Available online at www.sciencedirect.com

ScienceDirect

Neuromuscular Disorders ■■ (2015) ■■–■■

www.elsevier.com/locate/nmd

How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?

Charles Sèbiyo Batcho [a,b,c], Peter Y.K. Van den Bergh [a,d,e], Philip Van Damme [e,f,g], Anna J. Roy [h], Jean-Louis Thonnard [a,i], Massimo Penta [a,b,*] on behalf of the BNMDR Scientific Committee [e]

[a] *Institute of Neuroscience, Université Catholique de Louvain, Brussels, Belgium*
[b] *Arsalis SPRL, Glabais, Belgium*
[c] *Centre Interdisciplinaire de Recherche en Réadaptation et Intégration Sociale (CIRRIS), Université Laval, Quebec, Canada*
[d] *Neuromuscular Reference Centre, Cliniques Universitaires St-Luc, Université Catholique de Louvain, Brussels, Belgium*
[e] *Scientific Committee of the Belgian Neuromuscular Disease Registry, Belgium*
[f] *KU Leuven – University of Leuven, Department of Neurosciences, VIB – Vesalius Research Center, Experimental Neurology – Laboratory of Neurobiology, Leuven, Belgium*
[g] *University Hospitals Leuven, Department of Neurology, Leuven, Belgium*
[h] *Scientific Institute of Public Health, Brussels, Belgium*
[i] *Cliniques Universitaires Saint-Luc, Physical Medicine and Rehabilitation Department, Université Catholique de Louvain, Brussels, Belgium*

Received 30 July 2015; received in revised form 13 November 2015; accepted 10 December 2015

Abstract

This study aims to investigate the clinimetric properties of ACTIVLIM, a measure of activity limitations, when it is used in daily practice in a large nationwide representative cohort of patients with neuromuscular diseases. A cohort of 2986 patients was assessed at least once over 2 years in 6 national neuromuscular diseases reference centers. Successive Rasch analyses were conducted in order to investigate the scale validity, reliability, consistency across demographic and clinical sub-groups and its sensitivity to change. ACTIVLIM confirmed excellent fit to a unidimensional scale, with stable but 3-times more accurate item calibrations compared to the original publication. It showed a good reliability (R = 0.95), an appropriate targeting for 87% of the sample and an excellent invariance across age, gender, language and time. Despite some variations in the item difficulty hierarchy across diagnoses, ACTIVLIM exhibited a good capability to quantify small but significant changes in activity for various diagnostic groups. Overall, ACTIVLIM demonstrated very good clinimetric properties, allowing accurate quantitative measurement of activity limitations in both children and adults with a variety of neuromuscular diseases.

Keywords: Neuromuscular diseases; Activity limitations; Rasch analysis; Patient-reported outcome measure

1. Introduction

Questionnaire-based instruments have demonstrated their usefulness in various clinical assessments. They are more frequently incorporated in clinical practice in many fields and in health care policy planning [1–3]. For example, ACTIVLIM is being used by reference centers in Belgium to observe the course of neuromuscular disease (NMD) in adult and pediatric patients, through the Belgian NeuroMuscular Disease Registry (BNMDR) [4]. ACTIVLIM is a measure of activity limitations in everyday life in children and adults with NMD [5,6]. It consists of a questionnaire including 22 items that describe activities of daily living, among which four are specific for children and another four for adult patients. The remaining 14 items are common for all NMD patients. For each item, patients are asked to report their perceived difficulty in performing each activity without human or technical assistance on a three-category rating scale: "impossible"/"difficult"/"easy", scored as 0, 1 or 2, respectively. Activities unfamiliar to the patient that were not attempted in the last 3 months are scored as missing. The items are presented in 10 different random orders in order to avoid systematic biases related to a repeated item presentation sequence. The development and validation of the ACTIVLIM questionnaire was based on the Rasch model [7] which has become a state-of-the-art practice

* Corresponding author. Institute of Neuroscience, Université Catholique de Louvain, 53 Avenue Mounier, COSY-B1.53.04, B-1200 Brussels, Belgium. Tel.: +32 10 474425; fax: +32 10 474451.
E-mail address: Massimo.Penta@uclouvain.be (M. Penta).

http://dx.doi.org/10.1016/j.nmd.2015.12.004
0960-8966/

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

in scales development and evaluation in health measurement [8].

Questionnaires such as ACTIVLIM are more and more used to follow-up chronic conditions such as stroke, chronic obstructive pulmonary diseases or neuromuscular diseases. However, most of available questionnaires lack assessment of their sensitivity to change, also known as responsiveness [9,10]. Moreover, when used across a varied clinical case mix, it must be checked that questionnaire items are free from bias towards one or more diagnostic group(s), different stages of diseases, different age groups or, for instance, be systematically easier for females than for males. Essentially, in order to compare measures across diagnoses, stage of disease or across gender, it must be checked that the questionnaire items retain a consistent difficulty whatever individual patient is being assessed and between time of assessment [11,12]. The routine use of the ACTIVLIM through the BNMDR provides the opportunity to more deeply investigate the psychometric properties of this questionnaire as described in its original validation [5].

The present study aims to investigate the psychometric properties of the ACTIVLIM when used in clinical daily practice in a large cohort of patients with neuromuscular diseases with a representative case mix. More specifically, the aim of this study was to validate the use of ACTIVLIM as a patient-reported questionnaire in daily clinical practice in the BNMDR by (1) investigating the robustness of ACTIVLIM calibration by comparing the item hierarchy in the BNMDR with the original calibration, (2) testing the fit of ACTIVLIM data to a unidimensional construct allowing the quantitative measurement of activity limitations in patients with NMD, (3) investigating the invariance of ACTIVLIM between demographic and clinical sub-groups in the BNMDR, (4) examining the reliability in this sample and (5) measuring the patients' change in activity over one year.

2. Materials and methods

2.1. Ethics statement and data collection

The BNMDR project, which includes the ACTIVLIM data collection and analysis, has been approved by the medical ethics committees of the participating neuromuscular centers. Each patient signed a consent form to participate in the BNMDR. ACTIVLIM data were collected over 2 years (2011–2012) in 6 neuromuscular reference centers. The ACTIVLIM questionnaire is available with items presented in 10 different random orders to avoid any systematic effect due to item sequences. A set of 10 questionnaires with different orders was distributed before issuing another set. Patients received the questionnaire either directly from clinicians during visits at their neuromuscular reference centers or by mail. In the present study, 4146 valid records were included. This sample comprises 56% male, 92% adult patients, and 85% of patients were of Flemish origin (Table 1).

