui présentent la perte fonctionnelle la plus évidente ($-6,22 \pm 6,6$ points par an en moyenne pour D1) ce qui correspond bien à l'histoire naturelle de la pathologie avec une perte de marche signalée au bout de 10 ans d'évolution de la maladie.

Malgré la sévérité de l'atteinte fonctionnelle de ces patients et leur jeune âge, leur coopération a été optimale pour plus de 80% de l'échantillon, et il n'y avait pas de relation entre la sévérité de la maladie et le temps pour compléter le test.

La mesure fiable de ces changements sur plus d'un an à l'aide de la MFM (exprimée en un taux annuel de changement) nous a permis d'estimer le nombre de sujets nécessaires pour un essai clinique en fonction du type d'ASI, l'effet du traitement prévu et la puissance de l'étude. Afin de calculer un nombre de sujet nécessaire encore plus précis et utile pour la mise au point d'essais thérapeutiques, l'évolution de la fonction motrice des patients ASI inclus dans le groupe placebo d'un essai thérapeutique reste cependant à être étudiée. Cette évolution devrait être en effet être différente (biais de prise en charge, effet placebo) de l'histoire naturelle de la maladie.

A.2 NM-SCORE

La classification NM-Score est un outil valide et fiable permettant de classer les patients 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. Nous avons, comme pour la MFM, choisi de ne pas inclure la fonction motrice de la face. La classification NM-Score est basée sur la description clinique de la fonction motrice permettant de fournir « une image » des patients. Cette nouvelle classification a été bien accueillie par les nombreux praticiens et chercheurs qui l'ont déjà utilisée.

La classification NM-Score est un outil facile et rapide à utiliser pour décrire les performances motrices des patients. Elle 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. 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 ou non. En fonction de la pathologie ou de l'âge, un pattern spécifique peut être observé. Par exemple, dans l'ASI ou la dystrophie de type FSH, les NM-Score D1 et D2 sont les plus altérés, quel que soit l'âge des patients, car l'atteinte de la maladie est essentiellement proximale. Alors que dans la DMD le NM-Score D1 est altéré dès les premières années de la maladie, les NM-Score D2 et D3 sont altérés de façon beaucoup plus tardive (NM-Score D2 étant altéré plus précocement que NM-Score D3).

D'une manière générale, dans les MNM, le domaine D3 de la fonction motrice est le plus longtemps préservé (sauf pour les rares myopathies 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.

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 classement est utile à la fois à des fins de recherche et pour le suivi clinique des patients.

Deux tours d'enquête ont été nécessaires avant 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. Pour surmonter les difficultés liées à la variabilité du développement psychomoteur normal de l'enfant nous proposons de n'utiliser le NM-Score que pour des patients porteurs d'une MNM âgés d'au moins 6 ans.

La fiabilité inter-observateur de la classification NM-Score est bonne lorsqu'elle est utilisée à partir de l'interrogatoire 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 (0,770, 0,690 et 0,642 pour respectivement les domaines D1, D2, et D3) et autorise l'utilisation du NM-Score dans les essais thérapeutiques multicentriques. Pour comparaison, Palisano et al [106] ont rapporté pour le « Gross Motor Function Classification System » (GMF-CS), classification ayant inspiré ce travail, un κ de 0,75 chez les enfants âgés de 2 ans et plus. Finkel et al. [48] rapportent un κ pondéré de 0,61 (IC à 95% de 0,59 à 0,63) concernant le TIMP. Les comparaisons avec les autres outils sont rendues difficiles par l'utilisation des coefficients intra-classe (ICC) pour estimer la fiabilité inter-observateur. A titre d'exemple, l'ICC est estimé entre 0,96 et 0,99 pour le score de Brooke [24][25]. 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 a montré que, dans la plupart des cas, un des 2 cotateurs au moins n'avait pas respecté la primauté des critères majeurs sur les critères mineurs pour la cotation. Nous avons donc proposé dans la version finale de la classification (Annexe 6) d'isoler sur une seconde page les critères mineurs pour limiter leurs utilisations qu'en cas de doute entre 2 grades (à partir des critères majeurs).

Notre objectif était de développer une classification rapide et facile à utiliser au lit du malade. Avant d'utiliser la classification NM-Score aucune formation n'est indispensable, les utilisateurs sont seulement priés de lire attentivement l'introduction et les instructions de l'utilisateur et de faire quelques essais de cotations au préalable pour une meilleure connaissance de l'outil.

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, les 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. L'interprétation de ces corrélations est plus rendue délicate par des effectifs insuffisants de patients ayant une atteinte très sévère en D2 ou D3, le calcul des coefficients de corrélation étant très en lien avec l'effectif de la population. Lorsque l'on s'intéresse aux corrélations entre la MFM et le NM-Score, les coefficients de corrélations sont satisfaisants pour D1 et D2, moyen pour D3. La limite du coefficient de Spearman, dans ce cas, est qu'il ne prend pas en compte l'ordre de la classification. Les corrélations de type polychoriques ne pouvaient, cependant, pas être utilisées du faite de la nature des variables.

Les analyses de validité des construits en utilisant un modèle à variables latentes montrent que les indicateurs NM-Score D1 et D2 de l'échelle NM-Score représentent bien ce pour quoi ils ont été conçus. L'indicateur NM-Score D3, au contraire, représente mal ce pour quoi il a été construit car il est trop corrélé aux indicateurs NM-Score D1 et D2. Lorsque l'on s'intéresse aux liens entre les différents domaines du NM-Score, il apparaît clairement un lien dans notre échantillon puisque les patients auront tendances à avoir une atteinte plus sévère en D1 qu'en D2, et une atteinte moins sévère en D2 qu'en D3. Afin de surmonter cette difficulté méthodologique, il faudrait intégrer une variable supplémentaire, par exemple le temps d'évolution de la maladie depuis le diagnostic qui permettrait de prendre en compte le stade de la maladie et de ce fait l'interaction entre les domaines. C'est une des limites du modèle à variable latente proposé qui ne prend pas en compte la chronologie des atteintes des différents domaines. Une meilleure spécification du modèle permettra de mieux évaluer les validités des construits D2 et D3.

La classification NM-Score est une classification ordinale (ou catégorielle avec un ordre entre les catégories). De par sa nature, elle ne se veut pas linéaire, c'est-à-dire qu'il n'existe pas forcément le même écart de sévérité entre chaque grade. Un patient passant du grade 0 au grade 1 ne s'aggrave pas forcément de la même façon (grandeur) que s'il passait du grade 3 au grade 4. Si nous voulions vérifier le caractère linéaire de cette classification, nous serions de plus gênés par le caractère par définition non linéaire de la MFM.

La classification NM-Score est en mesure de décrire avec précision une population donnée en termes de fonction motrice et aider à constituer des groupes homogènes d'atteinte fonctionnelle motrice. Le NM-Score est par nature peu sensible au changement, comme les scores de Vignos et de Brooke il ne peut pas être utilisé comme une mesure de résultat pour étudier l'efficacité d'un traitement donné car le changement de la fonction motrice nécessaire pour passer d'un grade à l'autre est trop grand. Cependant, il fournit une description complète (sauf la face) de la fonction motrice d'un patient dans tous les domaines (la position debout et les transferts, la fonction motrice axiale et proximale et la fonction motrice distale).

