Measurement Properties of the Inclusion Body Myositis Functional Rating Scale

Sharfaraz Salam MRCP

Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK

sharfaraz.salam@nhs.net

Tara Symonds PhD

Clinical Outcomes Solutions, Folkestone, UK

tara.symonds@clinoutsolutions.com

Helen Doll DPhil

Clinical Outcomes Solutions, Folkestone, UK

helen.doll@clinoutsolutions.com

Sam Rousell MSc

Clinical Outcomes Solutions, Folkestone, UK

sam.rousell@clinoutsolutions.com

Jason Randall PhD

Clinical Outcomes Solutions, Folkestone, UK

jason.randall@clinoutsolutions.com

Lucy Lloyd-Price MSc

Clinical Outcomes Solutions, Folkestone, UK

lucy.lloyd-price@clinoutsolutions.com

Stacie Hudgens MA

Clinical Outcomes Solutions, Tucson, AZ, USA

stacie.hudgens@clinoutsolutions.com

Christina Guldberg MSc Pharm

Orphazyme A/S, Copenhagen, Denmark

christina.guldberg@hotmail.com

Laura Herbelin

Department of Neurology, University of Missouri, Columbia, MO, USA

lherbelin@health.missouri.edu

Professor Richard J. Barohn MD

Department of Neurology, University of Missouri, Columbia, MO, USA

rbarohn@health.missouri.edu

Professor Michael G. Hanna FMedSci

Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK

m.hanna@ucl.ac.uk

Arimoclomol in IBM Investigator Team of the Neuromuscular Study Group*

Professor Mazen M. Dimachkie MD†

Department of Neurology, University of Kansas Medical Center, Kansas City, KS, USA

mdimachkie@kumc.edu

Professor Pedro M. Machado PhD†

Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, London, UK; Department of Rheumatology, Northwick Park Hospital, London North West University Healthcare NHS Trust, London, UK; NIHR University College London Hospitals Biomedical Research Centre, University College London Hospitals National Health Service (NHS) Trust, London, UK

p.machado@ucl.ac.uk

* Group members listed at the end of the paper

† Joint last authors

Correspondence to: Professor Pedro M. Machado, MD PhD; Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, University College London, 8-11 Queen Square, WC1N3BG, London, UK; Tel: +442031087515; Email: p.machado@ucl.ac.uk.

Abstract word count: 250 words; **Manuscript word count:** 3980; **Tables:** 6; **Figures:** 2; **References:** 40.

Abstract

Objectives: To evaluate the validity, reliability, responsiveness, and meaningful change threshold of the Inclusion Body Myositis (IBM) Functional Rating Scale (FRS).

Methods: Data from a large 20-month multicentre, randomised, double-blind, placebo-controlled trial in IBM were used. Convergent validity was tested using Spearman correlation with other health outcomes. Discriminant (known groups) validity was assessed using standardised effect sizes (SES). Internal consistency was tested using Cronbach's alpha. Intra-rater reliability in stable patients and equivalence of face-to-face and telephone administration were tested using intraclass correlation coefficients (ICCs) and Bland-Altman plots. Responsiveness was assessed using standardised response mean (SRM). A ROC curve anchor-based approach was used to determine clinically meaningful IBMFRS change.

Results: Among the 150 patients, mean (SD) IBMFRS total score was 27.4 (4.6). Convergent validity was supported by medium to large correlations (r_s modulus: 0.42-0.79) and discriminant validity by moderate to large group differences (SES=0.51-1.59). Internal consistency was adequate (overall Cronbach's alpha: 0.79). Test-retest reliability (ICCs=0.84-0.87) and reliability of telephone versus face-to-face administration (ICCs=0.93-0.95) were excellent, with Bland-Altman plots showing good agreement. Responsiveness in the worsened group defined by various external constructs was large at both 12 (SRM=-0.76 to -1.49) and 20 months (SRM=-1.12 to -1.57). In ROC curve analysis, a drop in 2 IBMFRS total score points was shown to represent meaningful decline.

Conclusions: When administered by trained raters, the IBMFRS is a reliable, valid and responsive tool that can be used to evaluate the impact of IBM and its treatment on physical function, with a 2-point reduction representing meaningful decline.

Key messages

What is already known on this topic:

- The Inclusion Body Myositis Functional Rating Scale (IBMFRS) is a clinician-reported outcome measure to assess the functional status of Inclusion Body Myositis (IBM) patients. Despite being used both clinically and in the context of research, there is a paucity of literature on its psychometric properties.

What this study adds:

- The IBMFRS is a reliable, valid and responsive tool that can be used to evaluate the impact of IBM and future treatments. Furthermore, a 2-point reduction in the IBMFRS total score indicates a clinically meaningful decline in function.

How this study might affect research, practice or policy:

- The IBMFRS can be utilised as a robust outcome measure for IBM patients in clinical trials and identify patients demonstrating a significant clinical decline. This information is valuable for a broad spectrum of stakeholders, including patients, clinicians, researchers, pharmaceutical companies, and regulators.

INTRODUCTION

Inclusion body myositis (IBM) belongs to the idiopathic inflammatory myopathies (IIMs) class of muscle disease. It is associated with ageing and characterised by early involvement of the long finger flexors and quadriceps muscles; swallowing and respiratory function can also be affected.[1-3] IBM is a progressive and debilitating muscle-wasting disease, with increased risk of death from complications such as aspiration pneumonia and dysphagia.[4] Although a variety of drug trials have been completed in the last decade, IBM currently has no licensed treatment.[5-8]

Clinical Outcome Assessments (COAs) are crucial for measuring disease progression and the severity of IBM. Valid, reliable, and responsive COAs are imperative in clinical trials to gauge the response to potential treatments.[9-12] The IBM Functional Rating Scale (IBMFRS), established in 2008 as a disease-specific clinician-reported outcome measure (ClinRO), was adapted from the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS).

In a previous study, the content validity of the IBMFRS was confirmed along with the reliability of the measure.[13] Both patients and physicians agreed that the measure adequately captured core functional impacts of IBM. The study found good inter-rater reliability for face-to-face and video ratings, excellent intra-rater reliability for both modes, and excellent equivalence between face-to-face and phone administration.

The IBMFRS has gained popularity as an outcome measure in recent IBM clinical trials,[14-17] and it was used as primary endpoint in the recent large, multicentre, randomised, double-blind, placebo-controlled (RDBPC) trial of arimoclomol in IBM.[8] Additionally, the IBMFRS has been selected as the primary endpoint for two ongoing IBM RDBPC trials: one involving rapamycin/sirolimus (NCT04789070) and the other employing an anti-KLRG1 antibody (ABC008/Ulviprubart, NCT05721573). To date, however, the measurement properties of the IBMFRS have not been thoroughly investigated, particularly in large datasets.