2.2. Data analysis

The psychometric qualities of the ACTIVLIM were investigated with RUMM2030, a comprehensive Rasch analysis software package [13–17]. The Rasch model formulates the requirements for objective measurement of latent variables such as patient-reported limitations of activity. It requires that the response of a patient to an item solely depends on the patient activity and the item difficulty (also including the relative threshold difficulty of selecting between two consecutive responses). The ACTIVLIM data were analyzed with the rating scale Rasch model [13] that assumes equidistance between consecutive thresholds across all items. This model was chosen in order to comply with the original item design [5] and to simplify the clinical interpretation of patient response patterns [18,19]. The model estimates the location of the patients (i.e. their level of activity) and of the items and thresholds (i.e. their relative difficulty) on a common underlying unidimensional linear scale of activity limitations [20]. Based on the estimated locations, the expected answer of each patient to each item can be computed and compared with the responses actually reported by the patients in order to determine how well the observed data fit the model requirements of order, invariance and unidimensionality [7,17,20–24]. The level of significance was set to 0.05 and the Bonferroni adjustment was applied for multiple analyses.

The internal validity of ACTIVLIM was examined by testing the fit of the items to the Rasch model by dividing the BNMDR sample in 6 class interval of increasing ability. The deviation from the model expectation was assessed through standardized residuals (expected range between −2.5 and +2.5) [12], and chi-square statistics [17]. Since the statistical significance of any misfit depends on sample characteristics such as sample size and targeting [25] with larger samples systematically providing higher significance, a sample size of 400 (similar to the sample size of the original calibration [5]), was used for the assessment of chi-square statistical significance [26,27]).

The measurement range, item difficulty targeting to patient abilities and measurement error were assessed through a reliability analysis. The measurement range was determined as the region between the easiest and the most difficult item threshold along the underlying continuum of activity limitations. The floor and ceiling effects in the dataset were computed as the proportion of patients located, respectively, below and above the measurement range. The person separation index (PSI, range: 0 to 1) was computed as the ratio between the spread of the patient locations (as expressed by their standard deviations corrected for measurement error) and the measurement error, indicating the extent to which distinct levels of ability can be distinguished in the sample [12,23,28]. The local dependence, i.e. the redundancy that links responses to different items once the main underlying trait is factored out and that might artificially inflate reliability [12,29] was investigated by examining the residual correlation matrix. Items presenting residual correlations above 0.3 were considered as a potential source of local dependence [30].

The invariance, or consistency, of the item difficulties between demographic and clinical patient subgroups was investigated through differential item functioning (DIF) analysis, in order to investigate whether different subgroups within the sample (e.g. males vs. females) respond in a different

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

Table 1
BNMDR dataset overview.

Parameter and level	2011 (n = 1820)	2012 (n = 2326)	2011 & 2012 (n = 4146)
Gender			
Male, n (%)	1027 (56%)	1276 (55%)	2303 (56%)
Female, n (%)	793 (44%)	1050 (45%)	1843 (44%)
Age			
Children, age of 6 to 15 y, mean of 11.01 ± 2.95 y, n (%)	159 (9%)	166 (7%)	325 (8%)
Adults, age of 16 to 92 y, mean of 50.57 ± 17.79 y, n (%)	1661 (91%)	2160 (93%)	3821 (92%)
Language			
French, n (%)	375 (21%)	265 (11%)	640 (15%)
Dutch, n (%)	1445 (79%)	2061 (89%)	3506 (85%)
Item Orders			
1 to 10, range (mean)	117–356 (182)	224–250 (233)	344–592 (415)
Symptoms duration			
Range, y (n)	2–73 (440)	1–79 (528)	1–79 (968)
<8 y	207	273	480
≥8 y	233	255	488
Stage of disease, n			
Diagnosis	47	109	156
Symptomatic Ambulatory	1256	1577	2833
Wheelchair dependent	397	519	916
Prolonged survival	15	43	58
Missing	105	78	183
NIHDI classification, n for adults/children			
Hereditary Motor and Sensory Neuropathy (HMSN)	209/26	286/38	495/64
Myotonic Dystrophy type 1 (MD1)	238/14	274/15	512/29
Amyotrophic Lateral Sclerosis (ALS)	195/0	266/0	461/0
Hereditary Spastic Paraplegia (HSP)	93/3	140/5	233/8
Limb Girdle Muscular Dystrophy (LGMD)	88/4	109/6	197/10
Spinocerebellar Ataxia (SCA)	87/0	113/0	200/0
Duchenne Muscular Dystrophy (DMD)	41/54	52/41	93/95
Facioscapulohumeral Muscular Dystrophy (FSHD)	56/1	78/4	134/5
Chronic Inflammatory Demyelinating Polyradiculoneuropathy (CIDP)	50/0	71/0	121/0
Spinal Muscular Atrophy (SMA)	30/10	45/15	75/25
Other	574/47	726/42	1300/89

way to individual items, despite equal levels of activity. The DIF analysis was conducted by splitting the sample of patients according to the following demographic and clinical subgroups: gender (male vs. female), age (different groups of age classes), administrative language (Dutch vs. French) of the center where the patient was evaluated, type of disease (proximal vs. distal NMDs), symptom duration (less vs. more than the median duration of 8 years), stage of disease (only wheelchair dependent vs. symptomatic ambulatory), and diagnosis (across the 9 most prevalent diagnoses). Age classes were obtained using two approaches: (1) division of the sample in groups of 10 years with subjects of less than 16 years old gathered in the same age group and (2) dichotomous split of the adult sample on the median age of 52 years. Two methods were used for the DIF tests. For dichotomous factors (e.g. gender) the consistency of item hierarchy was tested by determining the item difficulty for each subgroup (e.g. one analysis for males only and one for females only), then comparing both sets of item difficulty on a XY plot [24] using standard errors normalized for large samples [26]. For polytomous factors (e.g. 9 most prevalent diagnoses), the invariance was tested by a two-way ANOVA on the residuals in order to determine if the residuals vary between subgroups (e.g. diagnoses) and/or between levels of activity [31]. Similarly, the ACTIVLIM item difficulty hierarchy in the present sample was compared to the original calibration [5] and between 2011 and 2012 in the BNMDR dataset.

The change in activity limitations over time was measured as the difference in patients' activity between 2011 and 2012 for those 1128 patients who were evaluated in both years. For patients assessed twice or more in each year, only the first assessment was taken into account for this analysis. Individual patients' change was computed according to the following *t*-score:

$$t = \frac{m_2 - m_1}{\sqrt{(SE_2)^2 + (SE_1)^2}}$$

to characterize the distribution of change across time, where m_1 = patient measure in 2011; m_2 = patient measure in 2012; SE_1 = patient measure standard error in 2011 and SE_2 = patient measure standard error in 2012. Then, the patient's change was qualified as "significant improvement" for $t > 1.96$, "improvement" for $0 < t \leq 1.96$, "stable" for $t = 0$, "deterioration" for $-1.96 \leq t < 0$ or as "significant deterioration" for $t < -1.96$.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

3. Results

3.1. Psychometric properties of the ACTIVLIM in the BNMDR

Among the 22 ACTIVLIM items, bathing and showering were the less scored activities in the BNMDR adult sample with, respectively, 5% and 9% of responses missing. The distribution of missing responses did not vary among the 10 item orders. The present analyses demonstrated acceptable overall fit statistics for both items (mean ± SD: −1.02 ± 2.82) and persons (mean ± SD: −0.42 ± 1.11), where fit statistics are standardized to a mean of 0 and a SD of 1. There was a non-significant item-trait interaction (chi-square = 127.8; df = 110; p = 0.12) indicating that the relative difficulties of the items were consistently ranked by the patients whatever their ability.