B PERSPECTIVES

B.1 ANALYSE DE LA MFM PAR LES MODELES DE LA THEORIE DE REPONSE AUX ITEMS

Dans les différents groupes de travail internationaux organisés ces dernières années sur les outils d'évaluation dans les MNM, l'accent est mis sur des analyses plus complètes utilisant, en particulier, des modèles de réponse à l'item, dont le modèle Rasch. Ces types d'analyse nécessitant des groupes de patients importants (> 300 patients), peu d'échelles ont pu en bénéficier à ce jour (essentiellement ACTIVLIM et ABILHAND). Si la MFM, que nous considérons comme un outil indispensable d'évaluation des fonctions motrices d'un patient, a un intérêt indiscutable pour ceux qui l'ont choisie et l'utilisent pour le suivi clinique des patients, cet intérêt n'est pas partagé par tous les professionnels et peu d'essais cliniques hors de France l'ont choisie comme principal critère d'évaluation. Il apparaît donc nécessaire, en utilisant les données en nombre suffisant fournies par la base de données de la MFM, de poursuivre l'analyse métrologique de cette échelle avec différents outils statistiques sur l'ensemble des données et par pathologie en utilisant des modèles de réponse à l'item. Une première étude utilisant les modèles Rasch a été réalisée sur des données MFM dans l'ASI par Cano et al [28] mais les résultats sont très discutables du fait d'une rigueur insuffisante dans les analyses réalisées.

Les résultats cumulés de ces analyses doivent permettre d'améliorer l'échelle et l'interprétation des résultats en fonction des différentes pathologies. Les équipes qui l'utilisent auront des données consolidées ; celles qui sont hésitantes à l'utiliser auront des informations supplémentaires. Grâce à l'analyse par pathologie, nous pourrions élaborer des recommandations et des plans expérimentaux type pour les laboratoires voulant utiliser la MFM comme outil d'évaluation lors d'essais thérapeutiques.

B.1.1 EVALUATION DE LA VALIDITE DES HYPOTHESES DES MODELES DE REPONSES AUX ITEMS

Dans le cadre des modèles à variables latentes généralisés et des modèles de la Théorie des Réponses aux Items TRI ou « Item Response Theory, IRT », l'étude des propriétés métrologiques de l'échelle MFM nécessite de vérifier au préalable que la MFM remplit les hypothèses fondamentales des modèles de type TRI à savoir l'unidimensionnalité, l'indépendance locale et la monotonie.

Unidimensionnalité

L'unidimensionnalité signifie que l'ensemble ou une partie des items d'un test mesure une seule variable aléatoire continue et non observée, dite variable latente. Une étude antérieure de l'échelle MFM a mis en évidence trois domaines [14]. L'unidimensionnalité de chaque domaine peut être étudiée par une Analyse Factorielle Confirmatoire (AFC).

Si l'hypothèse d'unidimensionnalité est valide, l'analyse peut se poursuivre dans le cadre des modèles unidimensionnels de la TRI. Pour chacun des trois domaines, plusieurs modèles unidimensionnels sont susceptibles d'ajuster les données. Trois analyses distinctes peuvent être réalisées.

Si l'hypothèse d'unidimensionnalité n'est pas valide, l'analyse est poursuivie dans le cadre des modèles à variables latentes généralisés.

Indépendance locale

La réponse à un item de la MFM est supposée être indépendante de la réponse à tous les autres items de la MFM, conditionnellement à la variable latente (la fonction motrice). Sinon, les estimations des mesures pourraient être biaisées. La dépendance locale entre les réponses des items est évaluée par l'examen des corrélations résiduelles entre les items et en utilisant des tests basés sur la TRI.

Monotonie

L'hypothèse de monotonie signifie que la probabilité théorique que le score à un item soit élevé (c'est-à-dire une meilleure aptitude à la réalisation de cet item) augmente quand la valeur de la variable latente (la fonction motrice) augmente. C'est une condition élémentaire pour les modèles unidimensionnels de la TRI dans le cas d'items ayant des réponses multiples et ordonnées, par exemple des items de type Likert. La monotonie peut être examinée par des graphiques : la probabilité observée (par rapport à la probabilité théorique) que le score d'un item soit supérieur ou égal à une valeur donnée doit être non décroissante en fonction du score total moins cette valeur. Ceci peut être, si besoin, complété par une approche de type TRI non paramétrique.

B.1.2 AJUSTEMENT AUX DONNEES D'UN MODELE DE REPONSES AUX ITEMS

Pour des tests dont les réponses aux items sont binaires, les deux principaux modèles de type TRI sont définis par le nombre de paramètres par item du test. Le modèle à un paramètre considère un seul paramètre par item, c'est un paramètre de difficulté [109]. Le modèle à deux paramètres considère un paramètre de difficulté par item et en plus un paramètre de discrimination par item [21]. Le modèle avec un troisième paramètre, paramètre de pseudo-chance qui traduit le fait que le patient peut deviner la bonne réponse à un item, n'est pas pertinent pour l'évaluation de la fonction motrice.

Les réponses des items de la MFM étant polytomiques avec les mêmes scores (0, 1, 2, ou 3), si les hypothèses des modèles de type TRI sont valides, un modèle approprié de la famille des modèles de Rasch est le modèle d'échelle de classement [8], son extension avec les paramètres de discrimination pouvant aussi être utilisée. Le modèle d'échelle de classement est approprié quand les scores des items ont les mêmes valeurs et significations pour tous les items; de plus, ce modèle suppose que la différence de difficulté entre deux scores consécutifs d'un même item est la même pour tous les items. Plus précisément, le modèle d'échelle de classement décompose chaque difficulté en deux termes : δ_i , dépendant de l'item i et τ_j dépendant du score j et ceci pour tout item... Néanmoins, l'hypothèse d'égalité des paramètres de discrimination de tous les items d'un test (fondamentale pour le modèle de Rasch) peut être relâchée. La pertinence de cette égalité est étudiée par une AFC. Les résultats de cette analyse doivent être examinés pour évaluer leurs pertinences cliniques.

En résumé, trois cas peuvent se présenter :

1- Les trois hypothèses fondamentales des modèles de type TRI sont valides et pour chaque domaine de la MFM, les paramètres de discrimination des items sont égaux. Une analyse de Rasch est effectuée et le score total de chaque domaine de la MFM peut être utilisé. La propriété d'objectivité spécifique (« specific objectivity »), fondamentale pour le modèle de Rasch, doit aussi être vérifiée.

2- Les trois hypothèses fondamentales des modèles de type TRI sont valides et pour chaque domaine de la MFM, les paramètres de discrimination des items ne sont pas tous égaux. Des modèles de type TRI avec des paramètres de discrimination sont utilisés. Une somme pondérée des scores est construite pour chaque domaine et les poids sont estimés à partir des paramètres de discrimination.

3- Une des trois hypothèses fondamentales des modèles de type TRI n'est pas valide et pour chaque domaine de la MFM, les paramètres de discrimination des items ne sont pas tous égaux. Des analyses ad hoc doivent être proposées ; par exemple, les modèles de type TRI non paramétriques, qui cherchent à relâcher les trois hypothèses fondamentales des modèles de type TRI paramétriques, peuvent être utilisés [121].

Dans chacun des trois cas ci-dessus, l'ajustement continue avec les étapes suivantes :

❖ **Réponses prédites vs réponses observées**

Les graphiques des résidus des modèles et les critères statistiques habituels sont utilisés. En particulier, les courbes caractéristiques des catégories des items (courbes observées et prédites) sont tracées pour évaluer la qualité de l'ajustement.