Our aim was to gather information on the measurement properties of the IBMFRS, namely validity, reliability, responsiveness, and interpretability (meaningful within-person change threshold), in patients recruited to the arimoclomol in IBM trial.[8]

METHODS

Study design and population

The arimoclomol trial was a randomised, double-blind, placebo-controlled trial conducted at specialist neuromuscular centres (NCT02753530). Eligible participants diagnosed with IBM, meeting any category of the European Neuromuscular Centre research diagnostic criteria 2011,[18] had to demonstrate the ability to rise from a chair unaided and walk at least 6 meters. The study spanned 20 months, featuring both in-person and remote visits, with the trial schedule and details having previously been published.[8, 19] Patients enrolled into the arimoclomol clinical trial were broadly representative of those from other clinical trials in IBM, including ongoing efficacy clinical trials in IBM (NCT04789070, NCT05721573). Participants gave informed consent to participate in the study before taking part.

Clinical Outcome Assessments

Inclusion Body Myositis Functional Rating Scale

The IBMFRS is a ClinRO measure used to determine participants' capability and independence in 10 functional activities.[13, 14, 16, 20] Each of 10 items (swallowing, handwriting, cutting food and handling utensils, fine motor tasks, dressing, hygiene, turning in bed and adjusting covers, sit to stand, walking, climbing stairs) are graded on a 5-point ordinal scale from 0 (unable to perform) to 4 (normal). The sum of the 10 items gives a value between 0 and 40, with higher scores representing less functional limitation (i.e., better health outcome). IBMFRS raters received initial training and certification before commencing the study, with mandatory yearly training and re-certification thereafter. Also, raters were provided with a written procedure on how to apply the scale. Sites were advised to consistently employ the same evaluator for IBMFRS administration at each visit.

Patient-Reported Outcome (PRO) Measures

A Patient Global Impression of Severity (PGIS) was included to measure the impact of the disease. The PGIS asked, "Considering all aspects of your inclusion body myositis and its impact on your day to day activities (e.g., dressing, walking, bathing) right now, would you say that the impact is currently…", and was scored from 0 to 5 (none to very severe): 0=none, 1=very mild, 2=mild, 3=moderate, 4=severe and 5=very severe.

A Patient Global Impression of Change (PGIC) was included to assess self-perceived change in the ability to conduct daily activities since the start of study medication. The PGIC was scored as follows: 0=very much worse, 1=much worse, 2=a little worse, 3=no change, 4=a little improved, 5=much improved, and 6=very much improved.

The Short Form 36-Item Survey (SF-36) measures health-related quality of life and was scored in accordance with existing guidelines for the instrument.[21] Scores range from 0-100, with higher scores representing better health status. The SF-36-Physical Functioning (SF36-PF) and the SF-36 Physical Component Summary (SF36-PCS) scores were used in our analyses.

The Health Assessment Questionnaire-Disability Index (HAQ-DI), a self-reported measure, was included to assess the level of functional ability; questions can be grouped in 8 categories of functioning: dressing and grooming, rising, eating, walking, hygiene, reach, grip, and usual activities. The score ranges from 0 to 3, with higher scores representing more disability.[22]

Performance Outcome (PerfO) Measures

Patients were assessed with the six-minute walk test (6MWT) and modified timed up and go (mTUG).[23] Hand grip strength was tested with a Jamar Dynamometer; the maximum result (in kg) for the strongest hand (as determined at baseline) was used in the analyses. Manual muscle testing (MMT) was used to assess the strength of 24 different muscles; the scores were converted to numerical values from 0 to 10 before a total score was calculated as an average across the 24 muscles.

Statistical analyses

Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) recommendations were followed to test and report measurement properties.[24, 25] Descriptive statistics were used to characterise the sample. All analyses were performed using SAS version 9.4. When applicable, the statistical significance tests were 2-sided, with a threshold of p<0.05.

Construct validity

Convergent validity is the degree to which the domains of a COA tool are associated with those of another tool known to measure the same construct. Convergent validity was assessed at baseline using Spearman correlations calculated between the IBMFRS total score and the HAQ-DI, SF36-PF, SF36-PCS, mTUG, 6MWT, hand grip strength and MMT. Correlations were considered weak if the resulting coefficient was <0.3; moderate if between 0.3 and <0.5; and strong if ≥0.5.[26]

Known groups (discriminant) validity is the ability of an instrument to discriminate between groups of individuals known to differ in terms of the construct of relevance, i.e., between clinically distinct groups hypothesized a priori. Known groups validity was assessed at baseline to determine whether the IBMFRS differed between groups based on the HAQ-DI score and the PGIS. Groups of HAQ-DI scores were created to indicate mild (score 0 to 1), moderate (score >1 to 2), and severe disability (score >2 to 3).[22] Four PGIS categories were considered: very mild (category 1), mild (category 2), severe (category 3), and very severe (category 4); there were no subjects at baseline that were scored either "none" (category 0) or "very severe" (category 5). Known groups validity was assessed using analysis of covariance (ANCOVA) adjusted for gender, age, and race; least squares (LS) means, and standard errors (SE) were derived from the ANCOVA model and two-sided p-values for the difference in means between adjacent groups were determined. In addition, standardized effect sizes (SES) were calculated by dividing the difference in scores between consecutive groups by the pooled group SD. Cohen's guidance was used to interpret the magnitude of SES: small (0.2 to <0.5), medium (0.5 to <0.8), and large difference (≥0.8).[26]

Reliability

Medical and health-related decisions by patients and clinicians rest on the assumption that differences among and within patients exist and have important implications. Survey instruments, therefore, are useful only to the degree to which they reliably and accurately reflect true psychological or health-related differences. Reliability reflects the extent to which differences in patients' observed scores are consistent with differences in their true scores as opposed to measurement error.

Cronbach’s alpha statistic at baseline was used to assess internal consistency reliability of the IBMFRS. An overall alpha value was calculated together with item-level alpha values showing the change in alpha following the exclusion of each item in turn. Alpha values ≥0.70 are considered adequate. Alpha values >0.90 may indicate an overly homogenous measure, or where the measure contains too many items, where items are redundant due to excessive similarity.

Intra-rater reliability was assessed between adjacent periods at which it was possible to determine stability, when the PGIS was also administered. The PGIS was administered every 4 months, and intra-rater reliability was measured at these timepoints in stable subjects who reported no change on the PGIS. Three periods were defined for identifying stable patients: between baseline and month 4, between month 4 and month 8, and between month 8 and month 12.