Regarding individual item fit (Table 2), 16 out of 22 items exhibited negative standardized residuals, indicating a predictable response with little randomness that is typically not considered as a threat to unidimensionality. Seven of the 16 items exhibited a fit residual lower than −2.5; however, this over fit had no threat to the model. Five of the 16 items demonstrated a low chi-square probability ($p < 0.05$), that however remained above the Bonferroni adjusted threshold of 0.002 [32]. In addition, only two items, namely (1) "Running" for children only and (2) "Standing for a long time" for adults only, exhibited a positive standardized residual over 2.5 with non-significant chi-square probabilities, i.e. a response that is more systematic than expected, but not significantly. This indicates that there was no significant misfit of the items to the Rasch model.

The average patient standardized residual (−0.42 ± 1.11) is relatively close to the ideal expected value of 0 ± 1 in case of perfect fit. A closer examination of individual person fit indicated that the standardized residual was below −2.5 in 62 records and above 2.5 in 65 records. This indicates that 97% of the total records (4019 out of 4146) fitted the Rasch model requirements.

The total range of measurement in this dataset varied from −5.3 to over +5.4 logits. The distribution of patients' abilities (Fig. 1) was well targeted to the distribution of item thresholds. Only 318 records (7.7%) were extremely low (i.e. patients had rated all items as "impossible") and only 221 records (5.3%) were extremely high (i.e. patients had rated all items as "easy"). This indicates that the ACTIVLIM demonstrated low floor and ceiling effects as the range of item difficulty was well targeted for 87% of the patients in the BNMDR dataset.

A PSI of 0.95 indicates a very good reliability in the BNMDR dataset showing that 6 strata of significantly different levels of ability can be distinguished in the functional range of measurement of ACTIVLIM [33,34]. The average standard error on the item difficulties was 0.05 logits.

Examination of the residual correlation matrix highlighted that most of the correlations were below 0.3, except for 3 pairs of items that presented weak to moderate residual correlations: items "Taking a bath" and "Stepping out from a bath tub" with r = 0.41; items "Walking downstairs" and "Walking upstairs" with r = 0.46 and items "Washing one's upper body" and "Wiping one's upper body" with r = 0.61.

3.2. Comparison of BNMDR and original calibration

The stability of the item difficulty hierarchy between the BNMDR and the original calibration is illustrated in Fig. 2. Although there were slight but non-significant changes in some items' location, all items lie within the 95% confidence interval of invariant calibration. The item "Hopping on one foot" has

Table 2
Calibration of ACTIVLIM for the BNMDR dataset.

Item and statement	Group*	Difficulty (logit)	SE (logit)	Fit Residual	Chi Square	p-value
a. Hopping on one foot	C	3.19	0.09	−0.35	0.74	0.981
c. Running	C	3.00	0.09	4.78	1.80	0.877
b. Carrying a heavy load	A	2.91	0.04	−1.28	4.95	0.423
d. Walking more than 1 km	A	2.14	0.04	−0.30	8.12	0.150
f. Standing for a long time	A	1.44	0.04	4.30	2.18	0.824
g. Stepping out of a bath tub	B	1.35	0.04	−4.60	5.28	0.383
e. Walking upstairs	B	0.96	0.03	−3.53	12.40	0.030
i. Taking a bath	B	0.94	0.04	−1.30	18.86	0.002
h. Walking downstairs	B	0.59	0.03	−5.33	3.17	0.673
j. Putting on a backpack	C	0.41	0.08	0.74	0.60	0.988
l. Walking outdoors on level ground	B	−0.08	0.03	−0.28	1.83	0.873
k. Dressing one's lower body	B	−0.09	0.03	−4.72	6.57	0.255
n. Taking a shower	B	−0.27	0.04	−3.94	11.62	0.040
m. Getting into a car	A	−0.63	0.04	−3.29	14.96	0.011
o. Wiping one's upper body	B	−1.40	0.04	−1.48	4.06	0.541
s. Washing one's upper body	B	−1.49	0.04	−1.48	4.97	0.419
q. Hanging up a jacket on a hatstand	B	−1.56	0.04	−2.34	12.69	0.026
r. Sitting on the toilet	B	−1.60	0.04	−3.24	4.67	0.457
p. Putting on a T-shirt	B	−1.75	0.04	−0.74	1.83	0.872
u. Closing a door	C	−2.35	0.11	1.09	2.65	0.754
t. Opening a door	B	−2.46	0.04	2.34	1.87	0.867
v. Washing one's face	B	−3.25	0.05	2.43	2.05	0.843

* A = adult only activity, C = child only activity and B = activity common to both children and adults.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

