Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Journal of **Comparative Effectiveness Research**

Development and preliminary evidence of the psychometric properties of the GNE myopathy functional activity scale

Jill Mayhew[1], Nicola Bonner*[,2], Rob Arbuckle[2], Alice Turnbull[2], Alexandra Bowden[1] & Alison Skrinar[1]

[1]Ultragenyx Pharmaceutical, 60 Leveroni Court, Novato, CA 94949, USA

[2]Adelphi Mill, Bollington, Cheshire, SK10 5JB, UK

* Author for correspondence: Tel.: +44 0 1625 577 478; Fax: +44 0 1625 577 328; nicola.bonner@adelphivalues.com

Aim: GNE myopathy, a rare, severe, progressive myopathy, presents with lower extremity distal muscle weakness. The GNE myopathy functional activity scale (GNEM-FAS) evaluates the impact of GNE myopathy on functioning in adults. This paper presents the psychometric validation of the GNEM-FAS. **Patients & methods:** Validation of the GNEM-FAS was performed using data from a randomized, double-blind, placebo-controlled Phase-II study (n = 46). **Results:** Domain score distributions were acceptable. Moderate inter-item correlations (typical range, 0.40–0.70), strong item convergent and discriminant validity and high internal consistency reliability (α = 0.88–0.92) supported the instrument structure. Test–retest reliability was strong (ICC range: 0.87–0.95). Scale scores distinguished among subjects with differing disease severity ($p < 0.05$). **Conclusion:** This study provides preliminary evidence of the GNEM-FAS as a valid, reliable assessment.

First draft submitted: 9 August 2017; Accepted for publication: 20 October 2017; Published online: 15 November 2017

Keywords: functional activity • GNE myopathy • psychometric validation

GNE myopathy, also known as hereditary inclusion body myopathy, is a rare, severe, progressive, autosomal recessive myopathy presenting in early adulthood with distal muscle weakness in the lower legs. The disease is rare, and there is limited accurate information to support the true prevalence, with one report of 1–3:1,000,000 worldwide [1]. It most commonly occurs in individuals of Persian–Jewish descent [2]. The genetic basis of GNE myopathy (GNEM) as a mutation in *GNE* was determined in 2001 [2]. All GNEM patients described in the original populations have mutations in the same gene [3]. A wide variety of *GNE* mutations have been identified including those that affect either the GNE or the kinase domains [3]. There is no apparent phenotypic difference based on the domain location of the mutation. Nearly, all mutations are missense mutations and retain variable degrees of residual enzyme activity. Certain mutations have a higher frequency in particular ethnic groups, although there is no absolute specificity [3] and there are no definitive or consistent population differences for these various genotypes. Thus, phenotype/genotype correlations are not established at this time and disease expression can be variable even within the same family.

A retrospective evaluation of MRI- and CT-scans from adults with GNEM revealed significant deterioration of muscle tissue in several muscle groups of the lower extremity, including the short head of the biceps femoris, tibialis anterior, soleus and gastrocnemius medialis, while the vastus lateralis remained relatively spared [4]. The loss of viable tissue observed on MRI is associated with a distinct pattern of profound weakness in several muscle groups, including the ankle dorsiflexors and plantarflexors, hip flexors and extensors, hip adductors and abductors and knee flexors, while the knee extensors remain relatively strong [3]. Sparing of the quadriceps relative to other muscles is a defining clinical feature of the disease that can be instrumental for differential diagnosis. Functional activities are eventually limited by the progressive upper and lower extremity muscle weakness, resulting in greater dependence and disability. Such deterioration typically occurs over a period of 10–20 years [3].

Future Medicine

During the planning of a GNEM, clinical development program existing measures of functional limitations and health-related quality of life were evaluated as potential clinical trial end points. Instruments considered for use in GNEM clinical trials included the inclusion body myositis function rating scale [5], the muscular dystrophy functional rating scale [6], the NorthStar Ambulatory Assessment [7], the Amyotrophic Lateral Sclerosis functional rating scale [8], the Outpatient Physical Therapy Improvement in Movement Assessment Log [9] and the WHO Disability Assessment Schedule [10].

A review of the face validity of these instruments suggested that they were well-developed instruments with some items relevant to the functional limitations experienced by patients with GNEM. However, for any given instrument, there are also redundant items and some concepts important to GNEM patients are not assessed. Furthermore, the response scales of these existing instruments are not adequate to capture the specific experience of GNEM patients whose disease burden, including the use of assistive devices and the need for caregiver assistance, can vary considerably. Finally, having been developed for use in other conditions, these existing measures lacked adequate content validity in the specific 'context of use' of GNEM studies. As such, none of the measures were deemed adequately specific to GNEM and the unique functional limitations experienced in these patients as a result of their distinct pattern of weakness [5–11] and thus, the need to develop an instrument specific to functional impairment in GNEM was identified. The GNEM functional activity scale (GNEM-FAS) was therefore developed. This article describes the development and initial psychometric testing of the GNEM-FAS.

Patients & methods

Development of the GNEM-FAS

A GNEM patient characterization study was conducted to inform and support selection and refinement of relevant outcome measures for evaluation in clinical trials. This was a mixed methods study involving both qualitative and quantitative testing conducted to identify and measure the functional impairments associated with GNEM. Participants were 12 subjects between 18 and 70 years of age and had a confirmed diagnosis of GNEM and no comorbid conditions that would interfere with administration or interpretations of the study procedures. These participants were a subset of participants taking part in a Phase I study of aceneuramic acid-extended release many of whom went on to participate in the Phase II study in which the GNEM-FAS was implemented. They were recruited from two treatment centers in the USA. Participant racial/ethnic background was informally collected at screening; given the higher incidence of GNEM in individuals of Persian–Jewish descent and suspicion that patients of Asian or sub-Saharan Indian descent have a more severe disease course and faster progression than patients of other racial or ethnic background, collecting this information was important to further understanding GNEM. Participants were required to sign an institutional review board-approved informed consent form specific to this study prior to undertaking any study activities. All activities were conducted according to the principles of the Declaration of Helsinki.

Evaluations were conducted by a licensed physical therapist with extensive experience in clinical research and assessment of patients with neuromuscular disorders. Patients who participated in the study were evaluated through patient interviews, observation and test performance to understand the impact of the disease on the functional abilities of subjects. Results of the qualitative and quantitative assessments led to the identification of the functional limitations that are most relevant and important to patients with GNEM. These included loss of muscle strength and the impact of lower and upper extremity muscle weakness on functional ability required for independent activities of daily living. The findings of the qualitative research supported the extraction of items evaluated in established neuromuscular scales (detailed in the introduction above) as well as the identification of concepts that are relevant and important to GNEM patients.