❖ **Ciblage**

Le ciblage échelle-échantillon concerne le lien entre les valeurs de la fonction motrice pouvant être mesurées par la MFM et les valeurs de la fonction motrice des patients de l'échantillon. Plus précisément sont examinées la distribution des paramètres de difficulté de la MFM et la distribution des valeurs de la variable latente (la fonction motrice) sur l'échantillon. Ceci est mis en œuvre en représentant ces deux distributions sur un même graphique : le continuum, l'absence de recouvrement et l'étendue commune des deux distributions sont examinés, et en particulier, la correspondance entre les difficultés des items et les valeurs de la variable latente des patients de l'échantillon [93]. Cette analyse montre donc comment l'échantillon est approprié pour évaluer l'échelle MFM et comment l'échelle MFM est adaptée pour mesurer la fonction motrice des patients de l'échantillon. Un meilleur ciblage implique une interprétation plus aisée et fiable des données [142]. Cette méthode, pertinente pour les modèles de Rasch, doit être adaptée si une des trois hypothèses fondamentales des modèles de type TRI n'est pas valide ou si l'hypothèse d'égalité des paramètres de discrimination des items est rejetée.

❖ **Courbes caractéristiques des catégories d'un item**

Pour chaque catégorie d'un item, la courbe caractéristique est la distribution conditionnelle de la réponse à cet item. Les courbes représentent donc la probabilité que la réponse à un item soit une catégorie donnée (en l'occurrence un score) conditionnellement à la valeur de la variable latente ; la probabilité est donnée sur un intervalle de valeurs de la variable latente. La qualité du fonctionnement d'un item est évaluée dans le sens que ses différents scores discriminent bien ou non différents niveaux de la variable latente [122]. Les graphiques des courbes caractéristiques des catégories sont produits pour chaque item et les items sont classés selon la qualité de leur fonctionnement.

❖ **Fiabilité**

La fiabilité quantifie la faiblesse de la part relative de l'erreur de mesure dans le score de l'échelle. Elle est estimée d'abord avec l'index PSI (« Person Separation Index ») [9] comparable au coefficient alpha de Cronbach [33] ; par contre, le PSI s'applique à la variable latente des modèles de type TRI et non au vrai score de la Théorie Classique des Tests (TCT) ou « Classical Test Theory , CTT ». Comme ces index sont dépendants de l'échantillon (biaisés), une méthode basée sur les modèles à équations structurelles [57] est aussi utilisée. Notons que ces index sont conçus pour une échelle dans sa globalité et non pour chaque item la composant.

Dans le cadre des modèles de type TRI, les concepts d'information d'un item et d'information de l'échelle étendent le concept de fiabilité. La fonction d'information d'une échelle ou d'un item est un index, habituellement présenté graphiquement, qui indique les valeurs de la variable latente pour lesquelles un test ou un item permet de bien différencier les patients. Autrement dit, la fonction d'information caractérise la précision de la mesure pour les patients à différents niveaux de la variable latente, une plus grande information indiquant une plus grande précision. Ces graphiques doivent être produits et doivent aider à identifier les items fonctionnant bien ou mal.

B.1.3 DIFFERENCES ENTRE GROUPES DANS LE FONCTIONNEMENT DES ITEMS ET DE L'ECHELLE

Il est important que les items de la MFM fonctionnent de manière similaire parmi différents sous-groupes de patients qui sont amenés à être comparés (par exemple sur l'âge, le sexe, le diagnostic, l'aptitude à la marche, les traitements). Le degré auquel le fonctionnement d'un item reste stable parmi différents sous-groupes est appelé le biais des items ou le fonctionnement différentiel d'un item (« Differential Item Functioning », DIF). Les deux sortes de DIF sont à évaluer : le DIF uniforme qui est analogue à un facteur de confusion, et le DIF non-uniforme qui est analogue à une interaction entre la variable latente et la variable définissant le sous-groupe. Ceci est mis en œuvre en examinant les courbes caractéristiques des items produites sur les sous-groupes et par la comparaison de modèles emboîtés [67][122].

B.2 LA VALIDATION TRANSCULTURELLE DE LA CLASSIFICATION NM-SCORE

Afin de faciliter la diffusion à l'international de la classification NM-Score, il nous reste à proposer une version en langue anglaise de cet outil.

Cette version anglaise devra fournir des mesures valides, dites équivalentes à la version originale, sans biais pouvant émaner des différences culturelles et langagières [54]. La validation transculturelle est le processus qui vise à démontrer qu'un instrument mesure les mêmes attributs dans des groupes culturels ou pays différents [127]. Elle vise à produire une adaptation de l'instrument original (incluant souvent une traduction) qui est équivalente à la version originale.

B.2.1 L'EQUIVALENCE

Dans un contexte de recherche transculturelle, l'équivalence réfère à l'évaluation non biaisée entre deux instruments de mesure traduits [29], afin que toute différence détectée soit le résultat de vraies différences entre les groupes testés et non le résultat de différences inhérentes à l'instrument utilisé [43].

Dans le processus d'adaptation et de traduction, Guillemain et al [60] recommandent aux traducteurs de considérer les dimensions suivantes :

- l'équivalence sémantique qui réfère à l'équivalence au niveau de la signification des mots ;
- l'équivalence idiomatique qui réfère à trouver des expressions équivalentes d'une langue à l'autre ;
- l'équivalence expérientielle qui réfère à ce que les situations évoquées correspondent au contexte de la culture cible ;
- l'équivalence conceptuelle qui réfère à la validité du concept exploré et aux expériences des individus de la culture cible.

Ces dimensions de l'équivalence sont à la base des processus de traduction et d'adaptation. Elles permettent de saisir les particularités entre les différentes approches d'adaptation transculturelle.

B.2.2 APPROCHES GENERALES D'ADAPTATION TRANSCULTURELLE

L'adaptation transculturelle se divise en deux composantes : la traduction et l'adaptation, c'est-à-dire une combinaison de la traduction littérale des mots et phrases d'une langue à l'autre et l'adaptation en regard des idiomes, du contexte culturel et du style de vie. Selon le cas, l'une ou les deux étapes (traduction, adaptation) doivent être réalisées. Van de Vijver et Leung [132] ont identifié trois approches générales à l'adaptation d'instruments de mesure : l'approche « decentering », la rétrotraduction et l'approche par comité.

❖ L'approche « decentering »

Elle implique la traduction d'un instrument de la langue originale vers la langue cible. Toutefois, la mesure originale est modifiée avant d'être traduite, ce qui implique que les composantes du test qui sont spécifiques à la culture originale soient modifiées ou enlevées. Ce processus est habituellement réalisé par une équipe d'individus multilingues et multiculturels qui connaissent le construit mesuré ainsi que l'instrument de mesure original [132]. Si le processus réussit, les deux versions résultantes (une dans chaque langage) seront généralement exemptes de langage et contenu spécifiques à une culture [54]. Selon Geisinger, cette approche est peu utilisée, car elle comporte plusieurs inconvénients, dont des coûts élevés ainsi que l'impossibilité d'utiliser les données recueillies à partir de la version originale dans une recherche transculturelle.

❖ La rétro traduction

Une personne ou une équipe traduit l'instrument de mesure original vers la langue cible. Ensuite, une seconde personne ou une équipe traduit la version obtenue dans la langue cible vers la langue originale, et ce sans avoir vu la version originale de l'instrument. À ce moment, les personnes planifiant l'utilisation de la traduction comparent la version originale de l'instrument avec la version rétro traduite en langue originale.