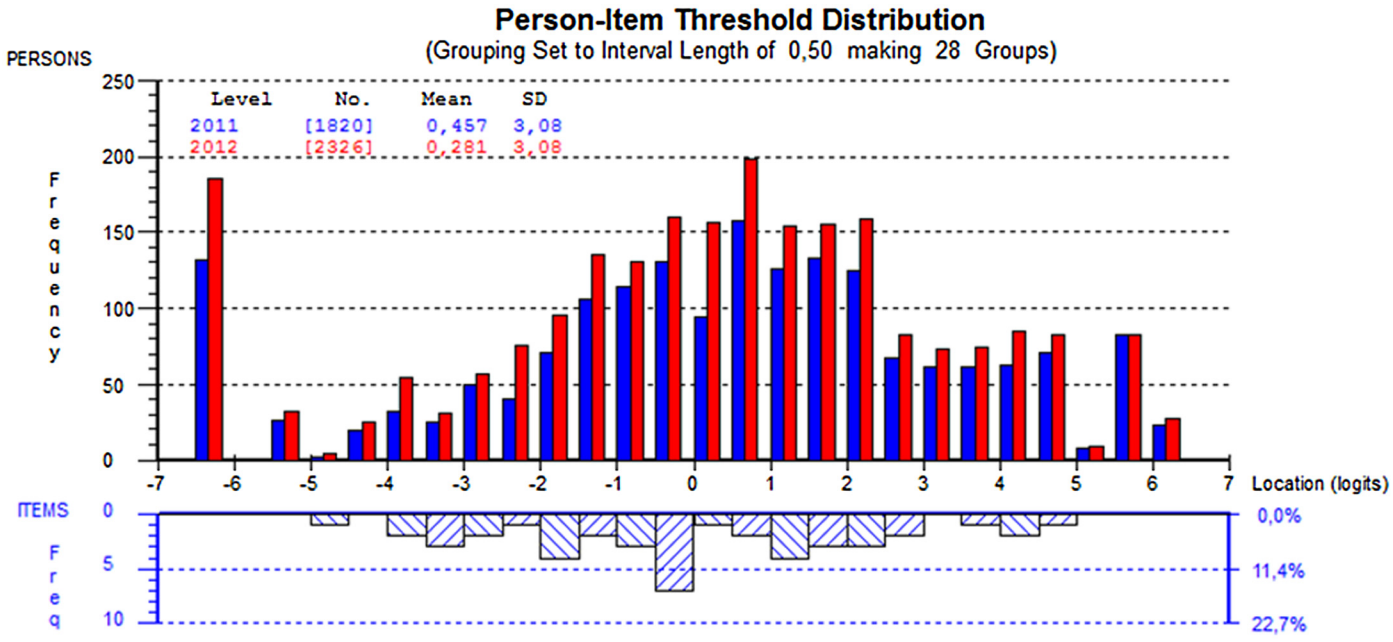

Fig. 1. Distribution of patient activity (top) in 2011 (blue) and 2012 (red) and of ACTIVLIM item thresholds (bottom). Positive locations indicate more able patients and more difficult items. An activity of zero is conventionally set at the average item difficulty. Each item is represented by two thresholds (one at the threshold between "impossible" and "difficult" responses and one at the threshold between "difficult" and "easy" responses), separated by 2.89 logits.

remained the most difficult item and item "Washing one's face" the easiest one.

In addition, there was no significant difference between 2011 and 2012 regarding the item difficulty hierarchy (Fig. 2). Therefore, data from both years were pooled together to evaluate the validity of ACTIVLIM in the BNMDR dataset.

3.3. Differential item functioning

The results of the DIF tests for dichotomous factors are presented in Fig. 3, where the item relative difficulties are compared between various complementary sub-groups of patients (e.g. males vs. females). Overall, the item difficulty hierarchy in the BNMDR dataset was not significantly different between males and females, between younger and older adults, between Dutch and French speaking patients, between patients with proximal or distal NMDs, between patients with recent or long-established symptoms and between ambulatory and wheelchair-dependent patients. However, minor exceptions were observed. For example, items e and h, namely "Walking upstairs" and "Walking downstairs", were perceived as more

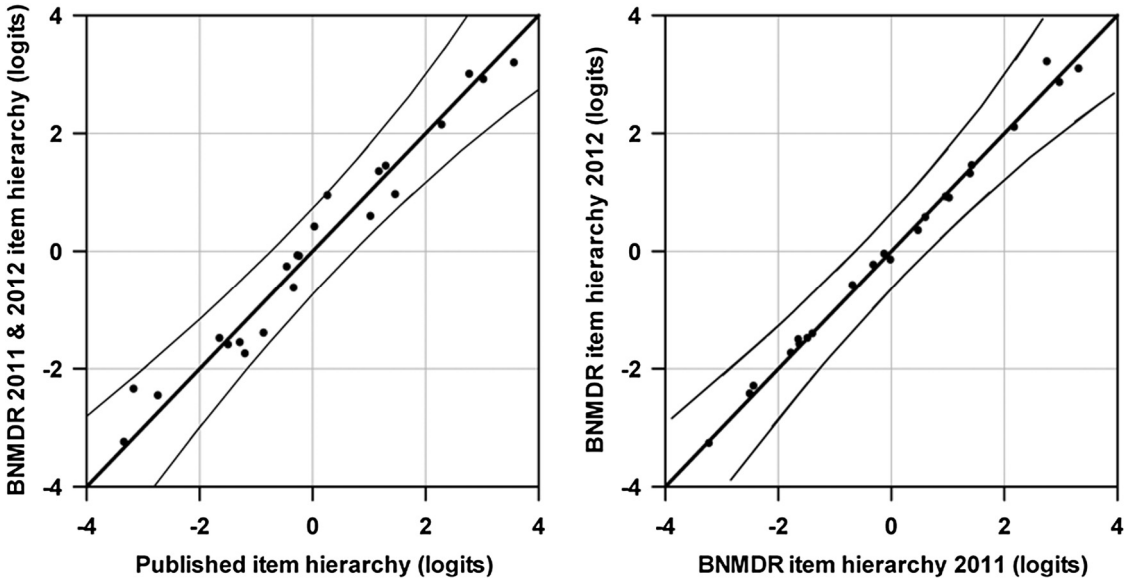

Fig. 2. Comparison of BNMDR and original calibration of ACTIVLIM (left panel) and item hierarchy in BNMDR 2011 and 2012 (right panel). More difficult items are plotted in the top/right part of each plot. All items (dots) lie within the 95% confidence interval (curved lines) of the ideal invariance (straight line), showing the same perceived item difficulty between conditions.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