The qualitative components included patient interviews featuring open-ended, semistructured questioning designed to elicit evidence of the experiences of patients living with GNEM. Subjects were also asked to perform activities related to the assessment of lower and upper extremity function including muscle strength tests using handheld dynamometry (HHD), walk tests (6-min walk test [6-MWT] and walking speed test), a sit-to-stand test and weighted arm lift tests. The symptoms and functional impairments identified during the patient interviews, performance-based activities and clinician observations were used to develop a conceptual framework relevant to patients with GNEM and supported the identification and development of GNEM-FAS items.

The version of the GNEM-FAS developed assesses 25 daily activities related to mobility (ten items), upper extremity function (eight items) and self-care (seven items). Activity levels for the 25 items are rated on a response scale of 0–4:

- 0: Unable or requires MAX assist of person (maximum assistance indicates that the patient is dependent on another person or two people to complete a movement or activity).
- 1: Requires MIN–MOD assist of person (minimum-to-moderate assistance is considered as meaning another person is required to safely perform the activity. The patient must be contributing to the activity or movement).
- 2: With devices and/or external support.
- 3: Slowly/some difficulty, may use orthoses for mobility domain, no external support or devices.
- 4: No limitations, no compensations, no devices.

Domain scores are calculated as the sum of the individual item scores within that domain. Total scores are calculated as the sum of the three domain scores.

The GNEM-FAS was developed as a Clinician-Reported Outcome assessment [12]. The GNEM-FAS is administered by a trained physical therapist as a clinical interview and involves a combination of clinical observation by the evaluator and clinical interview of the patient (for movements that cannot be readily observed in the clinic). Table 1 presents an overview of the GNEM-FAS items and response options.

Psychometric validation of the GNEM-FAS

Psychometric validation of the GNEM-FAS was performed using data from a randomized, double-blind, placebo-controlled Phase II study of aceneuramic acid-extended release) conducted with ambulatory adult GNEM patients across multiple sites in the USA and a single site in Israel [13]. Racial/ethnic background was collected as part of the case report form.

Clinical assessments included HHD, manual muscle testing, 6-MWT, walking speed test, sit-to-stand test, weighted arm-lift test and stair-climb test. These assessments were administered at screening, baseline and at seven additional visits over the 48-week duration of the study. Participant performance on these measures was used to categorize patients into groups for the known-groups validity assessment, and relationships between scores on these clinical assessments and the GNEM-FAS were measured as part of the concurrent validity assessment (Table 2). Additionally, the GNEM-FAS, inclusion body myositis–functional rating scale (IBM–FRS) and individualized neuromuscular quality of life (INQoL) questionnaire were administered at baseline and weeks 12, 24 and 48.

The quality of completion, item and scale-level response distributions, test–retest reliability, validity and responsiveness of the GNEM-FAS were evaluated to establish the properties of the GNEM-FAS as described below. Analyses were performed at the time points and using the *a priori* criteria for acceptability detailed in Table 2. All analyses were performed following an *a priori* statistical analysis plan and using Statistical Analysis Software® (SAS) version 9.3 (SAS Institute, NC, USA). Adelphi values performed the analyses on behalf of Ultragenyx.

Results

Development of the GNEM-FAS

12 GNEM patients (five male and seven female) with a median age of 42 years (30–62 years) participated in the qualitative and quantitative study that was conducted to gather evidence supporting the identification and development of clinical outcome assessments for GNEM. Participants were asked a series of open-ended, semistructured questions about their experience of living with GNEM in terms of symptom presentation and functional impact. Participants reported difficulties in mobility, upper extremity function and independent activities of daily living. Profound muscle weakness in the upper and lower extremities was quantified with HHD measurements. Patient comments during the interview questioning highlighted that this muscle weakness resulted in impaired motor function and quality of life. These functional impacts were further quantified with performance-based tests of walking capacity, ability to transfer independently and a reaching maneuver required for independent activities of daily living. Walking distance and speed measured during the 6-MWT and gait speed tests were less than 80% predicted values demonstrating difficulties in walking ability associated with GNEM. Difficulties in rising from sitting to a standing position, critical for independent performance, were also observed with many patients requiring the assistance of chair arms, assistive devices, stationary objects or a caregiver to complete the test. Muscle weakness and impact on the active range of motion of the shoulders and arms was also noted during the weighted arm-lift test. The findings of the qualitative research coupled with the results of the performance-based measures and observations led to the development of a conceptual framework (Table 3) and the generation of the items that were included in the GNEM-FAS.

Table 1. GNE myopathy functional activity scale.

Mobility	Activity level				
	Independent			Dependent	
	No limitations, no compensations, no devices	Slowly/some difficulty may use orthotics no external support	With devices and/or external support	Requires MIN–MOD assist of person	Unable or requires MAX assist of person
1. Turn in bed	4	3	2	1	0
2. Supine to sit	4	3	2	1	0
3. Sit to stand	4	3	2	1	0
4. Walking	4	3	2	1	0
5. Stepping up on curb	4	3	2	1	0
6. Climbing stairs	4	3	2	1	0
7. Reach to floor and recover (e.g., pick up object from floor)	4	3	2	1	0
8. Floor to standing	4	3	2	1	0
Additional mobility items			**Activity level**		
9. Running	Runs fast, no limitations, no compensations, no devices 4	3	2	1	Unable to run 0
10. Vertical jump (able to clear both feet)	Jumps vertically, no limitations, no compensations, no devices 4	3	2	1	Unable to jump vertically 0
Upper extremity			**Activity level**		
		Independent		**Dependent**	
	No limitations, no compensations, no devices	**Slowly/some difficulty, no modifications, no external support**	**With devices modifications and/or external support**	**Requires MIN–MOD assist of person**	**Unable or requires MAX assist of person**
11. Making fist (all fingers flexing fully)	4	3	2	1	0
12. Writing with pencil or pen	4	3	2	1	0
13. Hand to mouth (e.g., lifting filled glass or mug)	4	3	2	1	0
14. Cutting foods with utensils (e.g., meat with regular knife)	4	3	2	1	0
15. Carrying objects (e.g., bag of med weight groceries)	4	3	2	1	0
16. Opening doors	4	3	2	1	0
17. Opening drink bottles (e.g., small mouth water bottle)	4	3	2	1	0
18. Lifting objects overhead (e.g., med weight object on high shelf)	4	3	2	1	0
Self-care			**Activity level**		
		Independent		**Dependent**	
	No limitations, no compensations, no devices	**Slowly/some difficulty, no modifications, no external support**	**With devices modifications and/or external support**	**Requires MIN–MOD assist of person**	**Unable or requires MAX assist of person**
19. Brushing teeth	4	3	2	1	0
20. Brushing/washing hair	4	3	2	1	0
21. Dressing upper body (e.g., shirt overhead)	4	3	2	1	0
22. Dressing lower body (e.g., pants in standing)	4	3	2	1	0
23. Buttoning	4	3	2	1	0
24. Bathing (consider tub and shower)	4	3	2	1	0
25. Toileting	4	3	2	1	0

MAX: Maximum; MIN–MOD: Minimum-to-moderate.

