



Validity and Reliability of the 32-Item Motor Function Measure in 2- to 5-Year-Olds with Neuromuscular Disorders and 2- to 25-Year-Olds with Spinal Muscular Atrophy

Dylan Trundell · Stephanie Le Scouiller · Ksenija Gorni ·

Timothy Seabrook · Carole Vuillerot on behalf of the SMA MFM Study Group

Received: May 30, 2020 / Published online: August 27, 2020
© The Author(s) 2020

ABSTRACT

Introduction: To investigate the validity and reliability of the 32-item Motor Function Measure (MFM32) in individuals with neuromuscular disorders (NMD), including spinal muscular atrophy (SMA), aged 2–5 years, and in non-ambulant individuals with Types 2 or 3 SMA, aged 2–25 years.

Methods: Test–retest reliability (intraclass correlation coefficient [ICC]), internal consistency (Cronbach's alpha [α]), convergent validity (Spearman rank-order correlations), and known-groups validity (analysis of covariance comparing groups defined by the Clinical Global Impression of Severity [CGI-S] scale and Vignos grade) were calculated. The analysis was performed on a dataset provided by Hospices

Civils De Lyon, extracted from the multinational MFM32 database. A total of 165 individuals were included in the analyses, of whom 84 were in the NMD group (aged 2–5 years) and 81 were in the SMA group (aged 2–25 years).

Results: Strong evidence of test–retest reliability (ICC: 2- to 5-years' population = 0.94–0.95; 2- to 25-years' population = 0.97), internal consistency (Cronbach's α : 2- to 5-years' population = 0.96; 2- to 25-years' population = 0.95), convergent validity (2- to 5-years' population: CGI-S rho = – 0.84, Vignos grade rho = – 0.79; 2- to 25-years' population: CGI-S rho = – 0.49), and known-groups validity (all $P < 0.001$) were demonstrated.

Conclusions: These analyses provide supportive evidence of the validity and reliability of the MFM32 in younger individuals with NMDs, aged 2–5 years, and in non-ambulant individuals with Types 2 or 3 SMA, aged 2–25 years, supporting the use of the MFM32 across a wide age range.

Digital Features To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.12728270>.

D. Trundell (✉) · S. Le Scouiller
Roche Products Limited, Welwyn Garden City, UK
e-mail: dylan.trundell@roche.com

K. Gorni · T. Seabrook
Rare Diseases–Pharma Research and Early
Development, F. Hoffmann-La Roche, Basel,
Switzerland

C. Vuillerot
Department of Pediatric Physical Medicine and
Rehabilitation, Hôpital Mère Enfant, Centre
Hospitalier Universitaire (CHU) de Lyon–Lyon
University, Lyon, France

Keywords: 32-item Motor Function Measure; Clinician-reported outcome; Reliability; Spinal muscular atrophy; Validity

Key Summary Points

This study demonstrated that the 32-item Motor Function Measure (MFM32) is a reliable and valid measure for assessing motor function in younger individuals with neuromuscular disorders, aged 2–5 years and in non-ambulant individuals (those unable to walk) with Types 2 or 3 spinal muscular atrophy, aged 2–25 years.

The analyses provide supportive evidence for the use of the MFM32 across a wide age range.

INTRODUCTION

Spinal muscular atrophy (SMA) is a progressive neuromuscular disorder (NMD) with a broad range of severity [1, 2]. The phenotypic spectrum is historically classified into four subtypes (1–4, with 1 indicating most severe and 4 indicating least severe) based on age of onset and the maximum motor milestone achieved [1]. Due to the heterogeneity of the SMA population, outcome measures which detect change across the disease spectrum and differentiate between phenotypes are essential. While the use of such outcome measures may not be appropriate across all SMA types, for example, due to the severe impairment and age of patients with Type 1 SMA (not receiving a disease-modifying therapy), it is appropriate for those with Types 2 or 3 SMA, which reflect a more overlapping population within the continuum of motor function impairment [3].