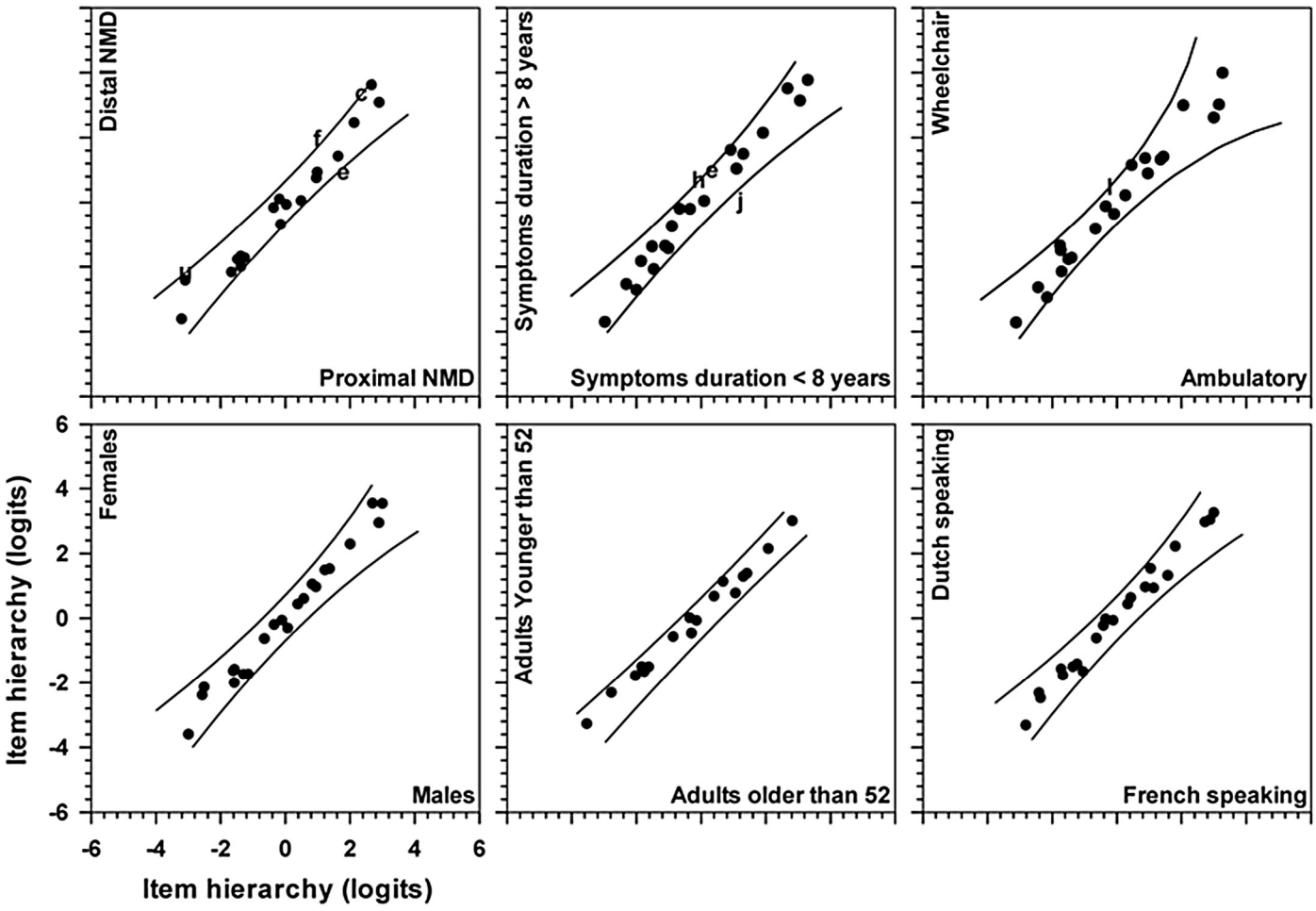

Fig. 3. Differential Item Functioning plots for 6 dichotomous sub-groups. In each plot, items (dots) lying within the 95% confidence interval (solid lines) of the ideal invariance have the same difficulty for both sub-groups. Outliers are identified by their labels. Proximal NMDs are Duchenne Muscular Dystrophy, Becker Muscular Dystrophy, Limb Girdle Muscular Dystrophy, Emery-Dreifuss Muscular Dystrophy, Polymyositis, Dermatomyositis, Adult Spinal Muscular Atrophy, X-linked Bulbo-Spinal Muscular Atrophy or Kennedy's disease; Distal NMDs are Distal Myopathy, Myotonic Dystrophy type 1, Distal Spinal Muscular Atrophy, Hereditary Spastic Paraplegia, Hereditary Motor and Sensory Neuropathy, Neuropathy associated with Paraproteinemia, Neuropathy associated with Plasma Cell Dyscrasia, Amyloidosis, Neuropathy in Systemic Disease, Other Neuropathies.

difficult by patients with symptom durations over 8 years than by patients with more recent symptoms; item l, namely "Walking outdoors on level ground", was perceived as more difficult by wheelchair-bound than by ambulatory patients and item f, namely "Standing for a long time", was perceived as more difficult by patients with distal rather than proximal NMDs.

The results for DIF tests involving polytomous sub-groups are presented in Fig. 4. Whereas, the overall item difficulty hierarchy was stable, and minor exceptions to the ideal invariance were observed across age sub-groups. For instance, some locomotion items, namely "Walking upstairs" and "Walking outdoors on level ground" were perceived as easier while some dressing and grooming activities, namely "Washing-" and "Wiping-one's upper body" and "Putting on a T-shirt" were perceived as more difficult by children than by other age groups. In addition, the slight divergence of item difficulties across the 9 most prevalent diagnoses suggests the presence of pathology-specific patterns of item difficulty. Two types of patterns are observed: (1) large deviations from the average item difficulty hierarchy are observed for diagnoses weakly represented in the sample; for instance, a large deviation in the difficulty of items "Hopping on one foot" and "Hanging up a jacket on a hatstand" is perceived by patients with Chronic Inflammatory Demyelinating Polyradiculoneuropathy (CIDP); (2) small deviations from the average item difficulty hierarchy are observed for diagnoses with larger sample size; for instance, a small deviation in the difficulty of medium difficult and easy items is perceived by patients with Amyotrophic Lateral Sclerosis (ALS).

3.4. Change in patients' activity level over time

The distribution of *t*-scores, representing individuals' change, is presented in Fig. 5. In the whole sample of 1128 patients, the level of activity has remained stable for 14% of the patients over one year. Activity significantly deteriorated in 14% of the patients while non-significant deterioration was observed in 38%. Non-significant improvement of activity was observed in

