THE QUALITY OF LIFE IN GENETIC NEUROMUSCULAR DISEASE QUESTIONNAIRE: RASCH VALIDATION OF THE FRENCH VERSION

ANTOINE DANY, PT, PhD,[1] AMANDINE RAPIN, MD, MSc,[1] BRICE LAVRARD, MD,[1] VIRGINIE SAOÛT, MD,[2] CHRISTIAN RÉVEILLÈRE, PhD,[3] GUILLAUME BASSEZ, MD, PhD,[3] VINCENT TIFFREAU, MD, PhD,[4] YANN PÉRÉON, MD, PhD,[5] SABRINA SACCONI, MD, PhD,[6] BRUNO EYMARD, MD, PhD,[7] MOUSTAPHA DRAMÉ, MD, PhD,[1] DAMIEN JOLLY, MD, PhD,[1] JEAN-LUC NOVELLA, MD, PhD,[1] JEAN-BENOIT HARDOUIN, PhD,[8] and FRANÇOIS C. BOYER, MD, PhD[1]

[1] Centre Hospitalier Universitaire de Reims, Hôpital Sébastopol, Service de Médecine Physique et Réadaptation, Centre de Référence des Maladies Neuromusculaires, EA 3797, 48, rue de Sébastopol, 51092, Reims Cedex, France
[2] Centre des Capucins, Angers, France
[3] Hôpital Henri Mondor, Créteil, France
[4] Hôpital Pierre Swynghedauw, Lille, France
[5] Hôpital Hôtel-Dieu, Nantes, France
[6] Centre Hospitalier Universitaire de Nice, Nice, France
[7] Institut de Myologie, Paris, France
[8] Université de Nantes, Nantes, France

Accepted 28 January 2017

ABSTRACT: *Introduction:* Slowly progressive, genetic neuromuscular diseases (gNMDs) often lead to important motor deficiencies and functional limitations. The Quality of Life in Genetic Neuromuscular Disease Questionnaire (QoL-gNMD) is a new health-related quality-of-life questionnaire developed for these patients. The purpose of the present study was to validate the French version of the QoL-gNMD and to calibrate its measurement system. *Methods:* Both the QoL-gNMD and a validated generic questionnaire (WHOQOL-BREF) were administered to patients. Validation was performed using item response theory. The partial credit model (Rasch) was used to calibrate each domain. *Results:* Three hundred fifteen adult patients were included. All 3 domains showed adequate psychometric properties (internal consistency: person separation index >0.77; repeatability: test–retest intraclass correlation coefficient >0.75, scalability coefficient >0.38) and fitted the partial credit model. The QoL-gNMD also demonstrated adequate concurrent validity with the WHOQOL-BREF. *Discussion:* The QoL-gNMD showed adequate psychometric properties and can be used in clinical settings. Although not anchor-based, the minimum detectable change tables help in interpreting score change.

Muscle Nerve **56**: 1085–1091, 2017

Slowly progressive neuromuscular diseases (NMDs), such as muscular dystrophies, myopathies or spinal muscular atrophies, involve a progressive loss of motor functions, resulting in a variety of outcomes ranging from mild muscle weakness to highly debilitating conditions. The age of onset varies greatly from the prenatal period to late adulthood. Therapeutic options are often limited and the long-term preservation of health-related quality of life (HRQL) is often the main focus of medical care. It is therefore particularly important to have HRQL assessment in these patients.[1]

HRQL measures are valuable tools to assess subjective perception of the disease, its progression, and the benefit yielded from rehabilitation or pharmacological treatments. Idler *et al.* showed that, when compared with classic objective clinical measures, these subjective measures could even be better predictors of mortality and, to a lesser extent, of functional limitations.[2] Public organizations (e.g., the European Medicines Agency and the Food and Drugs Administration) encourage or may even require developers of new health interventions to include HRQL assessments in their randomized clinical trials.[3,4]

So far, HRQL in NMD has been assessed using either generic or specific questionnaires.[5,6] Generic questionnaires lack the exploration of specific aspects of life often impaired by NMD, while they may include aspects of life that are irrelevant to patients with NMD. There are currently 3 adult NMD-specific HRQL questionnaires: the Quality of Life Profile Questionnaire, which has never undergone a quantitative validation study[7,8]; the Individualized Neuromuscular Quality of Life Questionnaire[9]; and the Neuromuscular Disease Impact Profile.[10] The latter 2 were developed using only classical test theory. These questionnaires do not therefore meet the current best methodological standards.[11,12] Moreover, they were developed using diseases that are either acquired or encompass major sensory alterations, which lead to very different disease evolution patterns. In the case of myotonic dystrophy type 1, there is a specific patient-reported health questionnaire, the