La qualité de la traduction est évaluée en fonction de la précision avec laquelle la version rétro traduite concorde avec la version originale. Selon Geisinger [54], cette approche permet entre autres d'évaluer la qualité de la traduction. Toutefois, elle ne permet pas la réalisation de changements à l'instrument afin de s'adapter aux différences culturelles. Selon van de Vijver et Leung [132], cette procédure s'attarde davantage à la sémantique qu'à la signification et à l'intelligibilité.

❖ L'approche par comité

Elle consiste à ce qu'un groupe d'individus bilingues traduise l'instrument de mesure de la langue originale vers la langue cible. Les membres du comité doivent également être familiers avec les deux cultures, le ou les construits mesurés et les principes généraux de la mesure. Bien que plus coûteux qu'un traducteur unique et pouvant occasionner des mésententes, le comité permet d'identifier les erreurs commises par un membre et permet l'échange d'expertises [132]. Puisque l'approche par comité ne permet pas une évaluation indépendante de son efficacité, Geisinger [54] recommande son utilisation avec la rétro traduction. De plus Flaherthy et al. [49] estiment que la clé pour établir l'équivalence sémantique est la technique de rétro traduction incluant la révision en comité d'experts.

B.2.3 METHODOLOGIE PROPOSEE POUR L'ADAPTATION ET LA VALIDATION TRANSCULTURELLE

Le tableau reprend les différentes étapes que nous proposons pour le processus de validation transculturelle de notre outil NM-Score.

Etapes	Calendrier	Version de la classification
1^{ère} traduction	12/2011-03/2012	VA1 (utilisée pour l'article)
2^{ème} traduction	04/2012-05/2012	VA2
Approche de type comité	06/2012	VA3
Retro traduction de VA3	07/2012-09/2012	VF3
Approche de type comité	09/2012	VA4
Etude de la fidélité inter observateur Etude de la validité contre critère	10/2012-03/2013	VA5 = VA finale

*Tableau 31 : Etapes de la validation transculturelle du NM-Score
VA : version anglaise VF ; version française*

Guillemin et al. [60] proposent des lignes directrices au processus d'adaptation transculturelle. La méthodologie proposée a été développée à partir d'une recension des écrits concernant la traduction de mesures de la qualité de vie ainsi qu'une étude de fidélité inter-juge. Elle comporte cinq étapes soit (a) la réalisation d'une ou de plusieurs traductions, (b) la réalisation d'une ou de plusieurs rétro traductions (c) l'élaboration d'une version finale par la révision en comité des traductions et des rétro traductions (d) la réalisation d'un prétest auprès de la population cible et (e) l'étude de la pondération des notations en fonction de la culture cible.

Guillemin et al. [60] rapportent que l'on peut croire que la réalisation complète de l'équivalence transculturelle confère l'équivalence aux propriétés psychométriques de la version source.

Vallerand [128], quant à lui, considère que la traduction d'un instrument de mesure n'assure en rien la fidélité et la validité de l'instrument dans la culture cible. À l'instar de Guillemain et al. [60], il recommande que plusieurs traductions et autant de rétro traductions soient réalisées et qu'elles le soient de façon indépendante les unes des autres. Ceci a pour but de mettre en relief les lacunes des traductions et permet de détecter les erreurs et les divergences d'interprétation. Guillemain et Vallerand s'entendent pour dire que le travail en comité multidisciplinaire (chercheurs, experts dans le domaine d'intérêt, personnes ayant participé aux traductions, linguistes, auteurs de l'instrument original, représentants de la culture cible) est essentiel pour diminuer les biais associés au chercheur seul. Les étapes ultérieures, proposées par Vallerand, visent la réalisation d'un prétest et l'étude des qualités psychométriques de la version expérimentale, plus spécifiquement l'étude de la validité concomitante et de contenu, l'analyse de la fidélité, l'évaluation de la validité de construit et l'établissement de normes.

Les différentes méthodologies présentées comportent de nombreuses similarités quant au processus d'adaptation transculturelle. Toutes proposent d'utiliser la rétro traduction et l'approche par comité suite à la traduction. Les spécificités de chacune des méthodologies concernant l'adaptation sont principalement d'ordre pratique. Toutefois, pour Vallerand [128] et Geisinger [54], la validation transculturelle d'un instrument de mesure nécessite la réalisation de plusieurs étapes qui s'étendent bien au-delà de l'adaptation transculturelle. L'étude des qualités psychométriques peut être effectuée tant dans l'approche TCT que dans la TRI. Les méthodologies varieront en fonction de l'approche choisie. Éventuellement, le processus doit intégrer des démarches visant à associer la version adaptée à la version source afin de conférer l'équivalence conceptuelle de l'adaptation réalisée.

CONCLUSIONS

Les progrès récents de la recherche fondamentale autorisent un espoir thérapeutique dans plusieurs pathologies neuromusculaires de l'enfance. Les essais thérapeutiques, obligatoirement multicentriques et parfois internationaux, obligent à une uniformisation des outils d'évaluations. Ces dernières années de nombreux outils ont été développés par différentes équipes, avec un effort fait de la part des professionnels pour s'intéresser à la précision de la mesure.

La MFM présente des qualités métrologiques intéressantes en termes de fidélité, de validité mais aussi de sensibilité au changement. Sa reconnaissance et sa diffusion internationale en plein essor sont favorisées par une méthodologie précise appliquée à son développement et à sa validation. La prochaine étape pour consolider son utilisation dans les essais thérapeutiques concerne son étude en appliquant des modèles issus de la théorie de réponse aux items (telle que l'analyse selon le modèle de Rasch).

Nous proposons à l'issue de ce travail un outil complémentaire aux différents outils déjà disponibles : le NM-Score. Cette classification, utile en recherche clinique comme pour le suivi habituel des patients, permet une description détaillée du statut fonctionnel des patients. Il s'agit d'une mesure des performances des patients dans leur vie quotidienne permettant une approche plus écologique des incapacités fonctionnelles. Sa fidélité inter-observateur satisfaisante autorise son utilisation multicentrique. Les NM-Score D1 et D2 sont corrélés de façon forte avec les gold standards utilisés (scores de Vignos, de Brooke, de la MFM) et l'analyse de la validité des construits pour ces 2 domaines confirment ce résultat. Pour le NM-score D3, les résultats sont plus difficiles à interpréter. Des analyses complémentaires seront nécessaires pour différencier entre imprécision de la mesure et limites en lien avec la nature même de la variable latente. La validation d'une version anglaise est la prochaine étape et autorisera, nous l'espérons, une utilisation étendue de cette classification.

L'exercice de la médecine nécessite une quantification précise de l'effet de nos actions thérapeutiques. Il doit concilier art et science comme Ambroise Paré l'avait perçu ; il insiste sur le fait, qu'en médecine, nous devons prendre en considération "*la théorie apprise à l'école, l'expérience et sa propre pratique*". La médecine est en constante évolution en "*Les arts*", dit Ambroise Paré "*ne sont encore si accomplis qu'on n'y puisse faire addition. Ils se parfent et polissent par succession de temps*". Selon le serment d'Hippocrate, le médecin doit "*appliquer ses capacités et son jugement à l'accomplissement de son art*".