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

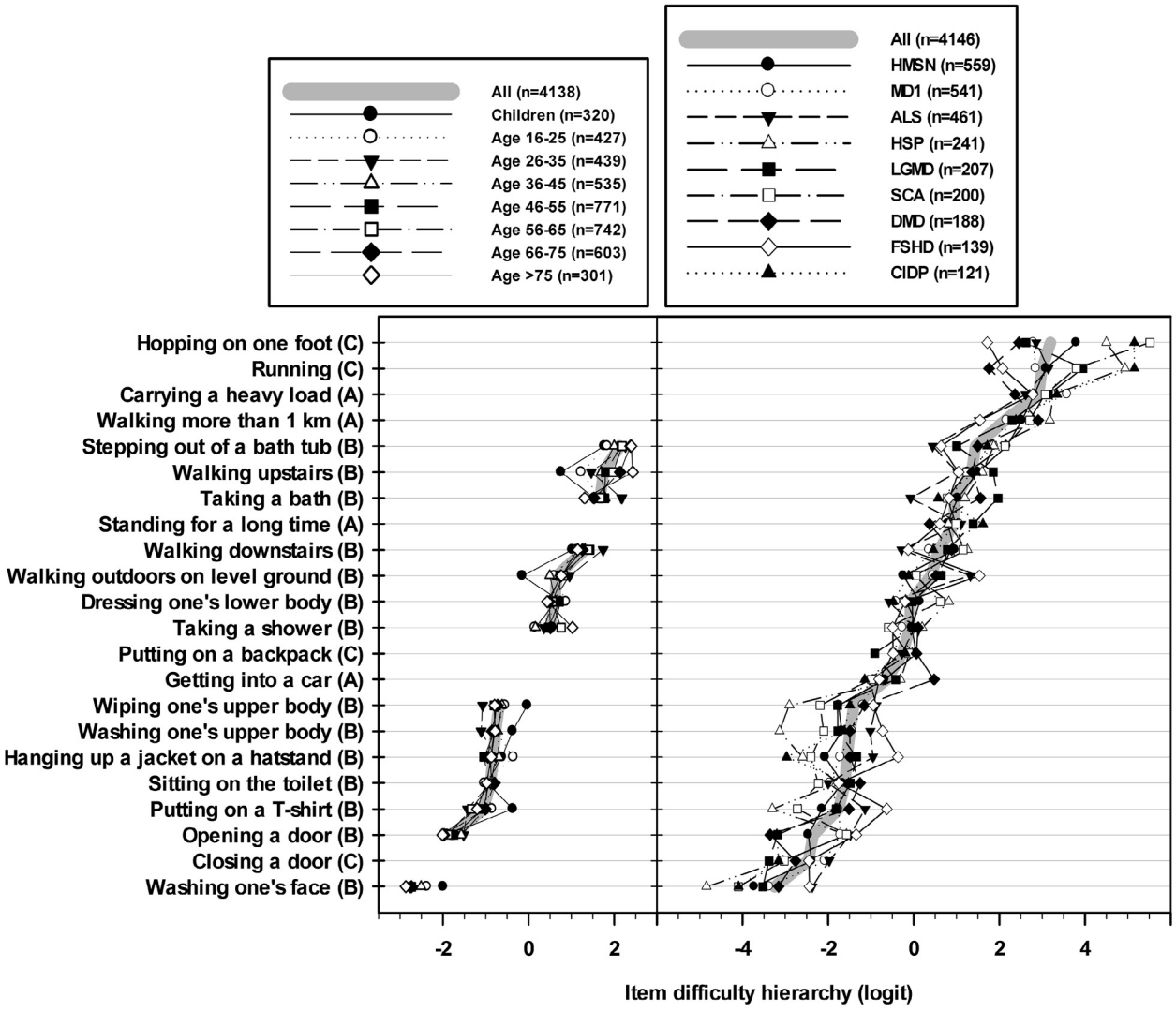

Fig. 4. Differential Item Functioning plots for age and diagnosis sub-groups. In each plot, item difficulties are plotted for each sub-group (symbols and lines) on top of the average item difficulty for the whole sample (thick gray line). Items are listed from top to bottom in decreasing difficulty order for the whole sample. Where symbols lie close to the average item difficulty, items have an invariant difficulty among sub-groups. Items with difficulties widespread between sub-groups display some DIF.

28% of the patients while 6% of them significantly improved. A comparable pattern of change was observed for all diagnostic sub-groups, except for ALS where larger deteriorations were observed over one year (t-score = -1.68 ± 2.17; $p < 0.001$) and for CIDP showing a more stable or improving condition (t-score = $+0.49 \pm 1.65$; $p > 0.1$).

4. Discussion

This study aimed at investigating the metric properties of the ACTIVLIM scale as it is used in daily practice in the BNMDR. Analyses showed a good overall and individual item and person fit to a unidimensional scale, a high reliability index and a good targeting between the scale and the population of interest. The item hierarchy was also stable over time and invariant across selected demographic and clinical factors. Analyses of sensitivity to change over one year revealed that the ACTIVLIM questionnaire has a good responsiveness, and is suitable as a tool for the follow-up of activity limitations in patients with neuromuscular diseases taking into account the dynamic of various NMD diseases. These results confirmed the validity of ACTIVLIM in adult and pediatric patients with NMD since the patient fit statistics, namely the standardized residuals, are acceptable for all diagnostic and age groups. This indicates that ACTIVLIM measures the limitation of activities in a consistent way across the diagnoses represented in the sample.

4.1. Consistency of item calibration over time

Although the metric of ACTIVLIM definition is consistent between the original calibration paper [5] and our study, minor hierarchical re-orderings appeared for some items. While the original item calibration was estimated from the response of 369 patients, the present calibration is based on a sample of 4146 records. Consequently, the average standard error on the

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

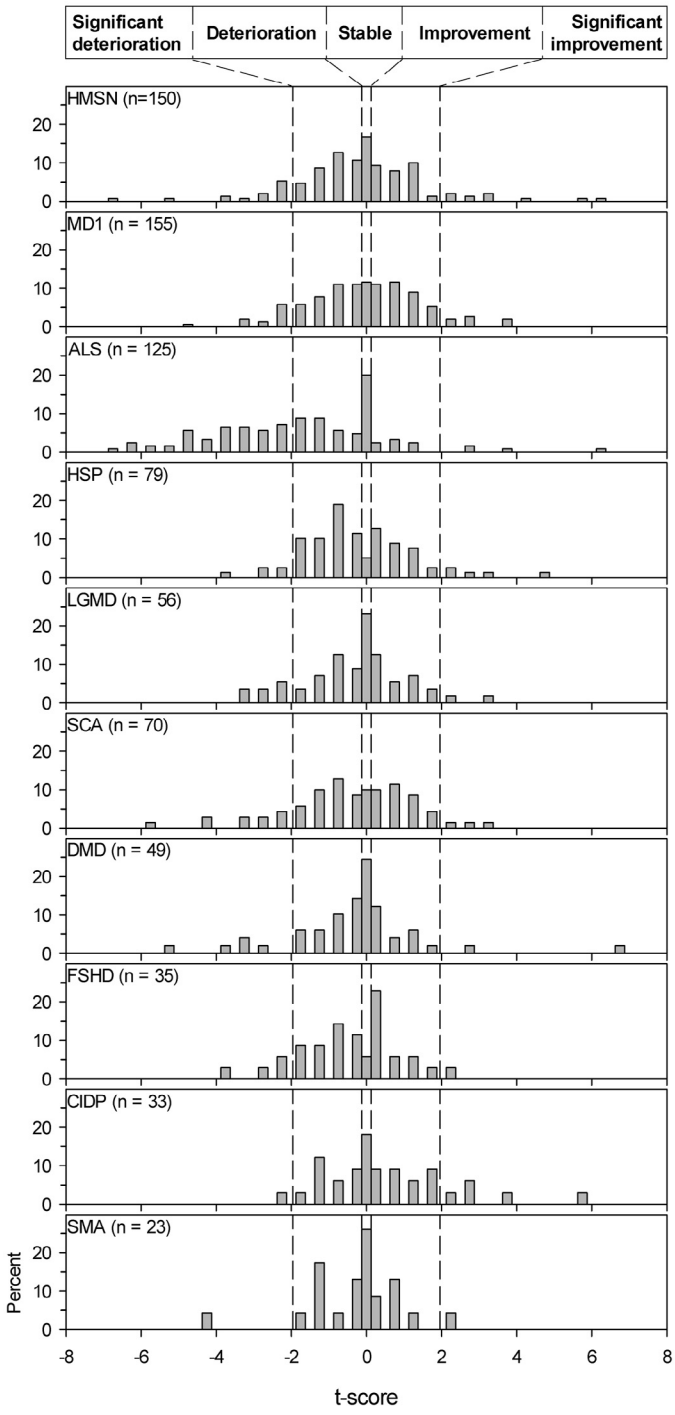

Fig. 5. Distribution of patients' change of activity from 2011 to 2012. Positive *t*-scores indicate an increase of activity for a patient over one year and negative values indicate a decrease of activity over one year.

item difficulty has dropped from 0.17 to 0.05 logits, showing that the BNMDR sample allowed an estimation of the item difficulty, i.e. the metric definition of the ACTIVLIM activity limitation scale, with an accuracy that is over 3-fold the accuracy of the original study. Although the minor re-ordering of items observed between the original [5] and the present studies might be due to the increased accuracy in the current study, it did not lead to a significant change in the item difficulty hierarchy. It is notable that the item hierarchy remained stable over time and despite differences in data collection contexts and sample size. Indeed, while in the original study, data were collected by a single rater, the present study used data collected through daily recordings in clinical practice with multiple raters. Altogether, the invariance of the item difficulty hierarchy demonstrates the robustness of the ACTIVLIM scale in the measurement of activity limitations in NMD patients.