The 32-item Motor Function Measure (MFM32) is a measurement of motor function that has been shown to be valid and reliable in individuals with NMDs aged ≥ 6 years [4]. Prior studies in SMA have demonstrated convergent and known-groups validity, and responsiveness in populations with Types 2 or 3 SMA [5, 6]. For use in clinical trials, it is important to

demonstrate that the measure is fit-for-purpose for use in the specific trial population.

Although a shorter version of the MFM (the 20-item MFM [MFM20]) has been used in children < 7 years [7], the 12 excluded items assess functions important for daily life. In clinical trials, the purpose of a functional scale is to assess changes following treatment intervention rather than to characterize cross-sectional scores. Thus, there is a practical and conceptual basis for inclusion of the 12 items in younger patients in clinical trials assessing treatment intervention, including those aged 2–6 years.

This study had two main objectives. The first was to extend the original validation by Berard et al. (2005) [4], by investigating the validity and reliability of the MFM32 in patients with NMDs, including SMA, aged 2–5 years. The second objective was to assess the validity and reliability of the MFM32 in a population representative of the risdiplam SUNFISH Part 2 study population (NCT02908685), in which the MFM32 is the primary endpoint (i.e. in individuals with Type 2 and non-ambulant Type 3 SMA, aged 2–25 years) [8].

METHODS

Analysis Population

This study was a retrospective analysis using data extracted from the MFM database, a multinational database containing MFM data collected in routine clinical practice. Two populations were studied: (1) patients with NMD (including SMA), aged 2–5 years; (2) patients with Type 2 or non-ambulant Type 3 SMA, aged 2–25 years. As the study population was not prospectively recruited, the analysis dataset reflects a convenience sample of individuals who provided their data to the MFM database.

Ambulatory status was defined by the clinician at each visit, based on questions on walking capacities acquired and loss of ambulation (defined as the ability to walk 10 m) or on MFM items 28 (score of 3), 29 (score of 3), and 30 (score of 2 or 3). The earliest visit with complete data, at which patients met the criteria for age (both populations) and ambulatory status (2- to

25-year age group only), was used for each analysis, with the exception of test–retest reliability, for which the two visits closest together (with available data) were used.

Standard Protocol Approvals, Registration, and Patient Consent

There was no protocol associated with the collection of data. This study was conducted in compliance with Good Clinical Practice guidelines, including International Conference on Harmonization guidelines [9] and consistent with the most recent version of the Declaration of Helsinki. In addition, all applicable local laws and regulatory requirements were adhered to throughout the study. All participants/a primary caregiver consented to the data being used for research activities. Ethical approval for the conduct of these analyses was granted by Comité d'Éthique du CHU de Lyon (No. 20–95).

Outcome Assessments

Three outcome assessments, namely, the MFM32, the CGI-S scale, and the Vignos functional grade, were included in the analysis.

MFM32

The MFM32 was used to assess motor function ability in both populations. The 32 items of this measure were scored using a 4-point Likert scale [4]: 0, cannot initiate the task or maintain the starting position; 1, performs the task partially; 2, performs the task incompletely or completely but imperfectly (with compensatory/uncontrolled movements or slowness); 3, performs the task fully and “normally”. The MFM32 includes three domains, with Domain 1 (D1) assessing standing transfers and ambulation; Domain 2 (D2) assessing proximal and axial function; and Domain 3 (D3) assessing distal function. The raw sum score of the 32 items (range 0–96) is then converted to a 0–100 scale, where lower scores indicate poorer functional ability.

CGI-S Scale

The CGI-S scale is a clinician-rated measure that evaluates overall disease severity through

assessment of the patient's history, psychosocial circumstances, behavior, and the impact of symptoms on the patient's ability to function. The CGI-S scale was rated by the clinician in both populations with four response options: mild, moderate, severe, and very severe. The CGI-S was not a mandatory assessment and thus was not completed for all patients or at all patient visits.