La quantification en médecine ne relève pas d'une approche mécanistique de l'humain ou le médecin n'est qu'un scientifique. Au contraire, il s'agit d'une approche prenant en compte la singularité de la personne humaine et de son fonctionnement. Dans un objectif de qualité des soins pour le patient, l'évaluation a été progressivement intégrée à l'exercice quotidien de la médecine, changeant les mentalités. Des preuves sont maintenant « exigées » pour justifier le choix de tel ou tel traitement, tout en tenant compte des préférences du patient. La mesure est nécessaire en médecine avant tout car l'humain mérite rigueur.

Dans les années à venir, il faudra rester vigilant pour que, malgré les progrès scientifiques annoncés, la pratique médicale face toujours appel au bon sens et aux qualités d'observation et

de discernement et que le patient, dans son caractère unique et complexe, reste au centre des préoccupations et des débats [78].

S'intéresser à l'évaluation et à la mesure en médecine, c'est avant tout faire preuve d'une rigueur indispensable aux décisions de soins touchant des personnes vulnérables aux besoins spécifiques.

REFERENCES

- [1] Abresch RT, Seyden NK, Wineinger MA. Quality of life. Issues for persons with neuromuscular diseases. *Phys Med Rehabil Clin N Am* 1998;9:233-248.
- [2] Abresch RT, Carter GT, Han JJ, McDonald CM. New clinical end points in rehabilitation medicine: tools for measuring quality of life. *Am J Hosp Palliat Care* 2009 Dec-2010 Jan;26(6):483-92.
- [3] Ahlström G, Gunnarsson LG. Disability and quality of life in individuals with muscular dystrophy. *Scand J Rehab Med* 1996;28:147-157.
- [4] Allsop KG, Ziter FA. Loss of strength and functional decline in Duchenne's dystrophy. *Arch Neurol* 1981;38:406-11.
- [5] Anastasi, A. Titre, Introduction à la psychométrie. Édition, Montréal: Guérin, 1994.
- [6] Andre JM. Handicap : Définitions, principes de traitement, processus de récupération. *Bull Acad Natl Med* 1994;178(7):1301-16.
- [7] Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000; 81(12 suppl 2): 15-20.
- [8] Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978; 43, 561-73.
- [9] Andrich D. An index of person separation in latent trait theory, the traditional KR20 index, and the Guttman scale response pattern. *Educational and Psychological Research* 1982;9(1), 95-104.
- [10] ATS committee on proficiency standards for clinical pulmonary function laboratories. Ats statement : guidelines for the 6 minutes walk test. *Am J respir Crit Care Med* 2002;166: 111-17.
- [11] Barois A, Mayer M, Desguerre I, Chabrol B, Berard C, Cuisset JM et al. A 4-year prospective, multicenter, longitudinal study (168 cases). *Bull Acad Natl Med* 2005 Jun;189(6):1181-98.
- [12] Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol* 2001 Dec;54(12):1204-17.
- [13] Benaïm C, Sacconi S, Fournier-Mehouas M, Tanant V, Desnuelle C. Analyse de la validité de la MFM en pratique clinique de consultation adulte d'un centre de référence pour maladies neuromusculaires. *Rev Neurol (Paris)* 2010 Jan;166(1):49-53.
- [14] Berard C, Payan C, Hodgkinson I, Fermanian J, and the MFM collaborative study group. A motor function measure scale for neuromuscular diseases. Construction and validation study. *Neuromuscul Disord* 2005;15: 463-70.

- [15] Berard C, Payan C, Fermanian J, Girardot F, groupe d'étude MFM. La mesure de fonction motrice, outil d'évaluation de la fonction motrice. *Rev Neurol (Paris)* 2006; 162: 485–93.
- [16] Berard C, Fermanian J, Payan C. Outcome measure for SMA II and III patients. *Neuromuscul Disord* 2008 Jul;18(7):593-4; author reply 594-5. Erratum in: *Neuromuscul Disord* 2008 Nov;18(11):911.
- [17] Beretta R. A critical review of the Delphi technique. *Nurs Res* 1996;3(4):79-89.
- [18] Bernier JJ, Pietrulewicz B. *La psychométrie: Traité de mesure appliquée*. Montréal : Gaétan Morin, 1997.
- [19] Bertini E, Burghes A, Bushby K, Estournet-Mathiaud B, Finkel AS, Hughes RAC [et al]. 134th ENMC International Workshop: Outcome measures and treatment of spinal muscular atrophy, 11-13 February 2005, Naarden, The Netherlands. *Neuromuscul Disord* 2005 Nov;15(11):802-16.
- [20] Bickenbach JE, Chatterji S, Badley EM, Ustün TB. Model of disablement, universalism and the ICIDH. *Social science and medicine*, 1999,48:1173-87.
- [21] Birnbaum A. Test scores, sufficient statistics, and the information structures of tests. In Lord L and Novick M eds., *Statistical Theories of Mental Test Scores*, AddisonWesley: Reading, MA, 1968. pp:425-435
- [22] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurements. *Lancet* 1986 Feb 8 ;1(8476):307–10.
- [23] Bönnemann CG, Rutkowski A, Mercuri E, Muntoni F, for the CMD Outcomes Consortium. 173rd ENMC international workshop: Congenital muscular dystrophy outcome measures 5–7 March 2010, Naarden, The Netherlands. *Neuromuscular Disorders* 2011; 21, 513-22.
- [24] 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 May-Jun;4(3):186-97.
- [25] 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.
- [26] 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; CD006282.

- [27] Bushby KM, Beckmann JS. The limb-girdle muscular dystrophies--proposal for a new nomenclature. *Neuromuscul Disord* 1995 Jul;5(4):337-43.
- [28] Cano S.J., Mayhew A., Joyce C., Kissel J.T. on behalf of the International Coordinating Committee for SMA Clinical Trials Rasch Task Force (2011). Are motor performance rating scales ready for clinical trials in Spinal Muscular Atrophy? Insights from a Rasch analysis. Submitted to *Neurology*.
- [29] Cella, DF, Lloyd SR, Wright BD. Cross-cultural instrument equating : Current research and future directions. Dans B. Spilker (Éd.), *Quality of life and pharmacoeconomics in clinical trials* (pp. 707-715). Philadelphia : Lippincott-Raven,1996.
- [30] Cella D, Chang CH. A Discussion of Item Response Theory and Its Applications in Health Status Assessment. *Medical Care* 2000; 38(9) supp 2: 66-72.
- [31] Charmet E, Bethoux F, Camels P, Gautheron V, Minaire P. MIF Môme : étude de la reproductibilité et analyse d'une population témoin d'enfants sains de 1 à 9 ans (échantillon de 167 sujets). *Ann Réadaptation Med* 1996;39:15-9.
- [32] Chung BH, Wong VC, Ip P. Spinal muscular atrophy: Survival pattern and functional status. *Pediatrics* 2004;114:548-53.
- [33] Cronbach L. Coefficient alpha and the internal structure of the test. *Psychometrika* 1951 ; 16(3) :297-334
- [34] Davis SE, Hynan LS, Limbers CA, Andersen CM, Greene MC, Varni JW et al. The PedsQL in pediatric patients with Duchenne muscular dystrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Neuromuscular Module and Generic Core Scales. *J Clin Neuromuscul Dis* 2010 Mar;11(3):97-109.
- [35] Dazord A. Prendre en compte le vécu intérieur pour définir la qualité de vie. In : Gardou C (ed), *La personne handicapée : d'objet à sujet, de l'intention à l'acte. Connaître le handicap, reconnaître la personne*. Ramonville-Sainte-Agne: Erès, 1999. p87-91.
- [36] de Lattre C, Payan C, Vuillerot C, de Castro D, Bérard C. Motor Function measure : validation of a short form (MFM-20) for young children with neuromuscular diseases. Submitted 29022012 *Dev Med Child Neurol*.
- [37] de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and a minimally important change. *Health Qual Life Outcomes* 2006;4:54.
- [38] Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis* 1986;39(11):897-906.