Table 2. Psychometric analyses performed and acceptability criteria.

Property	Description	Definition/test	Criteria for acceptability
Quality of completion	Evaluated to determine item and domain level missing data at all time points to understand if any items were problematic or unacceptable	At baseline: • Frequency and percentage of missing items per subject • Frequency and percentage of missing data per item • Number of subjects with one or more missing items	Missing GNEM-FAS item response <10%
Distribution of scores	Examined to evaluate if any items exhibited a skewed distribution, or if any particular response options were overly favored	At baseline: • Number and percentage of subjects choosing each response option	No evidence of substantial floor/ceiling effects or any one response being overly favored • Item level: ≤30% responses in the highest or lowest response category • Domain level: ≤10% responses for any response category
Inter-item correlations	Assessed to identify potential redundancy among items which were highly correlated and to assess the correlation of items within and between domains	• Inter-item Pearson correlation coefficients (r)	• Low correlations: $r < 0.4$, • Moderate correlations: r = 0.4–0.7, • High correlations; $r > 0.7$
Multitrait analysis	Conducted to investigate the assumption of grouping the GNEM-FAS items into domains	• Pearson's correlation coefficients (r) corrected for overlap to examine item-scale correlations	• Item convergent validity criterion: correlations between each item and its domain $r > 0.40$ [14,15] • Item discriminant validity criterion: items have a higher correlation coefficient with its own domain than with any other domain
Internal consistency reliability	Assessed to determine the extent to which items comprising a scale measure the same construct (i.e., homogeneity of the scale)	Cronbach's alpha coefficient (α)	An alpha coefficient ≥0.70 recommended for each multiitem score[16]
Test–retest reliability	Evaluated to establish the ability of the GNEM-FAS to give reproducible results when administered twice, over a given time period, to a population with stable disease	ICC coefficients baseline to week 12 and week 12–24 in subjects defined as 'stable'. Stability criteria as follows: • Change in 6-MWT <0.2 SD† • Change in the IBM–FRS ≤2.9[17] • Change in HHD total score <0.2 SD†	Strong test–retest reliability indicated by ICC scores ≥0.70 [16]
Known-groups validity	Used to assess the extent to which instrument scores were linked to variance in individuals' known health states (i.e., known-groups). Thus, allowing the GNEM-FAS to be evaluated for its ability to discriminate between groups that would be expected to differ	Assessed at baseline with known groups defined by: • 6-MWT result ≥200 m/6-MWT of <200 m at baseline • Use of a wheelchair or a scooter/no use of a wheelchair or a scooter • Walking with assistive devices/no use of assistive devices • Tertiles of the IBM–FRS total score • Tertiles (lowest, middle or highest) of HHD UEC, HHD LEC and HHD total composite scores. The UEC score was defined as the sum of the bilateral average force (kg) for gross grip, shoulder abductors, elbow flexors and elbow extensors. The LEC score was defined as the sum of the bilateral average force (kg) for hip flexors, hip extensors, hip abductors and knee flexors. A HHD total composite score was calculated as the sum of the UEC and LEC	Significantly different GNEM-FAS scores ($p < 0.05$) between the groups compared, using *t*-tests or analysis of variance tests (depending on the group levels)
Concurrent validity	Evaluated to determine the relationship of GNEM-FAS scores with those of other measures of related or similar constructs	Pearson correlation co-efficients (r) used to examine relationships with: • 6-MWT (%) • HHD LEC score • HHD UEC score • Sit-stand test • Arm lift test • Stair climb	Pattern of correlations should reflect the content or domains. Moderate or high correlations expected for domains with similar content, low correlations for domains with dissimilar content

†SD at baseline.

6-MWT: 6-min walk test; GNEM-FAS: GNE myopathy functional activity scale; HHD: Handheld dynamometry; IBM–FRS: Inclusion body myositis–functional rating scale; ICC: Intra-class correlation; LEC: Lower extremity composite; SD: Standard deviation; SEM: Standard error of the mean; SES: Standardized effect size; SRM: Standardized response mean; UEC: Upper extremity composite.

Table 2. Psychometric analyses performed and acceptability criteria (cont.).

Property	Description	Definition/test	Criteria for acceptability
Responsiveness (or sensitivity to change)	Evaluated to ascertain the extent to which the GNEM-FAS can detect change when clinical status has changed	Assessed baseline to week 24, week 24 to week 48, baseline to week 48 Improved/worsened groups defined by: • Subjects showing ± >0.1 SD[†] change in 6-MWT • Subjects showing ± >2.9 change in IBM–FRS score • Subjects showing ± >0.1 SD[†] change in HHD total composite score results Responsiveness statistics calculated as follows: 1. SES [18] $SES = \frac{(Mean_{T2} - Mean_{T1})}{SD_{T1}}$ 2. SRM $SRM = \frac{(Mean_{T2} - Mean_{T1})}{SD_{change}}$ 3. Guyatt's statistic $Guyatt = \frac{(mean_{T2} - Mean_{T1})}{SD_{change\_stable}}$	• Small change: 0.20 ≤ SES < 0.50 • Moderate change: 0.50 ≤ SES < 0.80 • Large change: SES ≥0.80 [19]
Estimation of clinically meaningful changes	Distribution based approaches were used to provide a statistical estimate of the smallest score change which could be considered to be clinically meaningful and important	• 0.5 SD of baseline scores [20] • SEM	• MID-D > 1 SEM [21]

[†]SD at baseline.
6-MWT: 6-min walk test; GNEM-FAS: GNE myopathy functional activity scale; HHD: Handheld dynamometry; IBM–FRS: Inclusion body myositis–functional rating scale; ICC: Intra-class correlation; LEC: Lower extremity composite; SD: Standard deviation; SEM: Standard error of the mean; SES: Standardized effect size; SRM: Standardized response mean; UEC: Upper extremity composite.