Vignos Grade

The Vignos grade is a single-item assessment of lower extremity function, rated by the clinician [10]. Lower limb function was assessed by Vignos grade in the 2- to 5-year-old study population only. Vignos grade has 10 response options: 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 (> 12 s for four standard steps); 4, walks unassisted and rises from chair but cannot climb stairs; 5, walks unassisted but cannot rise 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, is confined to bed. Vignos grade was not a mandatory assessment and thus was not completed for all patients or at all patient visits.

Analyses

Analyses were conducted on all patients with available data using SAS v9.4 statistical software (SAS Institute, Cary, NC, USA). Scoring manuals were used to determine appropriate methods for the handling of item-level missing data (e.g. MFM32 missing item data were imputed as 0). Where appropriate, the threshold for statistical significance was $P < 0.05$ (without adjusting for multiplicity), and suggested thresholds of acceptability were used to aid interpretation.

Sociodemographic Descriptive Statistics

Descriptive statistics were calculated for the patients' demographic characteristics (age

[mean, standard deviation], gender, disease [frequency and percentage by category]) in order to characterize the sample.

Reliability

Test–Retest Reliability

Test–retest reliability was conducted to assess the degree to which scores remain unchanged when measuring a stable individual characteristic on different occasions. Test–retest reliability of the MFM32 total score was assessed by comparing scores at two time points in stable patients. This is a common methodology for assessing reliability in outcome measures intended for use in clinical trials [11]. While short-term follow-up in all patients permits an assessment of test–retest reliability, it offers little in understanding how stable assessments are over longer periods of time. Moreover, if a patient's condition has changed between the two time points (regardless of length of interval period), these two time points should not be included in the analysis as we should expect that their scores differ. Where there are multiple patients fitting these criteria, a high test–retest reliability coefficient may actually reflect a lack of sensitivity (and, arguably, validity). Indeed, the assessment will no longer be one of test–retest reliability. This is generally not an issue when two time points are close together, but it cannot be ruled out. For this reason, it is important to select stable patients in a manner consistent with guidance from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative [12].

Stable patients were defined as: (1) patients in the 2- to 5-year-old population with no change in Vignos grade between two time points; and (2) patients with no change in CGI-S score between two time points (both populations). Intraclass correlation coefficient model 2,1 (ICC [2,1]), a two-way, random, single-measure analysis of variance (subject by visit) was calculated to assess the test–retest reliability. An ICC ≥ 0.7 was considered to be acceptable [13].

Internal Consistency Reliability

Internal consistency reliability was determined to assess the extent to which items within a scale or domain measure various aspects of the same characteristic or construct [14]. Internal consistency of the MFM32 was assessed by calculating Cronbach's α . A Cronbach's $\alpha \geq 0.7$ was considered to be acceptable [13, 14].

Validity

Convergent Validity

The validity of a measure can be evaluated by demonstrating its relationship (via correlation) with other measures. The more conceptually related the construct being measured, the greater the correlation should be. Convergent validity of the MFM32 total score was assessed using Vignos grade for the 2- to 5-year-old population, and Spearman rank-order correlations with CGI-S scores were used in both populations. Correlations > 0.4 were anticipated. To aid interpretation, the following thresholds were used: < 0.2 , weak; ≥ 0.2 to < 0.4 , modest; ≥ 0.4 to < 0.6 , moderate; ≥ 0.6 to 0.8 , strong; ≥ 0.8 , very strong [15, 16].

Known-Groups Validity

The validity of a measure can be demonstrated by its ability to discriminate between two groups known to differ for the variable of interest. Known-groups validity was assessed by comparing mean total MFM32 scores via analysis of covariance (controlling for age and gender) with groups defined by: Vignos grade (1–5 vs. 6–10) for the 2- to 5-year-old population and CGI-S score (mild/moderate vs. severe/very severe) for both populations. A significant difference ($P < 0.05$) between the groups was required to provide evidence of known-groups validity.