Equivalence of telephone versus face-to-face (F2F) in-clinic administration of the IBMFRS was measured in subjects who reported no change on the PGIS (i.e., stable subjects) between months 8 and 12. Most assessments at these time points were in-clinic, with the assessment at month 10 being by telephone. Patients who did not change according to the PGIS scores at month 12 versus month 8 were assumed to have not changed at month 10 also.

Agreement was assessed using intraclass correlation coefficients (ICC) with 95% confidence intervals (CIs). The ICCs were calculated using the Shrout-Fleiss reliability formula for calculating absolute agreement on a single (domain) score based on a 2-way mixed-effect ANOVA with a factor corresponding to modality (F2F vs telephone) and another related to patient. An ICC ≥0.70 or greater is considered desirable, with an ICC ≥0.80 considered to indicate excellent reliability. Agreement across the scale of the IBMFRS was also visualised by Bland-Altman plots. Due to the increased frequency of COVID-19 induced telephone visits after month 12, reliability was not measured after this time point.

Responsiveness

Responsiveness refers to the ability of an assessment to detect change where change exists. In the longitudinal hypothesis testing analysis, Spearman’s correlation coefficients were used to assess the degree of association between change on the IBMFRS and change on the reference measures. Correlations were considered weak if the resulting coefficient was <0.3; moderate if between 0.3 and <0.5; and strong if ≥0.5.[26]

In the magnitude of change analysis, PGIS, HAQ-DI, SF36-FCS, mTUG and MMT were used to stratify IBM patients into change groups according to the corresponding score. Paired t-tests were used to evaluate the within-group differences in IBMFRS change scores between two time points, and the standardized response mean (SRM) was calculated by dividing the mean change score between baseline and the subsequent time point by the SD of the change score. The magnitude of SRM was interpreted based on Cohen's recommendations as outlined above.[26]

Meaningful change threshold

The purpose of an anchor measure is to identify change on the COA measures that represents patient perceived improvement or deterioration. Two single-item measures (PGIS, PGIC) were selected to be evaluated as possible anchors. Both measures satisfy the recommendation that anchors should be less complex to interpret than the endpoint they are used to assess.

An anchor-based approach using PGIS and PGIS tools as external constructs was used to help determine meaningful change in the IBMFRS. ROC analyses were used to determine the optimal threshold for meaningful decline in the IBMFRS total score, referred to as the best cut point, i.e., the IBMFRS total score deterioration (change) that best discriminates between pre-defined binary outcomes on the PGIS and PGIC anchors.[27, 28] PGIC binary scoring was defined as (1) "worsened", comprising: "very much worse", "much worse," or "a little worse"; versus (0) "no change or improved", comprising: "no change", "a little improved," "much improved," or "very much improved". PGIS binary scoring was defined as (1) "worsened", comprising: >=1 category worsening within the scale; versus (0) "no change or improved", comprising: no change in PGIS category or >=1 category improvement within the scale. The distance to the (0,1) point of the ROC curve and the Youden's index were used to define best cut points, as these are the two methods that provide the best balance between sensitivity and specificity.[29-31]

RESULTS

Study Population

The total number of patients analysed at baseline was 150. Participant age ranged from 48 and 89, with a mean age of 67.2 (SD 8.1) (**Table 1**). There was a male preponderance (114 [76%]) in the population and most of the participants were of white race (143 [95.3%]). Mean IBMFRS total score was 27.4 (SD 4.6), reflecting overall moderate disability.

Table 1. Baseline clinical and demographic characteristics (N=150)

Sex	
Female	36 (24.0%)
Male	114 (76.0%)
Age, years	67.2 (8.1)
Age at diagnosis, years	63.4 (8.3)
Race	
Asian	3 (2.0%)
Black Or African American	1 (0.7%)
White	143 (95.3%)
Mixed	1 (0.7%)
Other	2 (1.3%)
Country	
UK	34 (22.7%)
USA	116 (77.3%)
CN1A antibody positive	78 (52%)
IBMFRS total score	27.4 (4.6)
HAQ-DI total score	1.18 (0.59)
6MWT, m	325.2 (100.3)
mTUG, m/s	0.50 (0.28)
MMT total score	7.7 (1.0)
Hand grip strength, kg	13.0 (11.0)
Knee extensor strength, kg	13.8 (13.2)
SF36-PF	34.3 (21.8)
SF36-PCS	37.2 (8.0)

Data are n (%) or mean (SD). Results for hand grip and knee extensor strength are for the stronger limb, as identified at baseline. When analysing the mTUG, the reciprocal value of the measured time multiplied by the planned total distance of 6 meters was used; this corresponds to analysing the velocity of the walking speed expressed in meters per seconds (m/s), including the time spent for standing up and sitting down again, and allows the adoption of the value "zero" for patients unable to perform the assessment. CN1A, cytosolic 5'-nucleotidase 1A; HAQ-DI, Health Assessment Questionnaire Disability Index; IBMFRS, Inclusion Body Myositis Functional Rating Scale; MMT, Manual Muscle Testing; mTUG, modified Timed Up and Go; SF36-PF, Short form-36 Health Survey physical functioning, SF36-PCS, Short Form-36 Health Survey physical component score; 6MWT, 6-min walk test.

Construct Validity

Convergent validity

The correlations between the HAQ-DI (r=-0.79), SF36-PF (r=0.53) and SF36-PCS (r=0.46) were medium to large, supporting convergent validity of the IBMFRS, with the HAQ-DI having the largest correlation. PerfO assessments, namely MMT total score (r=0.58), mTUG (r=0.64), 6MWT (r=0.62), and, to a lesser extent, Hand Grip Strength (r=0.42) showed moderate to strong correlations with the IBMFRS, again supporting convergent validity of the IBMFRS.

Known groups validity

Data to support known groups (discriminant) validity is presented in **Table 2**. IBMFRS scores decrease progressively the greater the severity as indicated by the PGIS and HAQ-DI. For patients categorised as mild vs very mild (SES=0.73, p=0.158) and moderate vs mild (SES=0.51, p=0.028) on the baseline PGIS scores, the SES values indicated a moderate difference versus the adjacent (lower) category. A large difference was noted for those patients with a score classified as severe vs moderate, with a SES of 0.98 (p=0.001). For patients categorised as moderate vs mild (SES=1.59) and severe vs moderate (SES=1.32) on the HAQ-DI (p<0.001 in both groups), large SES differences were observed.