Abbreviations: gNMD, genetic neuromuscular diseases; H, scalability coefficient; HRQL, health-related quality of life; IRT, item response theory; ICC, intraclass correlation coefficient; NMD, neuromuscular diseases; PSI, person separation index; QoL-gNMD, Quality of Life in genetic Neuromuscular Disease questionnaire; WHOQOL, World Health Organization Quality of Life questionnaire

Key words: item response theory; neuromuscular diseases; patient-centered outcomes; quality of life; Rasch model

Additional Supporting Information may be found in the online version of this article at the publisher's website

This project was supported by the French Muscular Dystrophy Association (Association française contre les myopathies/AFM-téléthon) and the Champagne-Ardenne Region (Programme ESSAIMAGE; principal investigator: François Constant Boyer).

Correspondence to: A. Dany; e-mail: adany@chu-reims.fr

Published online 6 February 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/mus.25598

Myotonic Dystrophy Health Index,[13] but it has not been validated for other NMDs.

Only questionnaires developed using advanced psychometric methods, derived from item response theory (IRT), including Rasch models, generate measures on an interval scale (i.e., a unit of measurement can be associated). Only this interval scale property[14] demonstrates that, for a given scale, when there are more than 2 measures, it is legitimate to compare the scale intervals between the measures. This is crucial, because, in clinical settings, a scale is commonly used by interpreting differences across many patients or variations across several timepoints. Another asset of IRT is that missing answers do not prevent the estimation of latent traits. This is why such a method is recommended in the current COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN).[12]

On the basis of these findings, the French Muscular Dystrophy Association (AFM-Téléthon) initiated the construction of a new HRQL questionnaire for adult patients with genetic NMDs (gNMDs), designed to meet the IRT assumptions. This questionnaire was initially named the QoL-NMD and later renamed the QoL-gNMD (to emphasize that it applies only to genetic diseases). In an initial qualitative study,[15] patients with gNMDs were questioned and audio-recorded using the focus group method.[16] The resulting verbatim reports were presented using the framework of the International Classification of Functioning, Disability and Health model,[17] These results enabled a group of experts (patients and specialists) to first construct a pool of items. The pool of items was then reduced using the Delphi method,[18] and finally by a pilot study enabling the removal of grossly malfunctioning items (e.g., floor effect). This yielded a pool of 64 items, small enough to be conveniently administered to patients.

This qualitative work was pursued with an exploratory quantitative study, enabling construction of the QoL-gNMD.[19] In this exploratory sample, each domain had good psychometric properties and met the IRT assumptions. This scale was translated into English with transcultural adaptation by 2 experts in questionnaire development, including a native English linguist. Both the French and English versions of the QoL-gNMD are freely available from the corresponding author. The QoL-gNMD psychometric properties remained to be validated in both languages.

The objective of the present study was 2-fold: to validate the QoL-gNMD on a confirmatory sample of French patients and to calibrate its measurement system.

METHODS

Participants. The QoL-gNMD and the short version of the generic World Health Organization Quality of Life questionnaire (WHOQOL-BREF)[20] were administered to patients recruited in 9 NMD referral centers (Angers, Créteil, Garches, Lille, Nancy, Nantes, Nice, Paris Institute of Myology, and Reims). Patients completed the questionnaires between January 2011 and January 2015. The Reims institutional review board approved the ethical aspects of the study. All patients gave written consent. Data collection was authorized by the French personal-data protection agency.

The QoL-gNMD is structured in 3 domains (Impact of Physical Symptoms, Self-Perception, and Activities and Social Participation) and includes 26 items with 2–4 response options. The WHOQOL-BREF is structured in 4 domains (Physical Health, Psychological, Social Relationships, and Environment) and includes 26 items with 5 response options. Both questionnaires include 2 general items that do not play a role in the domain measure calculation. These items are meant to familiarize patients with the instrument before they begin the real assessment. They may also be used as a rough but fast global assessment of HRQL change that needs to be specified further by domain measures.

Eligible patients had a genetic neuromuscular disease, confirmed by the appropriate diagnostic method (laboratory results or indisputable clinical findings and paraclinical tests). The disease was a pure or predominant motor deficiency and was not associated with any symmetrical sensory deficiency or autoimmune disease. Patients were at least 18 years old, and were excluded if they could not read or speak fluently. No minimum school-grade level was required.