Table 3. Conceptual framework in GNE myopathy (n = 12).

Physical function: mobility	Physical function: upper extremity function	Activities of daily living (ADL)
Impaired walking ability, i.e., distance and speed	Difficulty flexing fingers	Dressing, e.g., button shirts
Poor balance	Difficulty with reaching maneuvers	Walking while carrying objects, e.g., groceries
Trouble climbing steps/stairs	Difficulty lifting objects overhead	Balance while performing ADL, e.g., toothbrushing
Difficulty with transferring position including lying to sitting and sitting to standing	Fatigue	Opening jars
Inability to run	–	–
Inability to jump	–	–
Tripping/falls	–	–
Trouble moving/lifting legs	–	–
Fatigue	–	–

Psychometric evaluation of the GNEM-FAS

Psychometric analysis was conducted on GNEM-FAS data from 46 subjects with GNEM enrolled in the Phase II study (USA: n = 26, Israel: n = 20). Demographic characteristics of the study population are presented in Table 4.

Quality of completion was excellent with no missing GNEM-FAS data at baseline or any study time points. There was an absence of floor effects for all domain scores, however, a low level ceiling effect was evident for the upper extremity domain score (19.6%, n = 9) and the self-care score (15.2%, n = 7). Despite, this small ceiling effect these results indicate that the majority of subjects would be able to show score improvements at the domain and total score level (Table 5).

At the item-level mobility domain responses were mainly clustered around response categories 'three' and 'two' (see Table 5). Floor effects were seen for items nine ('Running') and ten ('Vertical Jump') with 91.3% (n = 42) and 82.6% (n = 38) of subjects, respectively, being rated at a '0' on the scale indicating no ability to perform these

Table 4. Descriptive analysis: baseline subject demographics used in psychometric analysis (n = 46).

Demographic characteristics		n (%) or statistic
Gender	Female	29 (63.0%)
	Male	17 (37.0%)
Age (years)	Mean (SD)	39.7 (10.5)
Race	White	38 (82.6%)
	Asian	8 (17.4%)
Descent	Persian–Jewish	28 (59.6%)
	Asian	9 (19.1%)
	European	6 (12.8%)
	Other	4 (8.5%)
Age at symptom onset (years)	Mean (SD)	27.6 (7.8)
Symptom duration (years)	Mean (SD)	12.2 (7.7)
6-MWT (m)	n	46
	Mean (SD)	271.1 (135.0)
	Min–max, m	71–629
HHD lower extremity composite score (kg)	n	45
	Mean (SD)	36.0 (25.3)
	Min–max	1.7–104.9
HHD upper extremity composite score (kg)	n	46
	Mean (SD)	40.4 (22.7)
	Min–max	2.4–109.6
HHD total composite score (kg)	n	45
	Mean (SD)	76.6 (45.2)
	Min–max	4.1–214.4

6-MWT: 6-min walk test; HHD: Handheld dynamometry; SD: Standard deviation.

Table 5. Descriptive statistics of the GNE myopathy functional activity scale domain and total scores at baseline (n = 46).

Statistic	Mobility score (40–0)	Upper extremity score (32–0)	Self-care score (28–0)	Functional activity total score (100–0)
Mean (standard deviation)	20.1 (6.9)	26.0 (6.1)	22.6 (4.9)	68.7 (16.1)
Min–max	6.0–38.0	9.0–32.0	5.0–28.0	22.0–94.0
Highest possible score† n (%)	0	9 (19.6%)§	7 (15.2%)§	0
Lowest possible score‡ n (%)	0	0	0	0

†Highest possible score reflects highest level of independence and skill that can be reported.
‡Lowest possible score reflects highest level of dependence that can be reported.
§Ceiling effect.

activities. This is not surprising, given the complexity of these motor activities for GNEM patients with muscle weakness. These two items were included to help capture the functioning of the small proportion of subjects who have the highest level of functioning.

For items in the upper extremity domain, responses were clustered around the higher end of the scale with a ceiling effect being evident for all items in the domain; that is for all items over 30% of subjects were rated as a 'four' indicating that they could complete the activities with no assistance. A similar result was evident for the self-care domain where a ceiling effect was evident for all but one of the items. As shown in Table 6 these item effects did not generally translate to the domain-level scores.

Confirmation of the structure of the GNEM-FAS

Construct validity of the GNEM-FAS was assessed by examining inter-item correlations, multitrait analysis and internal consistency reliability (Table 7 & Supplementary Table A). Inter-item correlations within a domain typically ranged from 0.40 to 0.70 (Mobility score = 0.18–0.83, upper extremity score = 0.28–0.73, self-care score = 0.36–

Table 6. Response distribution for GNE myopathy functional activity scale items at baseline (n = 46).