Table 2. Known-groups validity of the Inclusion Body Myositis Functional Rating Scale versus the PGIS and the HAQ-DI reference measures at baseline

Reference Measure	Category (score achieving category)	n	IBMFRS LS Mean (SE)	p-value	SES
Baseline PGIS	Very Mild [category 1]	6	32.9 (2.2)		
	Mild [category 2]	27	29.9 (1.9)	0.158	0.73
	Moderate [category 3]	89	27.7 (1.7)	0.028	0.51
	Severe [category 4]	14	23.5 (2.0)	0.001	0.98

Baseline HAQ-DI	Mild [score 0 to 1]	67	32.0 (1.2)		
	Moderate [score >1 to 2]	68	26.7 (1.2)	<0.001	1.59
	Severe [score >2 to 3]	13	22.5 (1.5)	<0.001	1.32

The LS mean and SE are derived from an ANCOVA adjusting for age, sex and race. The two-sided p-value is from the difference in means between adjacent groups derived from ANCOVA. The SES is calculated by dividing the difference in scores between consecutive groups by the pooled group SD. There were no subjects at baseline that were scored either "None" (category 0) or "Very Severe" (category 5). HAQ-DI, Health Assessment Questionnaire - Disability Index; IBMFRS, Inclusion Body Myositis Functional Rating Scale; LS, least squares; PGIS, patient global impression of severity; SE, standard error; SES, standardized effect size.

Reliability

Internal Consistency

The overall Cronbach's alpha coefficient was 0.79, with the coefficient after exclusion of each of the 10 items ranging from 0.75 to 0.81), which supports an adequate consistency of the IBMFRS (**Table 3**). The exclusion of swallowing resulted in the greatest increase in consistency, with an alpha coefficient of 0.81.

Table 3. Internal consistency for the Inclusion Body Myositis Functional Rating Scale at baseline (N=150)

Item	**Cronbach's alpha coefficient**
Overall	0.79
Item 1 - Swallowing	0.81
Item 2 - Handwriting	0.78
Item 3 - Cut Food/Handle Utensil	0.76
Item 4 - Fine Motor Tasks	0.78
Item 5 - Dressing	0.75
Item 6 - Hygiene	0.75
Item 7 - Turn Bed/Adjusting Clothes	0.77
Item 8 - Sit To Stand	0.78
Item 9 - Walking	0.76
Item 10 - Climbing Stairs	0.77

Intra-rater reliability

IBMFRS total scores taken at all the three defined periods achieved ICCs >0.80 and supported strong intra-rater reliability of the IBMFRS: stable patients between baseline and month 4 (n=78), ICC=0.84 (95%CI=0.77-0.90); between months 4-8 (n=77), ICC=0.85 (95%CI=0.77-0.90); and between months 8-12 (n=78), ICC=0.87

(95%CI=0.80-0.91). In addition, Bland-Altman plots showed a good agreement between IBMFRS total scores at first and second assessments in stable patients (**Figure 1**).

Equivalence of telephone versus face-to-face (F2F) in-clinic administration

In stable patients, the ICCs for equivalence were notably high at 0.95 (95% CI = 0.92-0.97) when comparing the in-clinic administration of the IBMFRS at 8 months versus over-the-telephone administration at 10 months. Similarly, a high ICC of 0.93 (95%CI=0.89-0.96) was observed when comparing the in-clinic administration at 12 months versus over-the-telephone administration at 10 months. Bland-Altman plots showed a good agreement between IBMFRS total scores at consecutive in-clinic and over-the-telephone assessments in stable patients (**Figure 2**).

Responsiveness

Longitudinal hypothesis testing

The correlations between IBMFRS change scores and change scores for different COAs calculated at months 12 and 20 are presented in **Table 4**. The most robust associations with IBMFRS change were observed with HAQ-DI change at month 12 and month 20, with corresponding coefficients of -0.50 and -0.54, respectively, followed by the mTUG change (coefficients of 0.36 and 0.41 at month 12 and month 20, respectively). Change score correlations with other COAs were weak to moderate (**Table 4**).

Table 4. Correlation between Change in the Inclusion Body Myositis Functional Rating Scale and Change in Reference Measures from Baseline to Months 12 and 20.

Reference measure	**Correlation at month 12 (IBMFRS)**	**Correlation at month 20 (IBMFRS)**
HAQ-DI	-0.50	-0.54
SF36-PF	0.25	0.27
SF36-PCS	0.19	0.22
mTUG	0.36	0.41
MMT	0.20	0.23
PGIS*	-0.31	-0.27

*PGIS correlation assessed with polyserial correlation and the rest of COAs assessed with Spearman's coefficient. Results for hand grip and knee extensor strength are for the stronger limb, as identified at baseline. When analysing the mTUG, the reciprocal value of the measured time multiplied by the planned total distance of 6 meters was used; this corresponds to analysing the velocity of the walking speed expressed in meters per seconds (m/s), including the time spent for standing up and sitting down again, and allows the adoption of the value "zero" for patients unable to perform the assessment. HAQ-DI, Health Assessment Questionnaire

Disability Index; IBMFRS, Inclusion Body Myositis Functional Rating Scale; MMT, Manual Muscle Testing; mTUG, modified Timed Up and Go; PGIS, Patient Global Impression of Severity; SF-36 PCS, Short Form-36 Health Survey physical component score; 6MWT, 6-min walk test.

Magnitude of change

Change in the IBMFRS by degree of change in the PGIS, HAQ-DI, SF36-PCS, mTUG and MMT, between baseline and months 12 and 20, is presented in **Table 5**. Analyses at 12 and 20 months yielded similar results. The greater the extent of worsening in the PGIS the greater the reduction, or worsening, in mean IBMFRS change by months 12 and 20, with the greatest IBMFRS drop observed in the markedly worsened group (at least 2 categories of worsening): mean reductions of -3.50 at month 12 and -3.83 at month 20 (both p=0.015), compared with very little change in the improved group: a mean increase, or improvement, of 0.25 at month 12 and a decrease of –0.09 at month 20. Moderate to large SRMs were observed for the 2 worsened groups at both time points, with the markedly worsened group having a large SRM: -≈1.50 at both time points.