4.2. Metric properties of ACTIVLIM

The range of non-extreme abilities measured in the BNMDR dataset varies from −5.3 to over +5.4 logits, while it ranges from approximately −5 to +5 logits in the original study [5], indicating that the BNMDR patient sample is slightly more spread than the original sample. Moreover, the Person Separation Index in the BNMDR (PSI = 0.95) and in the original study (0.96) were similar. This observation supports the reliability of the ACTIVLIM in the BNMDR dataset. Despite difference in data collection between this study and the original study [5], only a slight decrease in PSI (0.95 vs. 0.96) was observed. A reliability of 0.95 allows over 6 distinct strata of significantly different abilities to be distinguished [33]. This indicates that ACTIVLIM can discern subtle but statistically significant differences in clinical practice. However, the PSI might be inflated by response dependence [12]. Although local dependency is not reported in the original study, it has been shown that three pairs of items presented moderate residual correlations ($0.41 \leq R \leq 0.61$), which might have artificially impacted the PSI in the BNMDR. Indeed, correlated residuals are typically observed for items with comparable meanings (e.g. "Walking upstairs" and "Walking downstairs") and such redundant items may also artificially inflate reliability indices since they challenge the same aspect of the patient's ability. Alternatives to this potential issue consist either in deleting the redundant item(s) [35] or in analyzing the existing responses by merging the redundant items into one single item [17].

4.3. Invariance of the ACTIVLIM item difficulty hierarchy across demographic and clinical subgroups

The robustness of the ACTIVLIM calibration was established across various demographic and clinical subgroups. However, the largest variations in item difficulty hierarchy appeared between diagnoses. The interpretation of these differences is complicated by the case mix since large differences in diagnoses with small sample size may be less significant than small differences in diagnoses with large sample size. While the application of the common item hierarchy is supported by the good patient fit across all diagnostic groups, the observed diagnosis-specific pattern fits with the clinical picture. Activities that involve distal muscles, e.g. items "Closing a door", "Standing for a long time" and "Running" were rated as more difficult for distal than for proximal NMDs; they, indeed, require ankle stabilization or hand/arm activity. On the other hand, the item "Walking upstairs" was more difficult for proximal NMDs, requiring activity of proximal muscles such as psoas and quadriceps.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

Items "Walking upstairs" and "Walking downstairs" were more difficult for patients with symptoms lasting more than 8 years. This might be age-related since patients with long lasting symptoms will be adults or old adults rather than children and thus might be less inclined to climb stairs. Activities relative to grooming and dressing the upper body appeared easier for Hereditary Spastic Paraplegia (HSP), a disease that does not impact upper limbs, and for Amyotrophic Lateral Sclerosis (ALS) which mainly affects lower limbs. Similarly, activities that require the use of upper limbs presented a higher challenge for facioscapulohumeral muscular dystrophy (FSHD) patients whereas items requiring essentially the lower limb, e.g. "Hopping on one foot", "Running", "Walking more than 1 km" and "Standing for a long time", were relatively easier for these patients since FSHD mainly affects the neck and shoulders. The analysis of DIF across age revealed that children reported less difficulty in stairs climbing. This might be explained by their smaller sensitivity to fatigue as compared to adults. The easier outdoor locomotion might be explained by their larger interest in outdoor activities in general (e.g. playing). The observed differences in item difficulty hierarchy between diagnoses may raise the debate on generic versus disease-specific scales. Despite minor variations in ACTIVLIM item difficulty hierarchy across diagnoses, individual diagnoses demonstrated an overall good fit, with standardized residuals (mean ± SD) ranging from -0.67 ± 1.03 to 0.00 ± 1.28, when using a common item hierarchy for a variety of neuromuscular diseases. In addition, despite a total sample size of 4146 records, the small number of cases in individual diseases prevents any reliable disease-specific calibration of ACTIVLIM.

Finally, the scale showed no DIF for language, confirming the previous observation that ACTIVLIM is invariant regarding language and can be consistently used in both French and Dutch speaking communities [5].

4.4. Sensitivity to change of ACTIVLIM

The magnitude of clinical change over a relatively short time period (i.e., 1 year) in patients with neuromuscular disorders is dependent on the dynamics of the disease progression rate and on the availability of efficacious therapeutic interventions. In this respect, neuromuscular disorders can be divided in three main categories: slowly progressive, rapidly progressive, and those that are amenable to treatment with stabilization or improvement. The large majority of genetic neuromuscular disorders, in particular the large group of muscular dystrophies, spinal muscular atrophies, and hereditary neuropathies, exhibit slow and steady deterioration. In contrast, ALS is a rapidly progressive disorder, which is fatal over a couple of years. Therefore, in ALS, responsiveness is a measure of worsening, which is more pronounced in this disease than in the muscular dystrophies, the spinal muscular atrophies, and the hereditary neuropathies (Fig. 5). Immune-mediated neuromuscular disorders, including the inflammatory myopathies, myasthenia gravis, and inflammatory neuropathies, such as CIDP, are amenable to treatment. In these disorders, responsiveness usually is a measure of improvement (Fig. 5). Responsiveness can be addressed via group-level or individual-level approaches. Group-level approaches, such as the effect size and the standardized response mean determine to which extent a group of patients has changed over a given period [36–38]. Individual-level approaches, such as the *t*-score used in this study, are more advantageous from a clinical management perspective since they compare the individual patient's change with the measurement error and determine whether each individual patient has changed more than can be attributed to measurement error [10,39,40]. Using the classical cut-offs of 1.96 implies that a functional progress of more than approximately twice the measurement error is required to assert a significant change. Here in our study, calculation of *t*-scores allowed observing large and subtle differences among patients across various diagnoses. Fig. 5 clearly indicates that ACTIVLIM can measure different magnitudes of clinical change in various NMD diagnoses, even though specific questionnaires might be recommended for clinical trials that target specific diseases.

5. Conclusions

The BNMDR ACTIVLIM dataset demonstrates the validity of ACTIVLIM and the invariance of its item hierarchy across gender, age, language, symptoms duration, stage of disease and over time. Slight variations were observed across diagnostic groups. Patients' change in activity over one year indicated an overall slight, though significant, decrease in activity. ACTIVLIM allowed a longitudinal monitoring of functional status even in a large and heterogeneous cohort of NMD patients, providing specific quantification of change per individual patient and per diagnosis.