RESULTS

Patient Demographics

A total of 165 individuals with NMDs were included in the analyses (Table 1). The mean

Table 1 Patient demographic characteristics at baseline in both populations

Characteristic	2- to 25-years' population (N = 81)	2–5 years' population (N = 84)
Age, years		
Mean (SD)	11.76 (5.71)	4.87 (0.95)
Min–Max	2–25	2–5
Gender, n (%)		
Male	41 (50.62)	54 (64.29)
Female	40 (49.38)	30 (35.71)
Spinal muscular atrophy Type, n (%)		
2	63 (77.78)	See Table 2
3	18 (22.22)	
MFM32		
Mean (SD)	41.31 (15.49)	69.92 (20.55)
Min–Max	6.25–81.25	7.29–96.88

MFM32 32-item Motor Function Measure, *SD* standard deviation

age of the participants in the 2- to 5-years' population was 4.87 years, and there was a higher proportion of males (64.29%). The mean age of participants in the 2- to 25-years' population was 11.76 years, and there was a similar proportion of males and females. In the 2- to 5-years' population, a range of NMDs were present in the patient sample, with the most common being Duchenne muscular dystrophy (30.95%) and SMA (20.24%) (Table 2). In the 2- to 25-years' population, Type 2 SMA was predominant (77.78%) (Table 1). There was a broad range of MFM32 total scores in both subpopulations (2- to 5-years' population range 7.29–96.88; 2- to 25-years' population range 6.25–81.25), demonstrating a wide range of functional ability.

Test–Retest and Internal Consistency Reliability

The mean time between assessments for test–retest reliability was 348 days for the 2- to

Table 2 Summary statistics of neuromuscular disease in the 2- to 5-years' population

Disease	2- to 5-years' population (N = 84)
Spinal muscular atrophy Type	17 (20.24)
2	12 (14.29)
3	5 (5.95)
Duchenne muscular dystrophy	26 (30.95)
Congenital myopathy	14 (16.67)
Charcot-Marie-Tooth	6 (7.14)
Congenital muscular dystrophy	6 (7.14)
Myotonic dystrophy	4 (4.76)
Spinal-bulbar muscular dystrophy	4 (4.76)
Becker muscular dystrophy	2 (2.38)
Peripheral neuropathy	2 (2.38)
Limb-girdle muscular dystrophy	1 (1.19)
Myotonia congenita	1 (1.19)
Progressive muscular dystrophy	1 (1.19)

Values in table are presented as the number (of patients) with the percentage of total patient group in parenthesis

5-years' population and 305 days for the 2- to 25-years' population. An acceptable threshold of > 0.7 was reached for all reliability analyses (Table 3).

Convergent Validity

All validity analyses achieved the acceptable threshold of > 0.4 for the respective tests. The correlation between the MFM32 and the CGI-S was very strong in the 2- to 5-years' population (Spearman's $\rho = -0.84$, $P < 0.0001$, $n = 37$) compared with the modest correlation in the 2- to 25-years' population (Spearman's $\rho = -0.49$, $P < 0.001$, $n = 51$). In addition, in the former, the correlation between the MFM32 and Vignos grade was moderate to strong (Spearman's $\rho = -0.79$, $P < 0.0001$, $n = 47$). The correlations follow the expected pattern (i.e. higher MFM scores are associated with lower Vignos grade and CGI-S grades).

Table 3 Test–retest and internal consistency reliability of the total MFM32 score in both populations

Patient population	Test–retest reliability (CGI-S)		Test–retest reliability (Vignos)		Internal consistency	
	<i>N</i>	ICC (2,1)	<i>N</i>	ICC (2,1)	<i>N</i>	Cronbach's α
2–25 years	15	0.97	–	–	80	0.95
2–5 years	14	0.94	19	0.95	84	0.96

CGI-S Clinical Global Impression of Severity, *ICC* intraclass correlation coefficient model, *MFM32* 32-item Motor Function Measure

Known-Groups Validity

The results of the known-groups validity analyses in both populations are shown in Fig. 1a, b. Least square (LS) means followed the expected pattern (i.e. less severe patients had higher MFM32 total scores) for CGI-S groups mild/moderate versus CGI-S groups moderate/severe in both populations, and for Vignos grade Groups 1–5 and 6–10 in the 2- to 5-years' population. LS mean differences for CGI-S were 34.51 (95% confidence interval [CI] 25.8–43.2) for the 2- to 5-years' population and 25.56 (95% CI 11.8–39.3) for the 2- to 25-years' population. LS mean difference for Vignos grade was 39.24 (95% CI 29.9–48.6). Sample sizes in the groups were typically small. All analyses demonstrated significant differences ($P < 0.001$) between groups.