Item	Response category frequency (%)					
	4	3	2	1	0	Missing
Mobility domain						
1. Turn in bed	8 (17.4%)	16 (34.8%)	20 (43.5%)	2 (4.3%)	0	0
2. Supine to sit	10 (21.7%)	17 (37.0%)	18 (39.1%)	0	1 (2.2%)	0
3. Sit to stand	8 (17.4%)	18 (39.1%)	19 (41.3%)	1 (2.2%)	0	0
4. Walking	3 (6.5%)	21 (45.7%)	20 (43.5%)	2 (4.3%)	0	0
5. Stepping up on curb	4 (8.7%)	12 (26.1%)	20 (43.5%)	8 (17.4%)	2 (4.3%)	0
6. Climbing stairs	1 (2.2%)	7 (15.2%)	34 (73.9%)	2 (4.3%)	2 (4.3%)	0
7. Reach to floor and recover	10 (21.7%)	14 (30.4%)	18 (39.1%)	2 (4.3%)	2 (4.3%)	0
8. Floor to standing	1 (2.2%)	10 (21.7%)	23 (50.0%)	3 (6.5%)	9 (19.6%)	0
9. Running	0	4 (8.7%)	0	0	42 (91.3%)‡	0
10. Vertical jump	2 (4.3%)	5 (10.9%)	0	1 (2.2%)	38 (82.6%)‡	0
Upper extremity domain						
11. Making fist	35 (76.1%)†	7 (15.2%)	2 (4.3%)	0	2 (4.3%)	0
12. Writing with pencil or pen	34 (73.9%)†	10 (21.7%)	2 (4.3%)	0	0	0
13. Hand to mouth	37 (80.4%)†	8 (17.4%)	1 (2.2%)	0	0	0
14. Cutting foods with utensils	30 (65.2%)†	11 (23.9%)	2 (4.3%)	2 (4.3%)	1 (2.2%)	0
15. Carrying objects	16 (34.8%)†	15 (32.6%)	7 (15.2%)	2 (4.3%)	6 (13.0%)	0
16. Opening doors	26 (56.5%)†	13 (28.3%)	6 (13.0%)	1 (2.2%)	0	0
17. Opening drink bottles	21 (45.7%)†	12 (26.1%)	3 (6.5%)	5 (10.9%)	5 (10.9%)	0
18. Lifting objects overhead	16 (34.8%)†	9 (19.6%)	13 (28.3%)	1 (2.2%)	7 (15.2%)	0
Self-care domain						
19. Brushing teeth	37 (80.4%)†	3 (6.5%)	6 (13.0%)	0	0	0
20. Brushing/washing hair	32 (69.6%)†	8 (17.4%)	4 (8.7%)	1 (2.2%)	1 (2.2%)	0
21. Dressing upper body	26 (56.5%)†	13 (28.3%)	5 (10.9%)	1 (2.2%)	1 (2.2%)	0
22. Dressing lower body	9 (19.6%)	5 (10.9%)	29 (63.0%)	0	3 (6.5%)	0
23. Buttoning	31 (67.4%)†	11 (23.9%)	1 (2.2%)	1 (2.2%)	2 (4.3%)	0
24. Bathing	15 (32.6%)†	13 (28.3%)	16 (34.8%)	2 (4.3%)	0	0
25. Toileting	27 (58.7%)†	8 (17.4%)	10 (21.7%)	1 (2.2%)	0	0

†Ceiling effect
‡Floor effect

0.85). Correlations between items within the same domain were generally higher than correlations with items in other domains, thus supporting the hypothesized item-scale structure.

The GNEM-FAS items showed strong item convergent validity results, with all items exceeding the *a priori* correlation threshold of 0.40 with their own domain (Table 6). In addition, all items correlated more highly with the domain they were included in as compared with the other domains, providing equally strong evidence of item discriminant validity. While the three domains are clearly closely related, there are no concerns regarding redundancy and results support the assignment of all items to their respective domains.

Calculating Cronbach's alpha coefficients to assess internal consistency reliability generated strong results, with coefficients easily exceeding the *a priori* threshold required to demonstrate acceptable homogeneity of scale for all three domains (range: 0.88–0.92). Deleting each of the individual items in turn and recalculating Cronbach's alpha coefficients for the domain generally produced slightly lower values. This provides evidence that all items are contributing some level of unique variance to the domains scores and are related and does not provide a strong rationale for deletion of any items.

Test–retest reliability

Regardless of which of the three alternative approaches was used to define 'stability' (change in 6-MWT <0.2 standard deviation [SD]; ≤2.9 IBM-FRS score change; change in HHD <0.2 SD), almost the entire trial sample was defined as stable between baseline and week 12, providing a robust sample size for the test–retest reliability

Table 7. Summary statistics of the GNE myopathy functional activity scale domain structure at baseline (n = 46).

	Mobility score	Upper extremity score	Self-care score
Number of items in domain	10	8	7
Range of inter-item correlations[†]			
Mobility score	0.18–0.83[*]	-0.09–0.68	0.01–0.67
Upper extremity score	-0.09–0.68	0.28–0.73[*]	0.01–0.76
Self-care score	0.01–0.67	0.01–0.76	0.36–0.85[*]
Multitrait analysis[†,‡]			
Range of item-scale correlations	0.49–0.85	0.50–0.87	0.52–0.79
% of items meeting convergent validity criterion[§]	100%	100%	100%
% of items meeting discriminant validity criterion[¶]	100%	100%	100%
Internal consistency reliability			
Cronbach's alpha at baseline	0.92	0.88	0.89
Range of Cronbach's alpha values following individual item deletion	0.90–0.92	0.84–0.88	0.86–0.89

[†]All values are computed by Pearson's correlation.
[‡]Item-scale correlation corrected for overlap (relevant item removed from its scale for correlation).
[§]Correlation between each item and its own scale corrected for overlap should be at least 0.4.
[¶]Items showing significantly higher correlations with competing scales.
[*]Intra-domain item correlations.

Table 8. Correlations between GNE myopathy functional activity scale scores and measures of strength and functional capacity (n = 46).

	Mobility score	Upper extremity score	Self-care score	Functional activity total score
6-MWT (%)	0.83	0.56	0.54	0.74
HHD lower extremity composite score	0.85	0.49	0.51	0.70
HHD upper extremity composite score	0.62	0.66	0.59	0.70

All values are computed by Pearson correlation coefficients.
6-MWT: 6-min walk test; HHD: Handheld dynamometry.

analysis. High ICC reliability coefficients were observed within these stable samples across all domain and total scores (ICC range: 0.87–0.95; Supplementary Table B). Furthermore, there were no statistically significant within-group changes for any score and for any definition of the stable group. In summary, the findings provide evidence of strong test–retest reliability for the GNEM-FAS.

Known groups validity

All known groups analyses provided strong evidence of the discriminative validity of the GNEM-FAS domain scores and total score (Figure 1). There were statistically significant monotonic differences in mean scores across groups differing in disease severity as determined by paired *t*-testing or analysis of covariance (ANCOVA) analysis as appropriate. Larger mean differences were observed where groups were defined using variables assessing similar functional concepts, further supporting the validity of the domains. For example, the mobility domain score was better able to discriminate between groups based on the HHD lower extremity composite (LEC; $p < 0.001$) compared with groups based on the upper extremity composite (UEC) score ($p = 0.003$; Figure 1A).