Table 5. Change in the Inclusion Body Myositis Functional Rating Scale by degree of change in the PGIS, HAQ-DI, SF-36 Physical Domain, mTUG and MMT, between baseline and months 12 and 20.

	n	Mean change in IBMFRS (SD)	Median change in IBMFRS	SRM	p-value
Change Level PGIS					
Month 12					
Markedly worsened (>=2 category worsening)	6	-3.50 (2.35)	-3.50	-1.49	0.015
Worsened (1 category worsening)	29	-2.17 (3.36)	-2.00	-0.65	0.002
Stable (no category change)	69	-1.00 (3.23)	-1.00	-0.31	0.012
Improved (>=1 category improvement)	16	0.25 (2.65)	1.00	0.09	0.711
Month 20					
Markedly worsened (>=2 category worsening)	6	-3.83 (2.56)	-5.00	-1.50	0.015
Worsened (1 category worsening)	33	-3.45 (3.73)	-3.00	-0.93	<0.001
Stable (no category change)	71	-2.30 (3.77)	-1.00	-0.61	<0.001
Improved (>=1 category improvement)	11	-0.09 (2.47)	0.00	-0.04	0.905
Change Level HAQ-DI					
Month 12					
Severely worsened (>1 point worsening)	2	-2.50 (4.95)	-2.50	-0.51	0.605
Markedly worsened (0.5 to 1 points worsening)	21	-4.48 (2.94)	-4.00	-1.52	<0.001
Worsened (0.25 to 0.5 points worsening)	31	-2.13 (3.20)	-3.00	-0.67	<0.001
Stable (absolute changes of <0.25 points)	69	-0.49 (2.89)	-1.00	-0.17	0.161
Improved (>=0.25 points improvement)	12	0.42 (2.91)	0.50	0.14	0.629

Month 20					
Severely worsened (>1 point worsening)	4	-7.50 (1.00)	-7.00	-7.50	<0.001
Markedly worsened (0.5 to 1 points worsening)	38	-4.89 (3.39)	-5.00	-1.44	<0.001
Worsened (0.25 to 0.5 points worsening)	18	-3.11 (3.45)	-2.00	-0.90	0.001
Stable (absolute changes of <0.25 points)	57	-1.23 (3.02)	-1.00	-0.41	0.003
Improved (>=0.25 points improvement)	13	0.15 (2.12)	0.00	0.07	0.798
Change Level SF36-PCS					
Month 12					
First quartile of worsening	32	-2.66 (3.11)	-2.00	-0.86	<0.001
Second quartile of worsening	33	-1.64 (3.26)	-1.00	-0.50	0.007
Third quartile of worsening	33	-0.79 (3.38)	-1.00	-0.23	0.190
Fourth quartile of worsening	33	-0.91 (3.42)	-1.00	-0.27	0.137
Month 20					
First quartile of worsening	32	-4.09 (3.67)	-3.50	-1.12	<0.001
Second quartile of worsening	33	-2.48 (4.13)	-2.00	-0.60	0.002
Third quartile of worsening	33	-1.79 (2.70)	-1.00	-0.66	<0.001
Fourth quartile of worsening	32	-2.09 (3.60)	-2.50	-0.58	0.002
Change Level mTUG					
Month 12					
First quartile of worsening	31	-2.87 (2.25)	-3.00	-1.28	<0.001
Second quartile of worsening	32	-1.84 (3.73)	-2.00	-0.50	0.009
Third quartile of worsening	32	-1.09 (3.47)	-1.00	-0.32	0.084
Fourth quartile of worsening	32	-0.25 (2.92)	0.00	-0.09	0.631
Month 20					
First quartile of worsening	31	-4.32 (2.75)	-5.00	-1.57	<0.001
Second quartile of worsening	31	-2.74 (3.42)	-3.00	-0.80	<0.001
Third quartile of worsening	31	-2.19 (3.13)	-2.00	-0.70	<0.001
Fourth quartile of worsening	31	-0.77 (3.14)	-1.00	-0.25	0.180
Change Level MMT					
Month 12					
First quartile of worsening	31	-2.55 (3.36)	-3.00	-0.76	<0.001
Second quartile of worsening	32	-0.69 (2.86)	-1.00	-0.24	0.183
Third quartile of worsening	32	-2.72 (3.42)	-2.00	-0.80	<0.001
Fourth quartile of worsening	32	-0.38 (3.16)	0.00	-0.12	0.507
Month 20					
First quartile of worsening	31	-4.00 (2.79)	-4.00	-1.43	<0.001
Second quartile of worsening	30	-2.13 (4.52)	-1.00	-0.47	0.015
Third quartile of worsening	32	-2.22 (3.34)	-1.00	-0.67	<0.001
Fourth quartile of worsening	32	-1.84 (2.76)	-1.00	-0.67	<0.001

The SRM is calculated by dividing the mean change score between baseline and the subsequent time point by the SD of the change score. The p-value for each individual change group is derived from a paired (within samples) t-test assessing the difference over time. When analysing the mTUG, the reciprocal value of the measured time multiplied by the planned total distance of 6 meters was used; this corresponds to analysing the velocity of the walking speed expressed in meters per seconds (m/s), including the time spent for standing up and sitting down again, and allows the adoption of the value "zero" for patients unable to perform the assessment. HAQ-DI, Health Assessment Questionnaire Disability Index; IBMFRS, Inclusion Body Myositis Functional Rating Scale; MMT, Manual Muscle Testing; mTUG, modified Timed Up and Go; PGIS, Patient Global Impression of Severity; SF36-PCS, Short Form-36 Health Survey physical component score.

Again, with respect to categories of HAQ-DI change, the greater the increase, or worsening, in HAQ-DI score the greater the reduction, or worsening, in IBMFRS score. At month 20, the mean change scores reduced from a mean increase of 0.15 in the improved group, through –1.23 in the stable group, and –3.11 (p<0.001), -4.89 (p<0.001), and –7.50 (p<0.001) in groups with increasing HAQ-DI deterioration. A large SRM was observed for all 3 worsened groups (-0.90, -1.44, and –7.50) at month 20.

For the SF36-PCS; groups were stratified according to quartiles of worsening. The mean change significantly decreased in all quartiles, with the greatest mean reduction being observed in the first quartile which had the greatest degree of worsening in the PCS: -2.66 at 12 months, p<0.001, SRM=-0.855; and -4.09 at 20 months, p<0.001, SRM=-1.12.

The mTUG was also used to investigate responsiveness by stratifying patients according to quartiles of mTUG change. At both month 12 and month 20, the greater the worsening in the mTUG the greater the worsening in the IBMFRS, with the greatest IBMFRS drop observed in the first mTUG quartile with the greatest mTUG reduction: -2.87 at month 12, p<0.001, SRM=-1.278; -4.32 at month 20, p<0.001, SRM=-1.572.