DISCUSSION

This study provides strong evidence of the validity and reliability of the MFM32 total score in younger individuals with NMDs (including 17 individuals with Type 2 or non-ambulant Type 3 SMA) aged 2–5 years. Combined with the analyses reported by Berard et al. (2005) [4], there is evidence supporting the use of this scale in individuals with NMDs, including SMA, aged 2–62 years, both in clinical practice and in clinical studies. In addition, strong evidence of these measurement properties has also been provided for a Type 2 and non-ambulant Type 3 SMA population, aged 2–25 years, supporting the use of this scale in clinical practice and clinical studies, involving children, adolescents, and young adults.

Evidence of MFM32 test–retest and internal consistency reliability was demonstrated by high ICCs associated with a subset of patients with no change in global disease severity and lower limb function (CGI-S and Vignos grade) for the 2- to 5-years' population and no change in global disease severity (CGI-S) for the 2- to 25-years' population. The high Cronbach's α results evaluated in isolation may be indicative of item redundancy; however, in this instance the tasks required for each of the 32 items are, by design, sufficiently functionally distinct and, therefore, provide additional value when evaluating changes in motor function (for a list of items, see Bérard et al. 2005 [4]). The Spearman rank correlations between MFM32 and CGI-S followed the expected pattern in both populations with strong (2- to 5-years' population) and moderately strong (2- to 25-years' population) inverse correlations identified, thereby providing evidence of convergent validity. The MFM32 was also able to discriminate between groups defined by CGI-S scores (patients with mild/moderate vs. severe/very severe global status) in both populations and between lower limb function (Vignos grade 1–5 vs. 6–10) in the 2- to 5-years' population with statistical significance, providing evidence of known-groups validity.

Although the MFM20 has been validated for use in younger individuals aged < 7 years [7], the use of two different versions of the MFM in clinical trials is challenging as changes over time cannot be equated across populations of different ages due to differences in the number of items attempted, as well as the contribution of each item to the total score (100/96 vs. 100/60 per item). Often in rare diseases (where data may be limited), measurement properties