Concurrent validity

Concurrent or convergent validity was assessed by examining correlations between GNEM-FAS scores and 6-MWT, LEC and UEC scores at baseline. Moderate-to-high correlations (range: 0.49–0.85) were seen between the GNEM-FAS total score and domain scores and all of the measures of strength and functional capacity demonstrating a high level of agreement between the measures (Table 8). More importantly, there was a logical pattern of correlations,

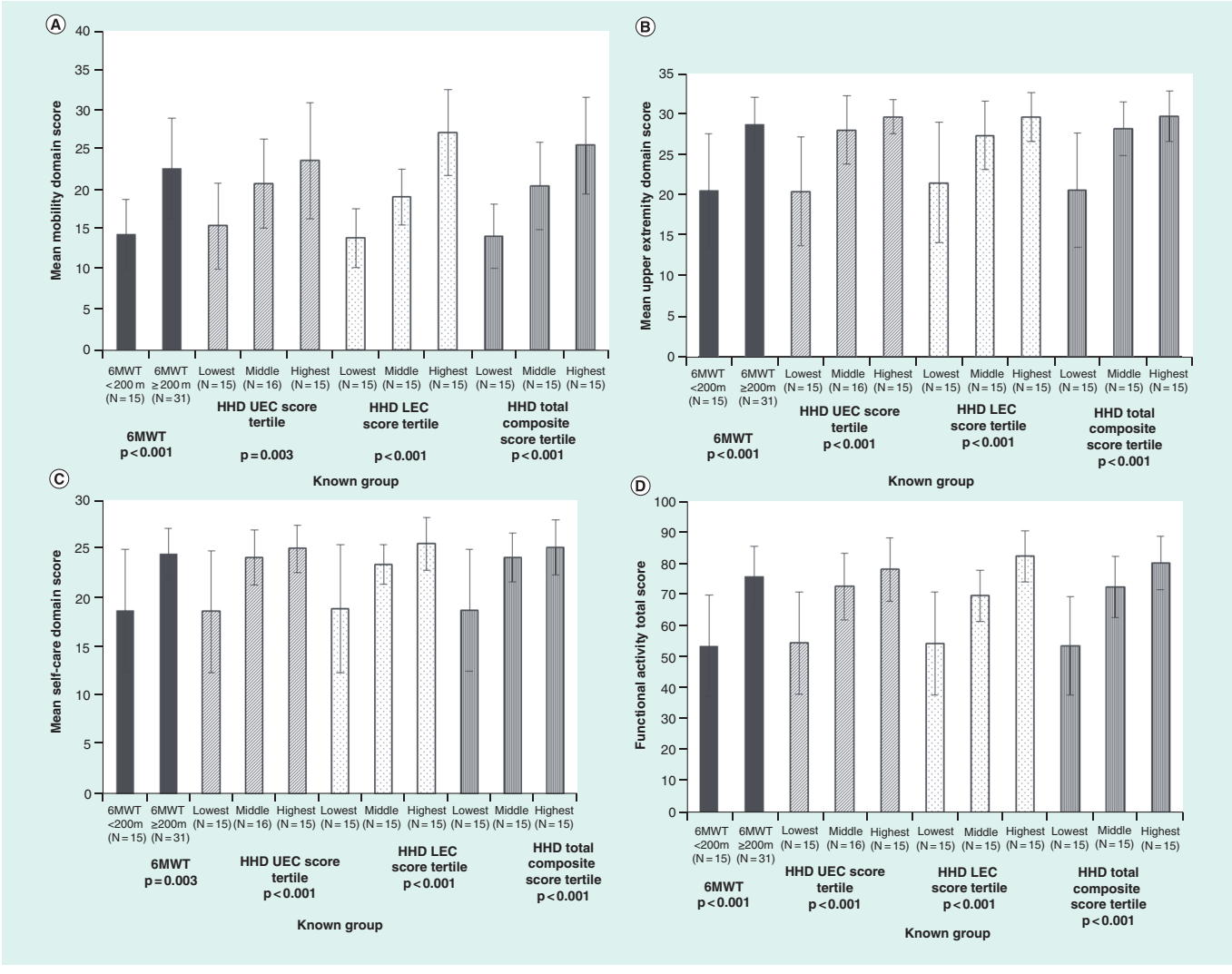

Figure 1. Known groups validity for GNE myopathy functional activity scale domain and total scores at baseline (n = 46).
6-MWT: 6-min walk test; HHD: Handheld dynamometry; LEC: Lower extremity composite; UEC: Upper extremity composite.

with the highest correlations seen between those assessments that would be expected to be more closely related, such as the GNEM-FAS mobility score and 6-MWT (r = 0.83).

Responsiveness

Due to the relatively stable nature of the cohort and the short duration of the study, few subjects were defined as 'improved' or 'worsened' in the responsiveness analysis. Due to these small sample sizes, responsiveness results can only be considered preliminary and must be interpreted with considerable caution. However, within that context, the responsiveness results were generally encouraging (see Table 9). Across all domain scores, for the 6-MWT and IBM–FRS defined change groups, mean score changes were generally negative in the 'worsened' group (indicating worsening in scores) with moderate-large effect sizes; negative in the 'no change' group, but with small or negligible effect sizes and positive (indicating improved scores) with small or negligible effect sizes in the 'improved' group. This pattern of mean change scores is as expected in an instrument sensitive to changes over time.

Given the lack of definitively appropriate anchors and the limitations described in the responsiveness analysis, distribution-based approaches were used to estimate the level of change that can be considered clinically meaningful for both the domain and total scores [14]. These analyses were based on calculation of the SEM and 0.5 SD at baseline. Results from the distribution-based analyses to estimate the levels of change in GNEM-FAS scores that can be considered important and clinically meaningful suggest changes in score for the mobility and upper extremity

Table 9. GNE myopathy functional activity scale responsiveness summary, mean score change baseline to week 24 (n = 46).

Responder definition	Responsiveness group	Mobility score				Upper extremity score				Self-care score				Functional activity total score			
		Worsened	Stable	Improved	p-value	Worsened	Stable	Improved	p-value	Worsened	Stable	Improved	p-value	Worsened	Stable	Improved	p-value
6-MWT[†]	n	5	36	4	0.148	5	36	4	0.161	5	36	4	0.588	5	36	4	0.167
	Mean change (SD)	-3.4 (4.3)	-0.8 (3.0)	0.3 (1.0)		-2.6 (4.7)	-0.6 (2.6)	1.3 (3.9)		-2.4 (5.4)	-1.2 (2.7)	-2.3 (2.1)		-8.4 (14.1)	-2.5 (5.7)	-0.8 (3.9)	
	SES	-0.81	-0.10	0.06		-0.49	-0.09	0.17		-0.79	-0.23	-0.48		-0.71	-0.15	-0.05	
IBM–FRS[‡]	n	18	23	5	0.051	18	23	5	0.095	18	23	5	0.060	18	23	5	0.015
	Mean change (SD)	-1.8 (2.9)	-0.9 (3.0)	2.0 (2.9)		-1.8 (3.2)	0.1 (1.8)	0.4 (5.5)		-2.6 (3.7)	-0.6 (2.6)	0.4 (1.8)		-6.1 (8.1)	-1.3 (4.7)	2.8 (8.3)	
	SES	-0.36	-0.11	0.78		-0.36	0.02	0.08		-0.75	-0.09	0.15		-0.54	00.06	0.41	
HHD total score[§]	n	4	33	6	0.768	4	33	6	0.708	4	33	6	0.354	4	33	6	0.606
	Mean change (SD)	-2.0 (1.8)	-1.0 (3.3)	-0.5 (3.4)		0.3 (1.7)	-0.8 (3.4)	0.0 (1.1)		-0.8 (1.9)	-1.7 (3.3)	0.2 (1.3)		-2.5 (4.1)	-3.5 (7.7)	-0.3 (4.6)	
	SES	-0.33	-0.14	-0.08		0.03	-0.13	0.00		-0.18	-0.34	0.08		-0.14	-0.22	-0.04	