- [39] Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991 Aug;12(4 Suppl):142S-158S
- [40] Domino G. Psychological testing. New Jersey : Prentice-Hall, 2000.
- [41] Dupage D. Qualité et santé. Première partie : qualité des moyens ou qualité des résultats ? *Medecine* 2008 ; oct :360-3.
- [42] Eagle M, Scott E, Main M et al. G.P.1.08 Steroids in Duchenne muscular dystrophy (DMD): Natural history and clinical evaluation using the North Star Ambulatory Assessment (NSAA); *Neuromuscular Disorders* 2007; 17(9): 774.
- [43] Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation and the Health Professions* 2005, 28(2), 212-232.
- [44] Fallissard B. Mesurer la subjectivité en santé : perspective méthodologique et statistique. Paris : Masson, 2001. 216p.
- [45] Fallissard B. Valeur scientifique des mesures de qualité de vie et autres mesures subjectives réalisées en recherche clinique ? E-mémoires de l'Académie Nationale de Chirurgie 2004;3(1):19-23.
- [46] Fermanian J. Validation des échelles d'évaluation en médecine physique et de réadaptation : comment apprécier correctement leurs qualités psychométriques. *Ann Readapt Med Phys* 2005;48(6):281-7.
- [47] Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health* 1984 Sep;74(9):979-83.
- [48] Finkel RS, Hynan LS, Glanzman AM, Owens H, Nelson L, Cone SR et al; AmSMART Group. The test of infant motor performance: reliability in spinal muscular atrophy type I. *Pediatr Phys Ther* 2008;20(3):242-6.
- [49] Flaherty JA, Gaviria FM, Pathak D, Mitchell T, Wintrob R, Richman JA et al. Developing instruments for cross-cultural psychiatric Research. *Journal of Nervous and Mental Disease* 1988; 176(5) : 257-63.
- [50] Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 1973;33, ;613-19.
- [51] Florence JM, Pandya S, King WM et al. Clinical Trials in Duchenne Dystrophy: Standardization and Reliability of Evaluation Procedures. *PhysTher* 1984; 64 (1): 41-45.

- [52] Forer S, Granger CV. Functional Independence Measure. Buffalo NY: The Buffalo General Hospital, State University of New York at Buffalo, 1987.
- [53] Geiger R, Strasak A, Trembl B, Gasser K, Kleinsasser A, Fischer V, et al. Six-minute walk test in children and adolescents. *J Pediatr* 2007;150:395–99.
- [54] Geisinger K. Testing and assessment in cross-cultural psychology. Dans J. R. Graham & J. A. Naglieri, (Éds.), *Handbook of psychology* (pp.95-117). New York: John Wiley & Sons, 2003.
- [55] Glanzman AM, Mazzone E, Main M, Pelliccioni M, Wood J, Swoboda KJ et al. The Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND): test development and reliability. *Neuromuscul Disord* 2010 Mar;20(3):155-61.
- [56] Glanzman AM, McDermott MP, Montes J, Martens WB, Flickinger J, Riley S et al; Pediatric Neuromuscular Clinical Research Network for Spinal Muscular Atrophy (PNCr); Muscle Study Group (MSG). Validation of the Children's Hospital of Philadelphia Infant Test of Neuromuscular Disorders (CHOP INTEND). *Pediatr Phys Ther* 2011 Winter;23(4):322-6.
- [57] Green SA, Yang Y. Reliability of summed item scores using structural equation modeling: an alternative to coefficient alpha. *Psychometrika* 2009; 74, 155-67.
- [58] Grootenhuis MA, de Boone J, van der Kooij AJ. Living with muscular dystrophy: health-related quality of life consequences for children and adults. *Health Qual Life Outcomes* 2007;5:31.
- [59] Groupe de travail 2 du Comité commun pour les guides en métrologie (JCGM/WG 2). Vocabulaire international de métrologie (VIM) : Concepts fondamentaux et généraux et termes associés, 2008. Disponibles sur http://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2008.pdf
- [60] Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures : Literature review and proposed guidelines. *Journal of Clinical Epidemiology* 1993;46, 1417-32.
- [61] Guinvarc'h S. Une échelle internationale de mesure de la fonction motrice pour les maladies neuro-musculaires 1. Revue de la littérature 2 Premières étapes de construction. Mémoire de DEA : université de Bourgogne 1998.
- [62] Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR; Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002 Apr;77(4):371-83. Review
- [63] Hambleton RK, van der Linden WJ. Advances in item response theory and applications: An introduction. *Applied Psychological Measurement* 1982; 6, 373-8.

- [64] Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement : Issues and Practice* 1993; 12, 38-47.
- [65] Hambleton RK. Emergence of Item Response Modeling in Instrument Development and Data Analysis. *Medical Care* 2000; 38 supp 2: 60-65.
- [66] Hambleton RK. Good practices for identifying differential item functioning. *Medical Care* 2006;44(11): S182-188.
- [67] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000 Oct;32(4):1008-15.
- [68] Hays RD, Morales LS , Reise SP. Item Response Theory and Health Outcomes Measurement in the 21st Century. *Medical Care* 2000; 38(9)Supp 2: 28–42.
- [69] Iannaccone ST; American Spinal Muscular Atrophy Randomized Trials (AmSMART) Group. Outcome measures for pediatric spinal muscular atrophy. *Arch Neurol* 2002 Sep;59(9):1445-50.
- [70] Iannaccone ST, Hynan LS. Reliability of 4 outcome measures in pediatric spinal muscular atrophy. *Arch Neurol* 2003;60:1130–6.
- [71] Iannaccone ST, Hynan LS, Morton A, Buchanan R, Limbers CA, Varni JW; AmSMART Group. The PedsQL in pediatric patients with Spinal Muscular Atrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Generic Core Scales and Neuromuscular Module. *Neuromuscul Disord* 2009 Dec;19(12):805-12
- [72] International classification of Impairments, Disabilities, and Handicaps:a manual of classification relating to the consequences of diseases. Geneva, Switzerland: World Health Organisation, 1980.
- [73] International statistical classification of diseases and related health problems, 10ème révision, Vol. 1-3. Geneva, Switzerland : World Health Organisation, 1992-1994.
- [74] International Classification of Functioning, Disability and Health-ICF: Geneva, Switzerland : World Health Organisation, 2001.
- [75] International Classification of Functioning, Disability and Health: children and youth version. Geneva, Switzerland: World Health Organisation, 2007.
- [76] Kissel JT, Scott CB, Reyna SP, Crawford TO, Simard LR, Krosschell KJ et al. SMA CARNIVAL TRIAL PART II: a prospective, single-armed trial of L-carnitine and valproic acid in ambulatory children with spinal muscular atrophy.; Project Cure Spinal Muscular Atrophy Investigators' Network. *PLoS One* 2011;6(7):e21296.