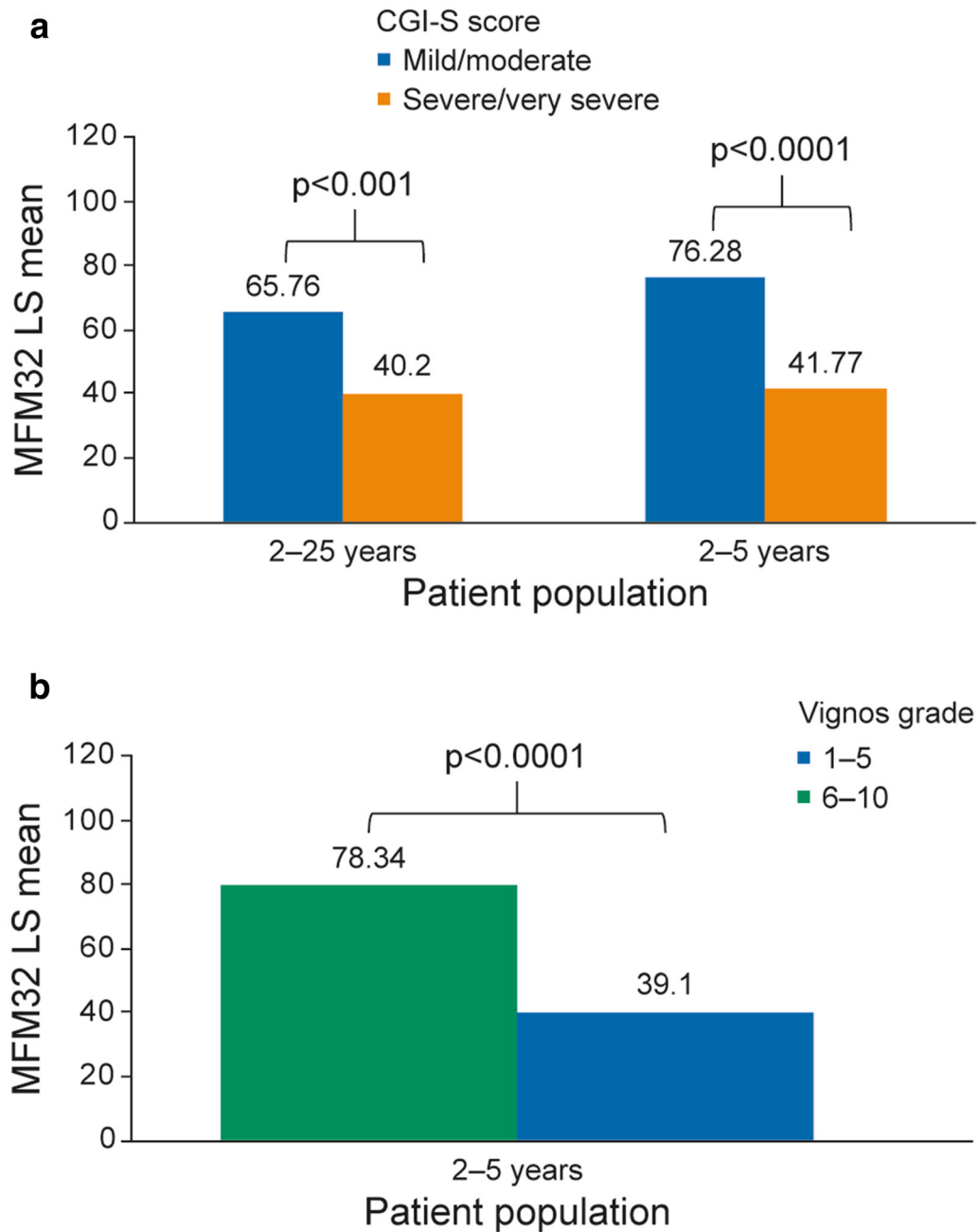


Fig. 1 a Known-groups validity in both populations defined by the CGI-S scale. **b** Known-groups validity in the 2- to 5-years' population defined by Vignos score. CGI-

S Clinical Global Impression of Severity, LS Least square, MFM32 32-item Motor Function Measure

are not assessed across narrow age ranges within a population. Indeed, the sample size for the 2- to 5-year-olds in our study with SMA was prohibitive for a dedicated analysis. Of note, a similar measure has also been used in clinical

trials in SMA, namely, the Hammersmith Functional Motor Scale – Expanded (HFMS-E), which is considered to be valid down to an age of 2 years based on an analysis across a population aged 2–45 years ($n = 70$) [17]. The HFMS-E

contains items analogous to those in the 12 items excluded by the MFM20, including raising both hands to the head. Thus, the evidence provided by our analysis of the 2- to 25-year's SMA population provides a level of validation similar to other commonly used scales. In addition, our analysis of the 2- to 5-years' NMD population provides a more targeted investigation of measurement properties in younger individuals.

In the context of a clinical trial lasting over a number of years, it is more optimal to use a single outcome assessment to evaluate changes over time. The use of both the MFM20 and MFM32 scales makes it difficult to compare the magnitude of change due to a treatment. Using a single measure of motor function, the MFM32, in clinical trials of SMA will allow comparisons of scores across a broader age range. For these reasons, the SUNFISH clinical trial, a multicenter, two-part, randomized, double-blind study assessing the safety and efficacy of the survival of motor neuron 2 splicing modifier risdiplam in patients with Type 2 or 3 SMA, aged 2–25 years, in comparison to placebo, used the MFM32 as the primary endpoint for all patients. Our study provides supportive evidence for the use of the MFM32 in patients aged ≥ 2 years as an alternative to a mixed use of the MFM20 and MFM32, addressing the issue based on the interpretation of two related but different scales.