[†]'No change' is defined as subjects with a change in 6-MWT results of no more than 0.2 SD at baseline. 'Improved' and 'Worsened' are defined as subjects showing a change in 6-MWT results of more than ±0.2 SD at baseline.
[‡]'No change' is defined as subjects with a change in IBM–FRS test results of no more than 0.2 SD at baseline.'Improved' and 'Worsened' are defined as subjects showing a change in IBM–FRS test results of more than ±0.2 SD at baseline.
[§]'No change' is defined as subjects with a change in HHD total composite score results of no more than 0.2 SD at baseline. 'Improved' and 'Worsened' are defined as subjects showing a change in HHD total score results of more than ±0.2 SD at baseline.
6-MWT: 6-min walk test; GNEM-FAS: GNE myopathy functional activity scale; HHD: Handheld dynamometry; SES: Standard effect size.

Table 10. Distribution based estimation of the level of changes in GNE myopathy functional activity scale scores that can be considered clinically important (n = 46).

Score	0.5 SD†	SEM‡
Mobility score (0–40)	3.45	1.98
Upper extremity score (0–32)	3.06	2.12
Self-care score (0–28)	2.46	1.66
Functional activity total score (0–100)	8.06	3.59

†1/2 SD is based on the standard deviation from baseline;
‡SEM is based on the standard deviation from baseline multiplied by the square root of one minus the internal consistency reliability from baseline.
GNEM-FAS: GNE myopathy functional activity scale; SD: Standard deviation; SEM: Standard error of measurement.

domain scores of approximately 2.0–3.0 points (Table 9). A slightly smaller magnitude of change of approximately 1.6–2.5 points is suggested for the self-care domain and a score change of approximately 3.6–8.0 is proposed as being important and meaningful for the functional activity total score. These estimates are in line with the domain score ranges since the analysis is based solely on a distributive assessment (Table 10).

Discussion

The analyses presented here provide support for the content validity, reliability and responsiveness to change of the GNEM-FAS for the evaluation of physical function in patients with GNEM. The items assessed in the GNEM-FAS were generated based the results of the qualitative patient interviews and performance based assessments in GNEM patients. This mixed methods approach to item generation and development of the conceptual framework identified that mobility, upper extremity function and independent activities of daily living were the domains of functioning most directly affected by GNEM.

The psychometric properties of the GNEM-FAS were evaluated using blinded data from a randomized double blind, placebo controlled trial of GNEM. It is acknowledged the sample size of 46 subjects is relatively small to form the basis of a psychometric evaluation of a clinical outcome assessment. However, for rare conditions such as GNEM, a flexible and practical approach must be taken. Nevertheless, all results should be interpreted with caution.

Analyses to establish the construct validity of the GNEM-FAS suggest it is well structured and provides highly consistent results over repeat administrations in a stable sample. Correlational analyses show logical relationships with other measures of strength and functional capacity while known groups comparisons suggest that GNEM-FAS scores can distinguish between groups of subjects that are known to differ in their functional capacity. There is also preliminary evidence that the instrument is sensitive to changes in underlying functional status.

Limited ceiling effects were evident for the upper extremity and self-care domain scores; however, for the majority of patients, the measure would be able to show some level of improvement following treatment, and the known groups results demonstrated that the instrument is still able to separate patients who differ in their strength and functional capacity.

Typically, either confirmatory or exploratory factor analysis [15], or modern psychometric techniques such as Item Response Theory [16] or Rasch analysis are used to evaluate the appropriateness of the item-scale structure of Clinical Outcome Assessments. However, as GNEM is a very rare disorder, the sample size of the trial population was not adequate to support those methods. Instead, multitrait analysis [17,18] and interitem correlations were used to explore the item-scale structure and confirmed that all items correlated highly with their own domain and more highly with that domain than any other, thus supporting the *a priori* conceptual framework. However, it is recommended that future studies build on these results through conduct of factor analysis or evaluating the item-scale structure through use of Item Response Theory methods where possible. Finally, analysis of the responsiveness of the GNEM-FAS was limited by the small number of subjects showing change within the study timeframe. However, in spite of this limited sample size, responsiveness results were generally promising and showed expected trends.

Given the lack of definitively appropriate anchors and the above limitations in the responsiveness analysis, only distribution approaches were used to estimate the level of change that can be considered clinically meaningful for both the domain and total scores [14]. The distribution-based approaches employed here still provide an estimation of a clinically meaningful change, as well as the smallest level of change that can be reliably detected. These analyses were based on calculation of the SEM and 0.5 SD at baseline, which have frequently and consistently

been demonstrated to be representative of MID across a range of instruments and populations [19]. This approach suggests the following levels of change can be considered clinically important: 2.0–3.0 points in the mobility and upper extremity domain scores, 1.6–2.4 points in the self-care domain and 3.6–8.0 for the total score. However, establishing the level of change that can be considered clinically important would require additional evaluation in a larger sample.

Conclusion

In summary, while further evaluation is required, the results reported here provide preliminary evidence that the GNEM-FAS is a valid and reliable assessment of functional activity in patients with GNEM. On the evidence so far evaluated, it appears appropriate for use as an efficacy end point in GNEM clinical trials providing context and support for changes in performance-based measures, and as a tool for use in clinical practice to monitor patient outcomes. Thus, the instrument can be of substantial value to aid understanding of this debilitating condition.