The ability of patients to perform activities of daily living was evaluated using the simplified Barthel index. This scale assesses feeding, bathing, grooming, dressing, bowel control, bladder control, toileting, chair transfer, ambulation, and stair climbing. Its measures range from 0 (very dependent) to 20 (independent).[21] Additional data, including age when symptoms first appeared, walking status, and potential use of mechanical ventilation device, were also collected.

Statistical Analyses. First, we validated the psychometric properties on a confirmatory sample of patients. Second, each domain was calibrated using an IRT model on all patients.

Reliability. IRT assumptions (unidimensionality, local independence, and monotonicity) were assessed using a non-parametric IRT model[22,23] (criterion scalability coefficient $H > 0.3$). Unidimensionality was reassessed using a parallel analysis on a principal component analysis.

Internal consistency was assessed using the person separation index (PSI),[24] which is comparable to Cronbach's α (criterion PSI >0.7). Test–retest reliability was assessed by the intraclass correlation coefficient (criterion ICC > 0.6) between 2 time-spaced administrations of the questionnaire. The time lapse was 1 month $\pm$ 7 days.

Validity. Associations between the QoL-gNMD and the WHOQOL-BREF were assessed using the Spearman correlation. The WHOQOL-BREF was not considered as a "gold standard," but rather as a point of comparison. Strong to moderate correlations (>0.4) were expected between each QoL-gNMD domain and the Physical Health domain in the WHOQOL-BREF as well as between the Self-Perception

domain in the QoL-gNMD and the Psychological domain in the WHOQOL-BREF. Weak correlations (<0.4) were expected between all other domains when comparing the QoL-gNMD and the WHOQOL-BREF.

IRT Calibration. In each domain, the measurement system is designed so that a higher measure represents a better quality of life. The rating metric of the QoL-gNMD was calibrated using the partial credit model (PCM).[25] To assess model fit, we computed a chi-square–based fit test (α risk = 5%).

To have a metric more appealing for the future users, all measures were rescaled to what is termed a T metric, with the mean set at 50 and standard deviation at 10. We estimated minimum detectable changes to determine which measure changes are above measurement errors and could thus reflect a change of patient's status over time.[26]

We investigated the presence of differential item functioning across genders, age groups, and walking status. Differential item functioning across generations was assessed by comparing 2 age groups. The cut-off age was the closest 5-by-5 number around the median age. To be considered as relevant, differential item functioning levels are to be higher than the minimal clinically important difference.[27] In the absence of a clinical effect size, salience was defined as higher than the median standard error of measurement for each domain, as described by Choi *et al.*[28]

Statistical analyses were performed using R programming language (R Foundation for Statistical Computing, Vienna, Austria).[28–31]

RESULTS

Participants. A total of 315 patients were recruited. Patients' characteristics are presented in Table 1. The majority of the patients were men. Diagnoses included various gNMDs, among which the most frequent were myotonic dystrophy type 1, facioscapulohumeral and limb-girdle muscular dystrophies, spinal muscular atrophies, and dystrophinopathies. There were small numbers of patients with congenital myopathies (4 nemaline, 3 centronuclear, 2 fiber-type disproportion, 1 multiminicore, 1 central core, and 4 unspecified), congenital muscular dystrophies (7 Ullrich, 1 Fukuyama-type, and 2 unspecified), muscular dystrophies (4 Emery–Dreifuss, 2 oculopharyngeal, and 2 unspecified laminopathies), glycogen-storage diseases (6 Pompe disease and 1 Cori–Forbes disease), and 9 other gNMDs (3 unspecified VCP mutations, 3 mitochondrial myopathies, 1 myofibrillar myopathy, 1 congenital myasthenic syndrome, and 1 unspecified myopathy). The majority of the patients were ambulatory. According to the simplified Barthel index, almost one third of the patients were independent for basic activities of daily living, whereas around one quarter had to rely substantially on a third party, with an index <10. One fifth of the patients used non-invasive mechanical ventilation and 4% were dependent on invasive mechanical ventilation. Almost all patients reached at least high school (95%), but only 42% had been to a college or a university. Approximately half the patients were single (51%), 36% were married, and 14% were living with a partner but not married.

Table 1. Clinical evaluation of patients.