The MMT PerfO was also used to stratify patients into change quartiles. In general, the greater the degree of worsening on the MMT, the greater the degree of worsening on the IBMFRS, but with the relationship being stronger at month 20, with the greatest IBMFRS drop observed in the first MMT quartile with the greatest MMT reduction: -4.00, p<0.001, SRM=-1.432.

Meaningful change threshold

When comparing PGIS and PGIC anchored dichotomous scores of worsening versus no change or improvement at 20 months (**Table 6**), the corresponding best cut point was a drop in 2 IBMFRS points, for both PRO anchors and for both threshold criteria (closest to (0,1) point, and Youden's index). Results at 12 months (**Table 6**) were similar for the PGIS anchor, while a drop in 1 IBMFRS point performed better for the PGIC anchor. Therefore, taking all the results into account, a drop in two IBMFRS points was the most consistent best cut point, and was thus taken to indicate a meaningful decline.

Table 6. ROC analyses to determine the optimal threshold for meaningful decline in the Inclusion Body Myositis Functional Rating Scale total score

PRO anchor	Threshold criterion	Best cut point for IBMFRS change at month 12	Best cut point for IBMFRS change at month 20
PGIS	Closest to (0,1) point	-2.0	-2.0
	Youden index	-2.0	-2.0
PGIC	Closest to (0,1) point	-1.0	-2.0
	Youden index	-1.0	-2.0

PRO, patient-reported outcome. PGIC, Patient Global Impression of Change; PGIS, Patient Global Impression of Severity.

DISCUSSION

This study evaluated the measurement properties of the IBMFRS in a cohort of 150 IBM patients who participated in a large IBM clinical trial.[8] It demonstrated the validity, reliability, and responsiveness of the IBMFRS in IBM. Equivalence between telephone and face-to-face administration was established, and a decrease of at least 2 points in the IBMFRS total score represented a meaningful change. Overall, the IBMFRS performed well in this study, with the high level of standardization in its administration being one of the contributing factors, which is critical when measures are being used in research studies such as clinical trials.

There is a growing need to find specific COAs to assess IBM patients both in clinical practice and research. The IBMFRS is a relatively simple and quick assessment to perform that only contains 10 items. Limited evidence[20, 32] has supported and contributed to the acceptance by regulatory authorities of the IBMFRS as the primary outcome measure in recent[8] and ongoing (NCT04789070, NCT05721573) efficacy clinical trials in IBM. Furthermore, the IBMFRS has been important in determining whether other potential COAs or biomarkers, for example quantitative MRI, are valuable in IBM.[33] Although we have previously assessed the IBMFRS using a Rasch based approach,[34] and showed content valid and reliability in a smaller IBM study,[13] there has been a pressing need for more detailed and robust psychometric evaluation of the IBMFRS scale.

The IBMFRS performed well compared with the other health domains used to assess construct validity in this study. IBMFRS scores correlated highly with PerfOs such as MMT scores, mTUG, and 6MWT, as expected, although hand grip strength achieved a weaker but still moderate correlation. This is likely to be the result of the other PerfOs including assessment of lower body strength (6MWT), both upper and lower body strength (mTUG), or the strength of multiple muscles (MMT), rather than just grip in isolation. HAQ-DI, SF36-PF and SF36-PCS achieved strong convergent relationships with the IBMFRS.

This study demonstrated that the IBMFRS has adequate internal consistency, with the overall score, and the score after exclusion of each of the 10 items, achieving a Cronbach’s alpha coefficient ≥0.75. The IBMFRS swallowing item was associated with the largest increase in alpha following its exclusion (0.81), suggesting that it is measuring a slightly different concept than the other IBMFRS items. It is generally accepted that at present there is a lack of reliable tools to assess dysphagia (difficulty or discomfort in swallowing) and bulbar dysfunction in IBM.[35-37]

When assessing intra-rater reliability, we demonstrated ICCs ranging from 0.84 to 0.87, while regarding equivalence of telephone versus F2F in-clinic administration the ICCs ranged from 0.93 to 0.95. These results reflect excellent intra-rater reliability and equivalence between remote telephone vs F2F administration of the IBMFRS. While our research group had recently demonstrated a similar observation, the study population in this previous report was considerably smaller (n=9).[13] Demonstrating equivalence between telephone and F2F administration is pertinent, particularly amidst the transition towards remote and telemedicine worldwide, largely as a consequence of the COVID-19 pandemic. Roy et al.[38] recently introduced the IBM personalized index calculator (IBM-PIC), a modified IBMFRS scale enabling online patient responses, with high equivalence to telephone-obtained IBMFRS scores (ICC=0.98), despite a small study size (n=35).

Overall, the IBMFRS tool demonstrated excellent responsiveness. For the severest groups (i.e., markedly worsened or first quartile of worsening) stratified according to all COAs tested, high IBMFRS score SRMs (>1.1) were achieved at 20 months. The higher SRMs, greater statistical significance, and stronger monotonic trends found at 20 vs 12 months reflects the greater worsening in IBMFRS observed at this time. In terms of longitudinal relationships, we found moderate and strong relationships between IBMFRS change score and mTUG and HAQ-DI change scores, respectively. The weak to moderate relationships with the other reference measures is likely to reflect the fact that these measures are generic in nature and thus not sufficiently aligned with the specific constructs measured by the IBMFRS. As also observed in other studies, we found a weak correlation between a change in IBMFRS and a change in MMT.[20]

ROC analysis anchored to PGIC and PGIS identified a two-point drop in the IBMFRS total score as indicative of meaningful decline. This finding has practical implications for monitoring disease progression in IBM patients clinically and selecting individuals for intensified surveillance. This cut-off also informs the design of future drug trials, particularly in defining target endpoints and outcomes based on a dichotomous IBMFRS-based variable to distinguish responders from non-responders.

This study has limitations. Patients were only recruited from the UK and US, hence studying the use of the IBMFRS across other countries internationally is needed. In addition, the great majority of the patients included were male and white, limiting the representativeness of the sample; however, it is known that IBM is more common among males (with an approximately 2:1 male-to-female ratio) and white people, and therefore the study population reflects the expected demographics of the disease in the UK and USA.[16, 20, 39, 40] The mean IBMFRS total score of the included patients at baseline indicated overall moderate disability. While psychometric analysis is typically not performed in separate severity groups, IBMFRS scores decreased progressively the greater the severity as indicated by the PGIS and HAQ -DI, suggesting that IBMFRS scores are able to measure disability across the spectrum. Finally, our investigations did not determine how IBMFRS total scores could be used to stratify disease severity and allow division of patients into groups such as for example mild, moderate, and severe.