Summary points

- The need to develop an instrument specific to functional impairment in GNE myopathy (GNEM) was identified.
- A GNEM patient characterization study informed and supported the selection and refinement of relevant outcome measures for evaluation in GNEM clinical trials.
- The symptoms and functional impairments associated with GNEM identified during the patient characterization study supported the identification and development of GNEM functional activity scale (GNEM-FAS) items.
- The GNEM-FAS assesses 25 daily activities related to mobility, upper extremity function and self-care.
- Psychometric validation of the GNEM-FAS was performed using data from a randomized, double-blind placebo-controlled Phase II study conducted with adult GNEM patients.
- The GNEM-FAS instrument structure was supported by moderate inter-item correlations (typical range, 0.40–0.70), strong item convergent and discriminant validity and high internal consistency reliability (α = 0.88–0.92).
- Test–retest reliability of the GNEM-FAS was strong (ICC range 0.87–0.95).
- Known-group analyses confirmed that GNEM-FAS scale scores distinguished among subjects with differing disease severity ($p < 0.05$).
- Responsiveness and minimal important difference analyses were limited by the sample size and the magnitude of the observed changes during the study period.
- This study provides preliminary evidence of the GNEM-FAS as a valid, reliable assessment of functional activity in GNEM.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/cer-2017-0062

Author contributions

All of the authors contributed to the design and conduct of the study and the development of the manuscript. J Mayhew was involved in the design and conduct of the patient characterization study and the initial development of the GNEM-FAS. N Bonner, A Turnbull and R Arbuckle were involved in the design, conduct, analysis and interpretation of the psychometric validation study. A Bowden was involved in the design and interpretation of the psychometric validation study. A Skrinar was involved in the design and conduct of the patient characterization study, initial development of the GNEM-FAS and design and interpretation of the psychometric validation study. All authors were involved in the drafting of the manuscript and have approved the manuscript prior to submission.

Financial & competing interests disclosure

This work was funded by Ultragenyx Pharmaceutical Inc. (CA, USA). J Mayhew, A Bowden and A Skrinar are employees of Ultragenyx. N Bonner and R Arbuckle are current employees of Adelphi Values, a health outcomes research consultancy contracted by Ultragenyx to perform the research. A Turnbull was an employee of Adelphi Values at the time this work was conducted. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1 Khademian H, Mehravar E, Urtizberea J *et al.* Prevalence of GNE p.M712T and hereditary inclusion body myopathy (HIBM) in Sangesar population of Northern Iran. *Clin. Genet.* 84(6), 589–592 (2013).

2 Eisenberg I, Avidan N, Potikha T *et al.* The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nat. Genet.* 29(1), 83–87 (2001).

3 Huizing M, Krasnewich DM. Hereditary inclusion body myopathy: a decade of progress. *Biochim. Biophys. Acta* 1792(9), 881–887 (2009).

•• **This reference is of considerable interest as it provides an overview of GNE myopathy functional activity scale (GNEM-FAS) from a preclinical and clinical perspective.**

4 Tasca G, Ricci E, Monforte M *et al.* Muscle imaging findings in GNE myopathy. *J. Neurol.* 259(7), 1358–1365 (2012).

5 Jackson C, Barohn R, Gronseth G, Pandya S, Herbelin L. Inclusion body myositis functional rating scale: a reliable and valid measure of disease severity. *Muscle Nerve* 37(4), 473–476 (2008).

6 Lue Y-J, Su C-Y, Yang R-C *et al.* Development and validation of a muscular dystrophy-specific functional rating scale. *Clin. Rehabil.* 20(9), 804–817 (2006).

7 Scott E, Eagle M, Main M, Sheehan J. The North Star Ambulatory Assessment: Poster presented at the Annual Meeting of the British Paediatric Neurology Association. *Dev. Med. Child Neurol.* 48(Suppl. 104), 27 (2006).

8 Montes J, Levy G, Albert S *et al.* Development and evaluation of a self-administered version of the ALSFRS-R. *Neurology* 67(7), 1294–1296 (2006).

9 Guccione AA, Mielenz TJ, Devellis RF *et al.* Development and testing of a self-report instrument to measure actions: outpatient physical therapy improvement in movement assessment log (OPTIMAL). *Phys. Ther.* 85(6), 515–530 (2005).

10 Üstün TB. Measuring health and disability: manual for WHO disability assessment schedule WHODAS 2.0. World Health Organization (2010). http://apps.who.int/iris/bitstream/10665/43974/1/9789241547598_eng.pdf

11 Vincent K, Carr A, Walburn J, Scott D, Rose M. Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL). *Neurology* 68(13), 1051–1057 (2007).

12 Walton M, Powers J, Hobart J *et al.* Clinical outcome assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment - Emerging good practices for outcomes research task force. *Value Health* 18(6), 741–752 (2015).

•• **This reference is of considerable interest as it presents the attributes and characteristics that affect the measurement properties of a Clinical Outcome Assessment (COA) instrument when used as an end point in a clinical trial and explores the clinical trial factors that can affect the measurement properties that should be considered when developing or selecting COAs.**

13 Lau H, Argov Z, Caraco Y *et al.* GO 27: Randomized controlled Phase 2 study of GNE myopathy subjects treated with placebo, 6 g or 3 g doses of extended-release sialic acid (SA-ER) and an extension study at a higher 12 g dose of combination extended and immediate release sialic acid (SA). *Neuromusc. Dis.* 24(9), 923 (2014).

• **This reference is of interest as it presents an overview of the methods and results of the Phase II study, the data which were used for the psychometric validation of the GNEM-FAS.**

14 Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med. Care* 27(3 Suppl.), S178–S189 (1989).

15 Thompson B. *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications.* American Psychological Association, MD, USA (2004).

16 Hambleton RK. *Fundamentals of Item Response Theory.* Sage Publications, CA, USA (1991).

17 Steiger JH. Tests for comparing elements of a correlation matrix. *Psych. Bull.* (87), 245–251 (1980).

18 Ware JE, Harris WJ, Gandek B, Rogers BW, Reese PR. *MAP-R for Windows: Multitrait/Multi-item Analysis Program-revised User's Guide.* Health Assessment Lab, Inc., MA, USA (1997).

19 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* 4(1), 54 (2006).

20. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J. Clin. Epidemiol.* 61(2), 102–109 (2008).

21. Turner D, Schünemann HJ, Griffith LE *et al.* The minimal detectable change cannot reliably replace the minimal important difference. *J. Clin. Epidemiol.* 63(1), 28–36 (2010).