Characteristics	Value
Number of patients	315
Male [*n* (%)]	182 (57.8)
Age (years)	
Median	42.8
25th–75th percentile	32–53
Range	18-80
Diagnosis [*n* (%)]	
Myotonic dystrophy type 1	88 (27.9)
Facioscapulohumeral muscular dystrophy	54 (17.1)
Limb-girdle muscular dystrophies	41 (13.0)
Spinal muscular atrophies	35 (11.1)
Becker muscular dystrophy	32 (10,2)
Duchenne muscular dystrophy	16 (5,1)
Congenital myopathies	15 (4.8)
Congenital muscular dystrophies	10 (3.2)
Muscular dystrophies	8 (2.5)
Glycogen storage diseases	7 (2.2)
Other neuromuscular diseases	9 (2.9)
Simplified Barthel index [*n* (%)]	
0–4	21 (6.7)
5–9	56 (17.8)
10–14	47 (14.9)
15–19	94 (29.8)
20	93 (29.5)
Omitted	4 (1.3)
Reported age when symptoms first appeared (years)	
Median	16
25th–75th percentile	6–30
Range	0–69
Walking status [*n* (%)]	
Came to consultation walking	185 (58.7)
Did not come to consultation walking	128 (40.6)
Omitted	2 (0.6)
Mechanical ventilation [*n* (%)]	
None	238 (75.6)
Non-invasive	64 (20.3)
Invasive	13 (4.1)

The validation of psychometric properties was performed on a subsample of 156 patients. The PCM parameters were estimated on 315 patients.

Impact of Physical Symptoms. Table 2 shows that 5 of 7 items in this domain had sufficient scalability coefficients. The items assessing micturition disorders and sleep quality had insufficient scalability coefficients. Despite these items, the domain had acceptable scalability and internal consistency. The parallel analysis led to the retaining of a single factor (see Fig. S1 in Supplementary Material, available online). There was excellent test–retest reliability (ICC $>$ 0.75).

There was no significant misfit on the PCM as detected by the chi-square–based fit test. Difficulty thresholds (Fig. 1, and Table S1 in Supplementary Material online) associated with items assessing the most basic body functions (micturition disorders, concentration deficit, and memory impairments)

Table 2. Psychometric properties

Domain	Item	Item scalability coefficient	Domain psychometric properties
1. Impact of Physical Symptoms	Muscle fatigue on waking	0.442 ± 0.049	Scale scalability coefficient: 0.38 ± 0.04
	Muscle fatigue caused by activities	0.414 ± 0.050	
	Sleep quality	0.261 ± 0.070	
	Pain during activities	0.446 ± 0.049	PSI = 0.77 (*N* = 154)
	Micturition disorders	0.228 ± 0.080	
	Concentration deficit	0.405 ± 0.055	Test–retest reliability: ICC = 0.75 (*N* = 107)
	Memory impairments	0.385 ± 0.047	
2. Self-Perception	Anxiety	0.449 ± 0.050	Scale scalability coefficient: 0.40 ± 0.04
	Morale	0.507 ± 0.044	
	Irritability	0.331 ± 0.059	
	Prospect of increasing involvement of persons close	0.380 ± 0.060	PSI = 0.81 (*N* = 146)
	Plans for the future	0.439 ± 0.054	
	Life control	0.361 ± 0.045	
	Perception of the way other people see them	0.338 ± 0.050	Test–retest reliability ICC = 0.79 (*N* = 99)
	Love life	0.376 ± 0.054	
3. Activities and Social Participation	Anxiety about going out alone	0.288 ± 0.061	Scale scalability coefficient: 0.41 ± 0.04
	Difficulty to move at home	0.444 ± 0.054	
	Difficulty getting around in other people's homes	0.485 ± 0.045	
	Anxiety to move alone	0.461 ± 0.047	PSI = 0.85 (*N* = 151)
	Risk of falling	0.392 ± 0.049	
	Lack of access to toilets	0.446 ± 0.051	
	Restriction in social life outside the home	0.403 ± 0.050	Test–retest reliability ICC = 0.80 (*N* = 105)
	Sensitivity to cold	0.381 ± 0.052	
	Participation in family life	0.444 ± 0.052	

Psychometric properties were assessed on a sample of 156 patients. PSI, person separation index. ICC, intraclass correlation; N, number of patients for the analysis (if less than 156).

were all below the mean latent trait value, indicating that these items were informative only among patients with major physical symptoms. The item assessing sleep quality and items assessing more active body functions (muscle fatigue on waking, muscle fatigue caused by activities, and pain during activities) were informative for all types of patients.

With 63% of the area under the domain characteristic curve below the mean latent trait value, the domain was more informative for patients with marked physical symptoms. We calculated that the minimum detectable change for the domain was between 11.57 and 16.76 on the T metric (see Table S2 in Supplementary Material online).

Self-Perception. Table 2 shows that all 8 items had sufficient scalability coefficients. The domain had good scalability and internal consistency ($H > 0.40$, $PSI > 0.80$). The parallel analysis led to the retaining of a single factor (see Fig. S1 in Supplementary Material online). There was excellent test–retest reliability.