Acknowledgements

We thank all patients for their willingness to participate in this study. The members of the BNMDR scientific committee are: Prof. A. Maertens de Noordhout, Dr. L. Servais (CHR de la Citadelle, Liège); Prof. Dr. L. De Meirleir (Universitair Ziekenhuis Brussel, Brussels); Dr. J. Haan, (Inkendaal Hospitaal, Wezembeek); Prof. Dr. P. De Jonghe, Prof. Dr. B. Ceulemans (Universitair Ziekenhuis Antwerpen, Antwerp); Prof. Dr. J. De Bleecker, Prof Dr. R. Van Coster (Universitair Ziekenhuis Gent, Gent); Prof. Dr. P. Van Damme, Dr. N. Goemans (Universitair Ziekenhuis Leuven, Leuven); Prof. Dr. P. Van den Bergh, Dr. F. Christiaens (Cliniques universitaires Saint-Luc, Brussels). Dr. V. Van Casteren, Mme A. Roy (Institut Scientifique de Santé Publique, Brussels).

PVD is a senior clinical investigators of the Flemish Fund for Scientific Research (FWO, Fonds voor Wetenschappelijk Onderzoek Vlaanderen, Belgium) and is supported by a grant from the KU Leuven (GOA/11/014), by the Interuniversity Attraction Poles (IUAP) program P7/16 of the Belgian Federal Science Policy Office, by the FWO-Vlaanderen under the frame of E-RARE-2, the ERA-Net for Research on Rare Diseases (PYRAMID) and by a EU Joint Programme – Neurodegenerative Disease Research (JPND) project (STRENGTH), as well as by the Belgian ALS League.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004

References

[1] Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009;13:iii, ix–x, 1–177.

[2] Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. Milbank Q 2014;92:754–75.

[3] Franklin PD, Harrold L, Ayers DC. Incorporating patient-reported outcomes in total joint arthroplasty registries: challenges and opportunities. Clin Orthop Relat Res 2013;471:3482–8.

[4] Roy AJ, Van den Bergh P, Van Damme P, Doggen K, Van Casteren V. Early stages of building a rare disease registry, methods and 2010 data from the Belgian Neuromuscular Disease Registry (BNMDR). Acta Neurol Belg 2015;115:97–104.

[5] Vandervelde L, Van den Bergh PY, Goemans N, Thonnard JL. ACTIVLIM: a Rasch-built measure of activity limitations in children and adults with neuromuscular disorders. Neuromuscul Disord 2007;17: 459–69.

[6] Vandervelde L, Van den Bergh PY, Goemans N, Thonnard JL. Activity limitations in patients with neuromuscular disorders: a responsiveness study of the ACTIVLIM questionnaire. Neuromuscul Disord 2009;19: 99–103.

[7] Rasch G. Probabilistic models for some intelligence and attainment tests. expanded ed. Copenhagen: Danish Institute for Educational Research; 1980 With foreword and afterword by Wright BD, editor. Chicago: Press TUoC.

[8] Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med 2012;44:97–8.

[9] de Bruin AF, Diederiks JP, de Witte LP, Stevens FC, Philipsen H. Assessing the responsiveness of a functional status measure: the Sickness Impact Profile versus the SIP68. J Clin Epidemiol 1997;50:529–40.

[10] Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. J Clin Epidemiol 2001;54:1204–17.

[11] Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004;42:I37–48.

[12] Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007;57:1358–62.

[13] Andrich D. A rating formulation for ordered response categories. Psychometrika 1978;43:561–73.

[14] Andrich D. Category ordering and their utility. Rasch Meas Trans 1996;9:464–5.

[15] Andrich D. Controversy and the Rasch model: a characteristic of incompatible paradigms? Med Care 2004;42:I7–16.

[16] Andrich D. Rating scales and Rasch measurement. Expert Rev Pharmacoecon Outcomes Res 2011;11:571–85.

[17] Andrich D, Sheridan B, Luo G. RUMM 2030. Perth, WA, Australia: Rumm Laboratory; 1997.

[18] Linacre JM. Comparing "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). Rasch Meas Trans 2000;14:768.

[19] Wright BD. Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? Rasch Meas Trans 1998;12:641–2.

[20] Wright BD, Masters GN. Rating scale analysis. Chicago: MESA Press; 1982.

[21] Bond TG, Fox CM. Applying the Rasch model. Fundamental measurement in the human sciences. J Educ Meas 2003;40:185–7.

[22] Penta M, Arnould C, Decruynaere C. Developper et interpreter une echelle de mesure. Applications du modele de Rasch. Hayen, Belgium: Mardaga; 2005.

[23] Smith EVJ, Conrad KM, Chang K, Piazza J. An introduction to Rasch measurement for scale development and person assessment. J Nurs Meas 2002;10:189–206.

[24] Wright BD, Stone MH. Best test design. Chicago: MESA Press; 1979.

[25] Wang WC, Chen CT. Item parameter recovery, standard error estimates, and fit statistics of the Winsteps program for the family of Rasch models. Educ Psychol Meas 2005;65:376–404.

[26] Tristan A. An adjustment for sample size in DIF analysis. Rasch Meas Trans 2006;20:1070–1.

[27] RUMM-Laboratory. Interpreting RUMM2030. Part 1: dichotomous data. Perth, Australia: 2009.

[28] Traub RE, Rowley GL. Understanding reliability. Educ Meas Issues Pract 1991;10:37–45.

[29] Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas 2008;9:200–15.

[30] Ramp M, Khan F, Misajon RA, Pallant JF. Rasch analysis of the Multiple Sclerosis Impact Scale MSIS-29. Health Qual Life Outcomes 2009;7:58.

[31] Hagquist C, Andrich D. Is the sense of coherence-instrument applicable on adolescents? A latent trait analysis using Rasch-modelling. Pers Individ Dif 2004;36:955–68.

[32] Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995;310:170.

[33] Fisher WPJ. Reliability, separation, strata statistics. Rasch Meas Trans 1992;6:238.

[34] Wright BD, Masters GN. Number of person or item strata: (4*Separation + 1)/3. Rasch Meas Trans 2002;16:888.

[35] Mills R, Young C, Nicholas R, Pallant J, Tennant A. Rasch analysis of the Fatigue Severity Scale in multiple sclerosis. Mult Scler 2009;15:81–7.

[36] Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care 1989;27:S178–89.

[37] Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. Med Care 1990;28:632–42.

[38] Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. Int J Integr Care 2002;2:e15.

[39] Batcho CS, Nduwimana I, Olouka J, Halkin V, Kpadonou TG, Thonnard JL. Measuring functional recovery in stroke patients: the responsiveness of ACTIVLIM-stroke. J Neurol Neurosurg Psychiatry 2014;85:1337–42.

[40] Cella D, Bullinger M, Scott C, Barofsky I. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. Mayo Clin Proc 2002;77:384–92.

Please cite this article in press as: Charles Sèbiyo Batcho, Peter Y.K. Van den Bergh, Philip Van Damme, Anna J. Roy, Jean-Louis Thonnard, Massimo Penta on behalf of the BNMDR Scientific Committee, How robust is ACTIVLIM for the follow-up of activity limitations in patients with neuromuscular diseases?, Neuromuscular Disorders (2015), doi: 10.1016/j.nmd.2015.12.004