- [77] Kohler M, Clarenbach CF, Böni L, et al. Quality of life, physical disability, and respiratory impairment in Duchenne muscular dystrophy. *Am J Respir Crit Care Med* 2005;172:1032-1036.
- [78] Kremer R. Médecine : concilier l'art et la science. Université catholique de Louvain. Disponible sur <http://sites.uclouvain.be/ama-ucl/concilierhuit.htm>
- [79] Krosschell KJ, Maczulski JA, Crawford TO, Scott C, Swoboda KJ. A modified Hammersmith functional motor scale for use in multi-center research on spinal muscular atrophy. *Neuromuscul Disord* 2006 Jul;16(7):417-26.
- [80] Krosschell KJ, Scott CB, Maczulski JA, Lewelt AJ, Reyna SP, Swoboda KJ; Project Cure SMA. Reliability of the Modified Hammersmith Functional Motor Scale in young children with spinal muscular atrophy. *Muscle Nerve* 2011 Aug;44(2):246-51.
- [81] Legrand-Persoz M, Persoz B, Boulvert E, Gros P, Dardel JP, Arhan P. Surveillance de l'évolution musculaire respiratoire dans la myopathie de Duchenne de Boulogne à évolution rapide : application au contrôle thérapeutique. *Ann Med Phys*1979; 23:1-14.
- [82] Lord JP, Lieberman JS, Portwood MM, Fowler WM Jr, Carson R. Functional ability and equipment use among patients with neuromuscular disease. *Arch Phys Med Rehabil* 1987; 68(1):8-9.
- [83] Lord JP, Portwood MM, Fowler WM, Lieberman JS, Carson R. Upper vs lower extremity functional loss in neuromuscular disease. *Arch Phys Med Rehabil* 1987; 68(6):348-52.
- [84] Lue YJ, Lin RF, Chen SS, Lu YM .Measurement of the functional status of patients with different types of muscular dystrophy. *Kaohsiung J Med Sci* 2009 Jun;25(6):325-33.
- [85] Lue YJ, Su CY, Yang RC, Su WL, Lu YM, Lin RF, Chen SS. Development and validation of a muscular dystrophy-specific functional rating scale. *Clin Rehabil* 2006 Sep;20(9):804-17.
- [86] Lyman HB. Test scores and what they mean ? Boston: Allyn and Bacon, 1998.
- [87] McDonald CM, Abresch RT, Carter GT, Fowler WM Jr, Johnson ER, Kilmer DD et al. Profiles of neuromuscular diseases. Duchenne muscular dystrophy. *Am J Phys Med Rehabil* 1995 Sep-Oct;74(5 Suppl):S70-92.
- [88] McDonald CM, Henricson EK, Han JJ, Abresch RT, Nicorici A, Atkinson L et al. The 6-minute walk test in Duchenne/Becker muscular dystrophy: longitudinal observations. *Muscle Nerve* 2010 Dec;42(6):966-74.
- [89] McDonald CM, Henricson EK, Han JJ, Abresch RT, Nicorici A, Elfring GL et al. The 6-minute walk test as a new outcome measure in Duchenne muscular dystrophy. *Muscle Nerve* 2010 Apr;41(4):500-10.

- [90] Mah JK, Thannhauser JE, Kolski H, Dewey D. Parental stress and quality of life in children with neuromuscular disease. *Pediatr Neurol* 2008;39:102–7.
- [91] Main M, Kairon H, Mercuri E, Muntoni F. The Hammersmith functional motor scale for children with spinal muscular atrophy: a scale to test ability and monitor progress in children with limited ambulation. *Eur J Paediatr Neurol* 2003;7(4):155-9.
- [92] Matillon Y. L'évaluation en médecine. Une démarche en plein développement. *Rev Prat* 1992;42(15):1917–19.
- [93] Mayhew A, Cano S, Scott E, Eagle M, Bushby K, Muntoni F (2011). Moving towards meaningful measurement: Rasch analysis of the North Star Ambulatory Assessment in Duchenne muscular dystrophy. *Developmental Medicine & Child Neurology* 2011; 53(6), 535-42.
- [94] Mazzone ES, Messina S, Vasco G, Main M, Eagle M, D'Amico A, et al. Reliability of the North Star Ambulatory Assessment in a multicentric setting. *Neuromuscul Disord*. 2009 Jul;19(7):458-61.
- [95] Mazzone E, Martinelli D, Berardinelli A, Messina S, D'Amico A, Vasco G et al. North Star Ambulatory Assessment, 6-minute walk test and timed items in ambulant boys with Duchenne muscular dystrophy. *Neuromuscul Disord* 2010 Nov;20(11):712-6
- [96] Mazzone E, Bianco F, Martinelli D, Glanzman AM, Messina S, De Sanctis R et al. Assessing upper limb function in nonambulant SMA patients: development of a new module. *Neuromuscul Disord* 2011 Jun;21(6):406-12.
- [97] Mazzone E, Vasco G, Sormani MP, Torrente Y, Berardinelli A, Messina S et al. Functional changes in Duchenne muscular dystrophy: a 12-month longitudinal cohort study. *Neurology* 2011 Jul 19;77(3):250-6.
- [98] Mercuri E, Messina S, Battini R, Berardinelli A, Boffi P, Bono R et al. Reliability of the Hammersmith functional motor scale for spinal muscular atrophy in a multicentric study. *Neuromuscul Disord* 2006;16(2):93-8.
- [99] Mercuri E, Bertini E, Messina S, Solari A, D'Amico A, Angelozzi C et al. Randomized, double-blind, placebo-controlled trial of phenylbutyrate in spinal muscular atrophy. *Neurology* 2007;68(1):51-5.
- [100] Mercuri E, Mayhew A, Muntoni F, Messina S, Straub V, Van Ommen GJ et al; TREAT-NMD Neuromuscular Network. 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 Nov;18(11):894-903.

- [101] Mercuri E. Choosing the right clinical outcome measure : From the patient to the statistician and back. *Neuromuscul Disord* 2011 July ;21 :16-9.
- [102] Montes J, McDermott MP, Martens WB, Dunaway S, Glanzman AM, Riley S et al; Muscle Study Group and the Pediatric Neuromuscular Clinical Research Network. Six-Minute Walk Test demonstrates motor fatigue in spinal muscular atrophy. *Neurology*. 2010 Mar 9;74(10):833-8.
- [103] Morin E. Le besoin d'une pensée complexe, in 1966-1996, La passions des idées », Magazine littéraire, Hors Série, décembre 1996.
- [104] Nelson L, Owens H, Hynan LS, Iannaccone ST. The gross motor function measure is a valid and sensitive outcome measure for spinal muscular atrophy. *Neuromuscul Disord* 2006;16:374–80.
- [105] O'Hagen JM, Glanzman AM, McDermott MP, Ryan PA, Flickinger J, Quigley J et al. An expanded version of the Hammersmith Functional Motor Scale for SMA II and III patients. *Neuromuscul Disord* 2007 Oct;17(9-10):693-7.
- [106] 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 Apr;39(4):214-23.
- [107] Paolaggi JB, Coste J. Le raisonnement medical de la science à la pratique clinique. Edition Estem 2001.
- [108] Pradhan S, Ghosh D, Srivastava NK, Kumar A, Mittal B, Pandey CM, et al. Prednisolone in Duchenne muscular dystrophy with imminent loss of ambulation. *J Neurol* 2006;253:1309–1316.
- [109] Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen: Danmarks Paedagogiske Institut, 1960.
- [110] Reitter B. Discussion on the CIDD protocol. Copenhagen, Seminar on evaluation of rehabilitation in a lifelong course, Mars 1998.
- [111] Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR and Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes* 2006, 4:70-5.
- [112] Rodary C, Pezet-Langevin V, Kalifa C. Qu'est ce qu'un bon outils d'évaluation ? *Arch Pediatr* 2001;8:744-50.