There was no significant misfit on the PCM as detected by the chi-squared–based fit test. Difficulty thresholds (Fig. 1, and Table S1 in Supplementary Material online) associated with the item assessing love life were all below the mean latent trait value, indicating that this item was informative only among patients with low self-perception. Two difficulty thresholds for this item were close and disordered, indicating that keeping the 2 responses "very often" and "often" may be pointless. All the other items were informative for all types of patients.

With 53% of the area under the domain characteristic curve below the mean latent trait value, the Self-Perception domain was comparably informative for all patients. We calculated that the minimum detectable change for the domain was between 10.60 and 15.56 on the T metric (see Table S3 in Supplementary Material online).

Activities and Social Participation. Table 2 shows that 8 of 9 items had sufficient scalability coefficients. The item assessing anxiety about going out alone had an insufficient scalability coefficient. Despite this, the domain had good scalability and internal consistency. The parallel analysis led to the retaining of a single factor (see Fig. S1 in Supplementary Material online). There was excellent test–retest reliability.

There was no significant misfit on the PCM detected by the chi-square–based fit test. Difficulty thresholds (Fig. 1, and Table S1 in Supplementary

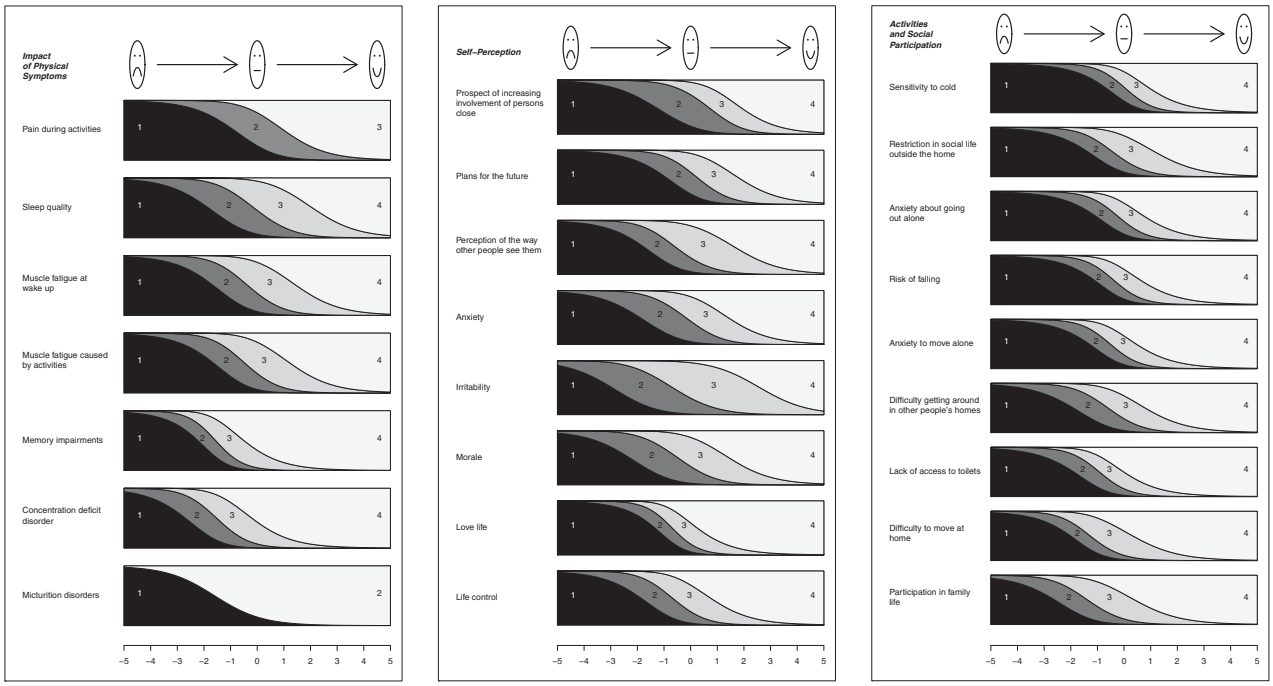

FIGURE 1. Probability of endorsement of each response. The latent trait is scaled on a theta metric with a mean set at 0 and standard deviation at 1.

Material online) associated with the item assessing the lack of access to toilets were all below the mean latent trait value, indicating that this item was informative only among patients with major restrictions in activities and social participation. All the other items were informative for all types of patients.

With 63% of the area under the domain characteristic curve below the mean latent trait value, the domain was more informative for patients with a major reduction in activities and social participation. We calculated that the minimum detectable change for the domain was between 9.20 and 15.69 on the *T* metric (see Table S4 in Supplementary Material online).