In conclusion this study lends support to the use of IBMFRS scale as valid, reliable, and responsive tool in monitoring disease progression in IBM when administered by trained raters. Evidence has been provided to propose a drop in at least 2 points in the IBMFRS total score to indicate a meaningful decline in disease status.

DECLARATIONS

Members of the Arimoclomol in IBM Investigator Team of the Neuromuscular Study Group

Anthony A. Amato MD (Department of Neurology, Brigham & Women's Hospital, Boston, MA, USA); Emma Ciafaloni MD (Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA); Miriam Freimer MD (Department of Neurology, The Ohio State Wexner Medical Center, Columbus, OH, USA; Summer B. Gibson MD (Neuromuscular Division, University of Utah School of Medicine, Salt Lake City, UT, USA); Sarah M. Jones MD (Department of Neurology, University of Virginia, Charlottesville, VA, USA); Todd D. Levine MD (Department of Neurology, HonorHealth, Phoenix, AZ, USA); Thomas E. Lloyd MD (Departments of Neurology and Neuroscience, Johns Hopkins University, Baltimore, MD, USA); Michael P. McDermott PhD (Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA); Tahseen Mozaffar MD (Division of Neuromuscular Disorders, University of California, Irvine, Orange, CA, USA); Aziz I. Shaibani (Nerve & Muscle Center of Texas, Baylor College of Medicine, Houston, TX, USA); Matthew Wicklund (Department of Neurology, University of Texas Health Science Center at San Antonio, TX, USA).

Funding

The arimoclomol trial was co-funded by a 4-year FDA Office of Orphan Products Development grant (grant/award no. R01FD004809) and Orphazyme A/S (grant/award no. N/A).

Conflicts of interest

S.S. is supported by a UCL Queen Square Institute of Neurology & Cleveland Clinic London MPhil/PhD Neuroscience Fellowship. T.S., H.D., S.R., J.R., L.L-P and S.H. are employees of Clinical Outcomes Solutions, a health research consultancy that was paid to conduct the measurement properties analyses reported in this study. C.G. was an employee of Orphazyme A/S, the pharmaceutical company that funded the measurement properties analyses reported in this study. L.H. has no disclosures to report. R.J.B. has received funding from the FDA Office Orphan Products Development grant for his role in the arimoclomol trial. M.G.H. receives research funding from the Medical Research Council UK and has previously acted as a consultant for Novartis and for Orphazyme A/S. MMD serves or recently served as a consultant for Abata/Third Rock, Abcuro, Amicus, ArgenX, Astellas, Cabaletta Bio, Catalyst, CNSA, Covance/Labcorp, CSL-Behring, Dianthus, Horizon, EMD Serono/Merck, Ig Society, Inc, Janssen, Medlink, Octapharma, Priovant, Sanofi Genzyme, Shire Takeda, TACT/Treat NMD, UCB Biopharma, Valenza Bio and Wolters Kluwer Health/UpToDate; MMD also received research grants or contracts or educational grants from Alexion/ AstraZeneca, Alnylam Pharmaceuticals, Amicus, Argenx, Bristol-Myers Squibb, Catalyst, CSL-Behring, FDA/OOPD, GlaxoSmithKline, Genentech, Grifols, Mitsubishi Tanabe Pharma, MDA, NIH, Novartis, Octapharma, Orphazyme, Ra Pharma/UCB, Sanofi Genzyme, Sarepta Therapeutics, Shire Takeda, Spark Therapeutics, The Myositis Association, and UCB Biopharma / RaPharma. P.M.M. has received honoraria from Abbvie, BMS, Celgene, Eli Lilly, Galapagos, Janssen, MSD, Novartis, Orphazyme Pfizer, Roche and UCB.

Data sharing

Data sharing requests can be submitted after 1 year following publication of the main study results, to the corresponding authors, who will provide a data access request form. Data sharing requests will be considered by the Trial Steering Committee on a case-by-case basis, and data will be shared if the request is considered reasonable, of scientific interest, and legally and ethically possible.

Contributors

The first draft of the manuscript was written by SS and PMM. All authors critically reviewed and commented on each draft of the manuscript. All authors approved the final manuscript for submission.

Ethics approval

The Arimoclomol trial study protocol was approved by the relevant Institutional Review Board (IRB)/Research Ethics Committee (REC), utilizing a single IRB review via the SMART IRB platform for the 11 US centres (University of Kansas Medical Center Human Research Protection Program, reference number STUDY00002461), and the Health Research Authority (HRA) approval process for the UK centre (London - Surrey Borders Research Ethics Committee, reference number: 18/LO/0696). The trial is registered with ClinicalTrials.gov, number NCT02753530, and is completed. The trial was conducted in accordance with the Declaration of Helsinki (October 2013) and its revisions as well as with the valid national laws of the participating countries and the Integrated Addendum to International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use ICH E6(R1): Guideline for Good Clinical Practice (GCP) E6 (R2) effective 14 June 2017, European Regulation No. 536/2014, and with the Commission Directives 1991/507/EEC and 2001/83/EC.

Acknowledgements

We wish to thank the patients with IBM who participated in this study; the Neuromuscular Muscle Study Group Executive Committee for their assistance in reviewing the arimoclomol trial study design; and all investigators, co-investigators, study coordinators and other staff involved in the arimoclomol trial. PMM and MGH are supported by the National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

REFERENCES

1. Machado PM, Ahmed M, Brady S, Gang Q, Healy E, Morrow JM, et al. Ongoing developments in sporadic inclusion body myositis. Curr Rheumatol Rep. 2014 Dec; 16(12):477.

2. Greenberg SA. Inclusion body myositis: clinical features and pathogenesis. Nat Rev Rheumatol. 2019 May; 15(5):257-272.

3. McLeish E, Slater N, Sooda A, Wilson A, Coudert JD, Lloyd TE, et al. Inclusion body myositis: The interplay between ageing, muscle degeneration and autoimmunity. Best Pract Res Clin Rheumatol. 2022 Jun; 36(2):101761.

4. Price MA, Barghout V, Benveniste O, Christopher-Stine L, Corbett A, de Visser M, et al. Mortality and Causes of Death in Patients with Sporadic Inclusion Body Myositis: Survey Study Based on the Clinical Experience of Specialists in Australia, Europe and the USA. J Neuromuscul Dis. 2016 Mar 3; 3(1):67-75.

5. Ahmed M, Machado PM, Miller A, Spicer C, Herbelin L, He J, et al. Targeting protein homeostasis in sporadic inclusion body myositis. Sci Transl Med. 2016 Mar 23; 8(331):331ra341.