STUDY LIMITATIONS

Due to the study design, there are several study limitations to be considered. Although the 2- to 5-years' population included patients with SMA (20%), patients with a range of NMDs were included in the analysis dataset. Investigation of the measurement properties (validity and reliability) of the MFM32 in a larger population of patients with SMA aged < 6 years would provide further supportive evidence for the use of the MFM32 total score in younger children with SMA. Furthermore, the 2- to 5-years' population contained a greater proportion of older individuals (i.e. few patients aged 2–3 years [15% of sample]). While this is a limitation, it is

one that applies to many scales frequently used within NMD populations, where measurement properties are not commonly assessed within narrow age ranges. Indeed, despite the limitations, this study provides a more targeted assessment than those typically conducted.

Several limitations relate to the retrospective nature of the study, including the availability and timing of suitable variables for use in the analysis. Data for more than one visit were not available for many patients within the sample (2- to 5-years' population: 45%; 2- to 25-years population: 75%). Additionally, the availability of CGI-S and Vignos grade data were limited, with no indication if the measures were missing at random or otherwise. The known-groups analyses assessed broad groupings (i.e. comparing walkers vs. non-walkers), and future studies would benefit from examining additional known-groups (e.g. walking assisted, walking unassisted, climbing stairs, etc.). Additionally, analyses were limited by the availability of suitable variables for assessing performance of MFM32 in relation to, for example, upper limb function. When considering areas for future research, anchor- and distribution-based analyses in a suitable sample (e.g. in a dataset that includes both a suitable anchor measure and a broad range of change in motor function) should be conducted to estimate a meaningful within-patient change threshold to support interpretation of the MFM32 data.

CONCLUSIONS

These analyses provide supportive evidence for broader use of the MFM32, with evidence of validity and reliability both in individuals aged 2–5 years with NMDs and those aged 2–25 years with Type 2 or non-ambulant Type 3 SMA. This is particularly important given treatment advances in NMDs (including ongoing and recently completed clinical trials) and the need for interpretable outcomes that cover both broad functioning and age ranges.

ACKNOWLEDGEMENTS

The authors would like to thank all individuals, and their families, who provided their data to the MFM database, as well as the site staff involved in collecting the data and the study investigators (Dalil Hamroun, Pascal Rippert, Sylvie Ragot-Mandry, Manuella Fournier Mehouas, Marguerite Munoz, H el ene Rauscent, Susana Quijano-Roy, Jean-Paul Gayraud, Vincent Tiffreau, Aleksandra Nadaj-Pakleza, Pascal Bonnet, Fran ois Rivier, Jon Andoni Urtizberea, V eronique Bourg, Anne Renders, Sybille Pellicieux, Laurent Servais, and Sylviane Peudenier). The authors would also like to thank Hannah Staunton, Roche, for her review of the draft manuscript.

Funding. This study and Rapid Service Fee were funded by F. Hoffmann-La Roche Ltd.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Author contributions. DT interpreted the data and drafted the manuscript. KG, SS, TS, and CV interpreted the data and revised the manuscript for intellectual content.

Disclosures. Dylan Trundell and Ksenija Gorni are shareholders and employees of Roche Products Ltd. Carole Vuillerot is a PI for Trophos and Roche clinical trials and has received consultancy fees from Roche, Biogen, and Avexis. Stephanie Le Scouiller is an employee of Roche Products Ltd. Timothy Seabrook is a stockholder and was an employee of Roche Products Ltd at the time of the study. He is currently affiliated with VectivBio.AG, Basel, Switzerland.

Medical Writing, Editorial, and other Assistance. Medical writing support was provided by Lindsey Weedon at MediTech Media and was funded by F. Hoffmann-La Roche Ltd.