Differential Item Functioning. We did not find any items with salient differential item functioning when comparing age groups (18–44 and 45–80 years), gender, and walking status.

Converting Sum Scores into IRT-Calibrated Measures. A conversion table (Table 3) enables easy translation of sum scores into IRT-calibrated measures. The IRT-calibrated measures can be interpreted as *T* scores with a mean set at 50 and a standard deviation set at 10. The conversion table applies only when all items in the domain have been answered.

Comparison with a Generic Questionnaire. All 3 domains of the QoL-gNMD were fairly well rank-correlated with the Physical Health domain on the WHOQOL-BREF (Impact of Physical Symptoms: 0.72; Self-Perception: 0.64; Activities and Social Participation: 0.56). The rank correlation between the Self-Perception domain on the QoL-gNMD and the Psychological domain on the WHOQOL-BREF was fair (0.52). There was a low rank correlation between the Activities and Social Participation domain on the QoL-gNMD and the Social Relationship domain on the WHOQOL-BREF (0.32).

DISCUSSION

The QoL-gNMD is a new adult gNMD-specific HRQL questionnaire that was validated using an IRT model of the Rasch family. From a clinician's perspective, the QoL-gNMD possesses a number of valuable qualities. First, with only 26 items, it is short and meets the "reasonable time" requirement of being able to be administered during a medical visit or a rehabilitation session. Second, it possesses only 3 domains, providing a profile of 3 measures, which can be easily used in clinical practice to summarize a patient HRQL condition. Third, it meets IRT assumptions and thus generates measures on interval scales. Last, it is specific to a relatively homogeneous subfamily of diseases that all progress slowly.

From a researcher's perspective, it is possible to compute IRT values from a large data set, using a dedicated software application that we developed using Shiny (RStudio, Inc., Boston, Massachusetts). This application automatically computes domain latent trait values from an appropriately formatted data file (available on request). In the presence of missing values, this application should be

Table 3. Conversion table

Sum	Impact of Physical Symptoms	95% CI	Self-Perception	95% CI	Activities and Social Participation	95% CI
0	24.01	13.63–34.38	23.50	12.70–34.31	24.21	13.85–34.56
1	26.89	17.33–36.44	26.62	16.63–36.61	26.88	17.40–36.35
2	29.38	20.39–38.37	29.31	19.97–38.66	29.14	20.34–37.95
3	31.63	23.02–40.23	31.70	22.86–40.53	31.13	22.85–39.41
4	33.72	25.36–42.08	33.84	25.42–42.27	32.90	25.03–40.77
5	35.71	27.49–43.94	35.82	27.71–43.93	34.51	26.97–42.05
6	37.67	29.50–45.84	37.66	29.79–45.52	36.00	28.73–43.27
7	39.61	31.43–47.79	39.40	31.71–47.09	37.40	30.34–44.46
8	41.58	33.32–49.84	41.08	33.51–48.66	38.72	31.84–45.60
9	43.59	35.20–51.98	42.72	35.21–50.23	39.98	33.24–46.73
10	45.68	37.11–54.25	44.35	36.85–51.84	41.20	34.56–47.85
11	47.87	39.08–56.67	45.98	38.45–53.50	42.39	35.82–48.96
12	50.20	41.12–59.28	47.62	40.04–55.21	43.56	37.03–50.08
13	52.68	43.27–62.10	49.31	41.62–56.99	44.71	38.21–51.22
14	55.37	45.57–65.17	51.04	43.24–58.84	45.86	39.35–52.38
15	58.29	48.04–68.53	52.83	44.88–60.77	47.03	40.47–53.58
16	61.50	50.73–72.27	54.69	46.58–62.79	48.21	41.58–54.83
17	65.08	53.66–76.50	56.62	48.34–64.90	49.42	42.69–56.14
18	69.15	56.88–81.42	58.65	50.17–67.13	50.67	43.81–57.53
19	—	—	60.78	52.07–69.50	51.99	44.94–59.03
20	—	—	63.05	54.05–72.05	53.38	46.11–60.65
21	—	—	65.48	56.12–74.83	54.87	47.31–62.44
22	—	—	68.13	58.31–77.94	56.50	48.58–64.43
23	—	—	71.07	60.66–81.49	58.31	49.94–66.68
24	—	—	74.44	63.24–85.64	60.34	51.41–69.28
25	—	—	—	—	62.68	53.05–72.32
26	—	—	—	—	65.44	54.92–75.97
27	—	—	—	—	68.78	57.12–80.44

The estimation of the partial credit model parameters was performed on a sample of 315 patients. For each domain, a higher value reflects a better quality of life. CI, confidence interval.

systematically used in place of the conversion table provided.