- [113] Roush J, Guy J, Purvis M. Reference values and relationship of the six minute walk test and body mass index in healthy third grade school children. *Internet J Allied Health Sci Pract* 2006; 4:1–6.
- [114] Russell DJ, Rosenbaum PL, Cadman DT, Gowland C, Hardy S, Jarvis S. The gross motor function measure: a means to evaluate the effects of physical therapy. *Dev Med Child Neurol*. 1989 Jun;31(3):341-52.
- [115] Russman BS, Buncher CR, White M, Samaha FJ, Iannaccone ST. Function changes in spinal muscular atrophy II and III. The DCN/SMA Group. *Neurology* 1996; 47: 973-6.
- [116] Russman BS. Spinal muscular atrophy: clinical classification and disease heterogeneity. *J Child Neurol* 2007; 22: 946-51.
- [117] Sackett DL, Rosenberg W, Gray M, Haynes B, Richardson S. Evidence based medicine: what it is and what it isn't. *BMJ* 1996 ; 312 (7023) :71-2.
- [118] Scott E, Mawson SJ. Measurement in Duchenne muscular dystrophy: considerations in the development of a neuromuscular assessment tool. *Dev Med Child Neurol* 2006 Jun;48(6):540-4.
- [119] Scott E, Eagle M, Mayhew A, Freeman J, Main M, Sheehan J and al; The North Star Clinical Network for Paediatric Neuromuscular Disease. Development of a Functional Assessment Scale for Ambulatory Boys with Duchenne Muscular Dystrophy. *Physiother Res Int* 2011 Sep 23. doi: 10.1002/pri.520. [Epub ahead of print].
- [120] Scott OM, Hyde SA, Goddard C, Dubowitz V. Quantitation of muscle function in children: a prospective study in Duchenne muscular dystrophy. *Muscle Nerve* 1982; 5(4):291-301.
- [121] Sijtsma K, Junker, BW. Item response theory: Past performance, present developments, and future expectations. *Behaviormetrika* 2006;33 (1): 75-102.
- [122] Skrondal A, Rabe-Hesketh S. Generalized Latent Variable Modeling. Multilevel, Longitudinal, and Structural Equation Models (Interdisciplinary Statistics). Chapman and Hall / CRC, 2004.
- [123] Steffensen BF, Hyde S, Lyager S, Mattsson E. Validity of the EK scale: a functional assessment of non-ambulatory individuals with Duchenne muscular dystrophy or spinal muscular atrophy. *Physiother Res Int* 2001;6(3):119-34
- [124] Steffensen BF, Lyager S, Werge B, Rahbek J, Mattsson E. Physical capacity in non-ambulatory people with Duchenne muscular dystrophy and spinal muscular atrophy. A longitudinal study. *Dev Med Child Neurol* 2002; 44:623-32.
- [125] Sumsion T. The Delphi Technique: an adaptive research tool. *Br J Occup Ther* 1998;61(4):153-6

- [126] Swoboda KJ, Scott CB, Reyna SP, Prior TW, LaSalle B, Sorenson SL et al. Phase II open label study of valproic acid in spinal muscular atrophy. *PLoS One*. 2009;4(5):e5268.
- [127] Tang ST, Dixon J. Instrument translation and evaluation of equivalence and psychometric properties: The Chinese sense of coherence scale. *Journal of Nursing Measurement* 2002 ;70(1) : 59-76.
- [128] Vallerand, R. J. Vers une méthodologie de validation transculturelle de questionnaires psychologiques: Implications pour la recherche en langue française. *Psychologie Canadienne* 1989 ; 30 : 662-80.
- [129] Vandervelde L, Van den Bergh PYK, Thonnard JL et al. ACTIVLIM: A rasch-built measure of activity limitations in children and adults with neuromuscular disorders. *Neuromuscul disord* 2007; 17:459-69.
- [130] Vandervelde L, Dispa D, Van den Bergh PYK et al. Comparison between self-perceived and observed activity limitations in adults with neuromuscular disorders. *Arch phys med rehab* 2008;89:1720-3 .
- [131] Vandervelde L, Van den Bergh PY, Penta M, Thonnard JL. Validation of the ABILHAND questionnaire to measure manual ability in children and adults with neuromuscular disorders. *J Neurol Neurosurg Psychiatry* 2010 May;81(5):506-12.
- [132] Van der Vijver F, Leung K. *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA : Sage, 1997.
- [133] Varni JW, Seid M, Rode CA. The PedsQLTM: measurement model for the Pediatric Quality of Life Inventory. *Med Care* 1999;37:126–39.
- [134] Varni JW, Seid M, Kurtin PS. PedsQLTM 4.0: reliability and validity of then Pediatric Quality of Life InventoryTM Version 4.0 Generic Core Scales in healthy and patient populations. *Med Care* 2001;39:800–12.
- [135] Velozo 0A, Forsyth K, Kielhofner G (2006). Objective measurement : The influence of item response theory on reasearch and practice. Dans G. Kielhofner, (Ed.), *Research in Occupational Therapy : Methods of Inquiry for Enhancing Practice* (pp. 177-200). Philadelphie : F. A. Davis.
- [136] Vignos PJ, Spencer GE, Archibald KC. Management of progressive muscular dystrophy of childhood. *J Am Med Assoc* 1963; 184:89-96.
- [137] Vincent KA, Carr AJ, Walburn J, Scott DL, Rose MR. Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL). *Neurology* 2007 Mar 27;68(13):1051-7.

- [138] Vuillerot C, Hodgkinson I, Bissery A, Schott-Pethelaz AM, Iwaz J, Ecochard R, D'Anjou MC, Commare MC, Berard C.. Quality of Life of Adolescents with Neuromuscular Diseases: Here's What They Say. *J Adol Health* 2009;46(1):70-6.
- [139] Vuillerot C, Girardot F, Payan C, Fermanian J, Iwaz J, De Lattre C, Berard C. Monitoring changes and predicting loss of ambulation in Duchenne muscular dystrophy with the Motor Function Measure. *Dev Med Child Neurol* 2010 Jan;52(1):60-5.
- [140] Wang CH, Finkel RS, Bertini ES et al. Consensus statement for standard of care in spinal muscular atrophy. *J Child Neurol* 2007; 22: 1027-49.
- [141] Wong BL, Hynan LS, Iannaccone ST. A randomized, placebo-controlled trial of creatine in children with spinal muscular atrophy. *J Clin Neuromuscul Dis* 2007;8:101–10.
- [142] Wright BD, Masters G. Rating scale analysis: Rasch measurement. Chicago: MESA, 1982
- [143] Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil* 1989;70:857-60.
- [144] Young HK, Lowe A, Fitzgerald DA, et al. Outcome of noninvasive ventilation in children with neuromuscular disease. *Neurology* 2007;68:198–201.
- [145] Zerres K, Rudnik-Schoneborn S. Natural history in proximal spinal muscular atrophy. Clinical analysis of 445 patients and suggestions for a modification of existing classifications. *Arch Neurol* 1995; 52: 518-23.
- [146] Zerres K, Wirth B, Rudnik-Schoneborn S. Spinal muscular atrophy--clinical and genetic correlations. *Neuromuscul Disord* 1997; 7: 202-7.

ANNEXES