Compliance with Ethics Guidelines. This study was conducted in compliance with Good Clinical Practice guidelines, including International Conference on Harmonization guidelines and consistent with the most recent version of the Declaration of Helsinki. In addition, all applicable local laws and regulatory requirements were adhered to throughout the study. All participants/a primary caregiver consented to their data being used for research activities. Ethical approval for the conduct of these analyses was granted by Comit e d' ethique du CHU de Lyon (No. 20-95).

Data Availability. The datasets generated during and/or analyzed during the current study are not publicly available due to the data belonging to the MFM database. For access to the MFM database, please direct enquiries to Hospices Civils de Lyon.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Mercuri E, Bertini E, Iannaccone ST. Childhood spinal muscular atrophy: controversies and challenges. *Lancet Neurol.* 2012;11(5):443–52.

2. D'Amico A, Mercuri E, Tiziano FD, Bertini E. Spinal muscular atrophy. *Orphanet J Rare Dis.* 2011;6:71.
3. Arnold WD, Kassar D, Kissel JT. Spinal muscular atrophy: diagnosis and management in a new therapeutic era. *Muscle Nerve.* 2015;51(2):157–67.
4. Berard C, Payan C, Hodgkinson I, Fermanian J, Group MFMCS. A motor function measure for neuromuscular diseases. Construction and validation study. *Neuromuscul Disord.* 2005;15(7):463–70.
5. Vuillerot C, Payan C, Iwaz J, Ecochard R, Berard C, MFM Spinal Muscular Atrophy Study Group. Responsiveness of the motor function measure in patients with spinal muscular atrophy. *Arch Phys Med Rehabil.* 2013;94(8):1555–611.
6. Chabanon A, Seferian AM, Daron A, et al. Prospective and longitudinal natural history study of patients with Type 2 and 3 spinal muscular atrophy: baseline data NatHis-SMA study. *PLoS ONE.* 2018;13(7):e0201004.
7. de Lattre C, Payan C, Vuillerot C, et al. Motor function measure: validation of a short form for young children with neuromuscular diseases. *Arch Phys Med Rehabil.* 2013;94(11):2218–26.
8. ClinicalTrials.gov. A study to investigate the safety, tolerability, pharmacokinetics, pharmacodynamics and efficacy of RO7034067 in Type 2 and 3 spinal muscular atrophy (SMA) participants (SUNFISH). 2019. <https://clinicaltrials.gov/ct2/show/NCT02908685?term=SUNFISH&rank=1>. Accessed July 2020.
9. [No authors listed]. ICH Harmonized Tripartite Guideline. Guideline for good clinical practice. *J Postgrad Med.* 2001;47(3):199–203.
10. Yi-Jing L, Rong-Fong L, Shun-Sheng C, Yen-Mou L. Measurement of the function status of patients with different types of muscular dystrophy. *Kaohsiung J Med Sci.* 2009;25:325–33.
11. Matza LS, Thompson CL, Krasnow J, Brewster-Jordan J, Zyczynski T, Coyne KS. Test-retest reliability of four questionnaires for patients with overactive bladder: the overactive bladder questionnaire (OAB-q), patient perception of bladder condition (PPBC), urgency questionnaire (UQ), and the primary OAB symptom questionnaire (POSQ). *Neurourol Urodyn.* 2005;24(3):215–25.
12. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs); user manual. 2018. <https://www.cosmin.nl>. Accessed July 2020.
13. Nunally JC, Bernstein IH. The assessment of reliability. *Psychometric Theory.* New York: McGraw-Hill; 1994. p. 248–92.
14. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
15. Bartz, AE. Basic statistical concepts (4th ed.). 1999. Upper Saddle River, NJ: Prentice Hall, Inc.
16. Swinscow DTV. Correlation and regression. *Statistics at square one.* 9th ed. London: BMJ Publishing; 1997. p. 75–84.
17. Glanzman AM, O'Hagen JM, McDermott MP, et al. Validation of the expanded hammersmith functional motor scale in spinal muscular atrophy type II and III. *J Child Neurol.* 2011;26(12):1499–507.