Our results show that the QoL-gNMD has overall good psychometric properties. The low scalability of the item assessing micturition disorders was expected, because it had already shown barely sufficient scalability in the exploratory analysis. It had been retained because it is considered as clinically important by experts. The low scalabilities of the items assessing sleep quality and anxiety about going out alone were unexpected but did not much alter the overall psychometric properties of their respective domains. A PCM fitted the data convincingly and enabled each domain to be calibrated. The minimum detectable change tables provide a clinically useful means of interpreting change for individual patients. These tables can help clinicians and investigators identify differences for individual patients that are large enough to reflect what a patient may recognize as a status change and motivate a modification of care.[32]

Each domain of the QoL-gNMD was substantially correlated with the Physical Health domain of the WHOQOL-BREF. There was a moderate correlation between the Self-Perception domain on the QoL-gNMD and Psychological domain on the WHOQOL-BREF, and a weak correlation between the domains of Activities and Social Participation on the QoL-gNMD and Social Relationship on the WHOQOL-BREF. Thus, despite the additional patients, comparisons between the QoL-gNMD and the WHOQOL-BREF led to very similar results to those obtained in our previous study.[19] This confirms that overall psychological well-being and social life are not necessarily determined by the disease severity.

An IRT model with item-specific discrimination parameters[33,34] could have been used to improve the model fit. Discrimination parameters enable the strength of the relation between the items and the latent trait being measured to vary across items. This possibility was disregarded, however, because, from a practical viewpoint, using a "non-Rasch" IRT model implies having multiple latent trait values associated with each domain measure. This is a serious drawback for use in clinical settings, because the estimation of the latent trait values cannot be performed easily by clinicians, and requires onsite available software. Moreover, because we had a fairly small sample (<500), estimating a complex model is not recommended.[35,36]

The present study had several limitations. First, distribution-based indices, such as minimum detectable changes, are not anchored to externally meaningful variables. Thus, future efforts to develop anchor-based indices that can reliably identify for example clinically important changes would be valuable but would require years of follow-up. Second, although only the most severe gNMDs lead to respiratory failure, this medical condition has such a detrimental effect on patients' HRQL that we started the development of an optional domain on mechanical ventilation. The development and validation of this domain will take time, however, because there are few eligible patients and participation is challenging in view of their medical conditions. Third, exploring differential item functioning between diagnoses would be very relevant, but the low prevalence of most gNMDs prevents such analysis. Fourth, a final limitation is the language specificity. This version is only applicable in French-speaking countries. Because cultural nuances in the way French is spoken across the numerous French-speaking countries of the world are likely to be only minor, only small adjustments would be necessary for transcultural adaptation, making this version applicable to many patients. An English version of the QoL-gNMD is already available but still requires cross-validation.

In conclusion, this has study led to the validation and IRT calibration of the French version of the QoL-gNMD. It can be conveniently administered to patients with gNMD in French-language rehabilitation services. A conversion table enables easy transformation of sum scores into IRT-calibrated measures. Minimum detectable change tables help interpreting score change. Anchor-based indices of the minimal important difference remain to be investigated.

These findings were presented on September 2015 at Société Française de Médecine physique et de Réadaptation, Montpellier, France.

REFERENCES

1. Carter GT, Han JJ, Abresch RT, Jensen MP. The importance of assessing quality of life in patients with neuromuscular disorders. Am J Hosp Palliat Care 2007;23:493.
2. Idler E, Russell L, Davis D. Survival, functional limitations, and self-rated health in the NHANES I Epidemiologic Follow-up Study, 1992. First National Health and Nutrition Examination Survey. Am J Epidemiol 2000;152:874–883.
3. Bottomley A, Jones D, Claassens L. Patient-reported outcomes: assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. Eur J Cancer 2009;45:347–353.
4. Speight J, Barendse SM. FDA guidance on patient reported outcomes A prompt for the industry to raise scientific standards. BMJ 2010;340: c2921.
5. Padua L, Aprile I, Frusciante R, Iannaccone E, Rossi M, Renna R, *et al.* Quality of life and pain in patients with facioscapulohumeral muscular dystrophy. Muscle Nerve 2009;40:200–205.
6. Kruitwagen-Van Reenen ET, Wadman RI, Visser-Meily JM, van den Berg LH, Schröder C, van der Pol WL. Correlates of health related quality of life in adult patients with spinal muscular atrophy. Muscle Nerve 2016;54:850–855.
7. Ahlström G, Gunnarsson LG. Disability and quality of life in individuals with muscular dystrophy. Scand J Caring Sci 1996;28: 147–157.
8. Nätterlund B, Ahlström G. Activities of daily living and quality of life in persons with muscular dystrophy. J Rehabil Med 2001;33: 206–211.
9. 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;68:1051–1057.
10. Bos I, Kuks J, Wynia K. Development and testing psychometric properties of an ICF-based health measure: the Neuromuscular Disease Impact Profile. J Rehabil Med 2015;47:445–453.
11. Bann CM, Abresch RT, Biesecker B, Conway KC, Heatwole C, Peay H, *et al.* Measuring quality of life in muscular dystrophy. Neurology 2015;84:1034–1042.
12. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, *et al.* Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Med Res Methodol 2010;10:82.
13. Heatwole C, Bode R, Johnson N, Dekdebrun J, Dilek N, Heatwole M, *et al.* Myotonic Dystrophy Health Index: initial evaluation of a disease-specific outcome measure. Muscle Nerve 2014;49:906–914.
14. Stevens SS. On the theory of scale of measurement. Science 1946; 103:677–680.
15. Dany A, Rapin A, Réveillère C, Calmus A, Tiffreau V, Morrone I, *et al.* Exploring quality of life in people with slowly-progressive neuromuscular disease. Disabil Rehabil 2016;1–11.
16. Kitzinger J. Qualitative research: introducing focus groups. BMJ 1995;311:299–302.
17. World Health Organization (WHO). The international classification of functioning, disability and health. Geneva: WHO; 2001.
18. Rayens MK, Hahn EJ. Building consensus using the policy Delphi method. Policy Polit Nurs Pract 2000;1:308–315.
19. Dany A, Barbe C, Rapin A, Réveillère C, Hardouin JB, Morrone I, *et al.* Construction of a Quality of Life Questionnaire for slowly progressive neuromuscular disease. Qual Life Res 2015;24:2615–2623.
20. Skevington SM, Lotfy M, O'Connell AK. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004;13:299–310.
21. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL index: a reliability study. Int Disabil Stud 1988;10:61–63.
22. Sijtsma K, Molenaar IW. Introduction to nonparametric item response theory. Thousand Oaks, CA: Sage; 2002.
23. Mokken RJ. A theory and procedure of scale analysis. Berlin, Germany: De Gruyter; 1971.
24. Wright BD, Stone MH. Measurement essentials, 2nd ed. Wilmington, DE: Wide Range, Inc.; 1999. http://www.rasch.org/measess/me-all.pdf. Accessed July 20, 2016.
25. Masters GN. A Rasch model for partial credit scoring. Psychometrika 1982;47:149–174.
26. 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 minimally important change. Health Qual Life Outcomes 2006;4:54.
27. Crane PK, Gibbons LE, Narasimhalu K, Lai JS, Cella D. Rapid detection of differential item functioning in assessment of health-related quality of life: the functional assessment of cancer therapy. Qual Life Res 2007;16:101–114.
28. Choi SW, Gibbons LE, Crane PK. lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. J Stat Softw [online] 2011;39:1–30. http://www.jstatsoft.org/v39/i08/. Accessed July 20, 2016.
29. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Core Team. 2013. http://www.R-project.org/. Accessed July 20, 2016.
30. van der Ark LA. Mokken Scale Analysis in R. J Stat Softw [online] 2007; 20:1–19. http://www.jstatsoft.org/v20/i11/. Accessed July 20, 2016.
31. Rizopoulos D. ltm: An R package for latent variable modelling and item response theory analyses. J Stat Softw [online] 2006;17:1–25. http://www.jstatsoft.org/v17/i05/. Accessed July 20, 2016.
32. Kozlowski AJ, Cella D, Nitsch KP, Heinemann AW. Evaluating individual change with the Quality of Life in Neurological Disorders (Neuro-QoL) Short Forms. Arch Phys Med Rehabil 2016;97:650–654.e8.
33. Samejima F. Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph Supplement 1969;34.
34. Muraki E. A generalized partial credit model: application of an EM algorithm. Appl Psychol Meas 1992;16:159–176.
35. Reise SP, Yu J. parameter recovery in the graded response model using MULTILOG. J Educ Meas 1990;27:133–144.
36. Lai JS, Nowinski C, Victorson D, Bode R, Podrabsky T, McKinney N, *et al.* Quality-of-life measures in children with neurological conditions: pediatric Neuro-QOL. Neurorehabil Neural Repair 2012; 26:36–47.