6. Benveniste O, Hogrel J-Y, Belin L, Annoussamy M, Bachasson D, Rigolet A, et al. Sirolimus for treatment of patients with inclusion body myositis: a randomised, double-blind, placebo-controlled, proof-of-concept, phase 2b trial. Lancet Rheumatol. 2021; 3:e40-48.

7. Hanna MG, Badrising UA, Benveniste O, Lloyd TE, Needham M, Chinoy H, et al. Safety and efficacy of intravenous bimagrumab in inclusion body myositis (RESILIENT): a randomised, double-blind, placebo-controlled phase 2b trial. Lancet Neurol. 2019 Sep; 18(9):834–844.

8. Machado PM, McDermott MP, Blaettler T, Sundgreen C, Amato AA, Ciafaloni E, et al. Safety and efficacy of arimoclomol for inclusion body myositis: a multicentre, randomised, double-blind, placebo-controlled trial. Lancet Neurol. 2023 Oct; 22(10):900-911.

9. Alfano LN, Focht Garand KL, Malandraki GA, Salam S, Machado PM, Dimachkie MM. Measuring change in inclusion body myositis: clinical assessments versus imaging. Clin Exp Rheumatol. 2022 Feb; 40(2):404-413.

10. Laurent D, Riek J, Sinclair CDJ, Houston P, Roubenoff R, Papanicolaou DA, et al. Longitudinal Changes in MRI Muscle Morphometry and Composition in People With Inclusion Body Myositis. Neurology. 2022 Aug 30; 99(9):e865-e876.

11. Roy B, Lucchini M, Lilleker JB, Goyal NA, Naddaf E, Adler B, et al. Current status of clinical outcome measures in inclusion body myositis: a systematised review. Clin Exp Rheumatol. 2023 Mar; 41(2):370-378.

12. Rider LG, Aggarwal R, Machado PM, Hogrel J-Y, Reed AM, Christopher-Stine L, et al. Update on outcome assessment in myositis. Nat Rev Rheumatol. 2018 May; 14(5):303–318.

13. Symonds T, Randall J, Lloyd-Price L, Hudgens S, Dimachkie MM, Guldberg C, et al. Study to Assess Content Validity and Interrater and Intrarater Reliability of the Inclusion Body Myositis Functional Rating Scale. Neurol Clin Pract. 2023 Aug; 13(4):e200168.

14. Jackson CE, Barohn RJ, Gronseth G, Pandya S, Herbelin L, Muscle Study Group. Inclusion body myositis functional rating scale: a reliable and valid measure of disease severity. Muscle Nerve. 2008 Apr; 37(4):473–476.

15. Cedarbaum JM, Stambler N. Performance of the Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS) in multicenter clinical trials. J Neurol Sci. 1997 Oct; 152 Suppl 1:S1-9.

16. Cortese A, Machado P, Morrow J, Dewar L, Hiscock A, Miller A, et al. Longitudinal observational study of sporadic inclusion body myositis: implications for clinical trials. Neuromuscul Disord. 2013 May; 23(5):404-412.

17. Nagy S, Khan A, Machado PM, Houlden H. Inclusion body myositis: from genetics to clinical trials. J Neurol. 2023 Mar; 270(3):1787-1797.

18. Rose MR, ENMC IBM Working Group. 188th ENMC International Workshop: Inclusion Body Myositis, 2–4 December 2011, Naarden, The Netherlands. Neuromuscul Disord. 2013 Dec; 23(12):1044–1055.

19. Dimachkie MM, Hanna MG, Machado PM, Herbelin L, Pasnoor M, Jawdat O, et al. Phase II Study of Arimoclomol in IBM FDAOOPD (Orphan Products Division) R01. RRNMF Neuromuscular Journal. 2021; 2:3:104-142.

20. Sangha G, Yao B, Lunn D, Skorupinska I, Germain L, Kozyra D, et al. Longitudinal observational study investigating outcome measures for clinical trials in inclusion body myositis. J Neurol Neurosurg Psychiatry. 2021 Apr 13.

21. Ware J, K Snow, M Kosinski, and B Gandek. SF-36 Health Survey Manual and Interpretation Guide. In: Institute BTH, ed. New England Medical Center
Hospitals 1993.

22. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes. 2003 Jun 9; 1:20.

23. Lowes LP, Alfano L, Viollet L, Rosales XQ, Sahenk Z, Kaspar BK, et al. Knee extensor strength exhibits potential to predict function in sporadic inclusion-body myositis. Muscle Nerve. 2012 Feb; 45(2):163-168.

24. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul; 63(7):737-745.

25. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol. 2020 Dec 3; 20(1):293.

26. Cohen. Statistical power analysis for the behaviors science. 2nd Edition ed. New Jersey: Laurence Erlbaum Associates: Hillsdale, 1988.

27. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes. 2006 Aug 22; 4:54.

28. Turner D, Schunemann HJ, Griffith LE, Beaton DE, Griffiths AM, Critch JN, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. J Clin Epidemiol. 2009 Apr; 62(4):374-379.

29. Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. Biometrics. 1997 Sep; 53(3):823-837.

30. Youden WJ. Index for rating diagnostic tests. Cancer. 1950 Jan; 3(1):32-35.

31. Machado PM. Measurements, composite scores and the art of 'cutting-off'. Ann Rheum Dis. 2016 May; 75(5):787-790.

32. Goyal NA, Greenberg SA, Cauchi J, Araujo N, Li V, Wencel M, et al. Correlations of disease severity outcome measures in inclusion body myositis. Neuromuscul Disord. 2022 Oct; 32(10):800-805.

33. Morrow JM, Sinclair CD, Fischmann A, Machado PM, Reilly MM, Yousry TA, et al. MRI biomarker assessment of neuromuscular disease progression: a prospective observational cohort study. Lancet Neurol. 2016 Jan; 15(1):65-77.

34. Ramdharry G, Morrow J, Hudgens S, Skorupinska I, Gwathmey K, Currence M, et al. Investigation of the psychometric properties of the inclusion body myositis functional rating scale with rasch analysis. Muscle Nerve. 2019 Aug; 60(2):161-168.

35. Ambrocio KR, Garand KLF, Roy B, Bhutada AM, Malandraki GA. Diagnosing and managing dysphagia in inclusion body myositis: a systematic review. Rheumatology (Oxford). 2023 Apr 28.

36. Focht Garand KL, Malandraki GA, Stipancic KL, Kearney E, Roy B, Alfano LN. Paucity of bulbar function measures in inclusion body myositis trials. Reply to: Current status of clinical outcome measures in inclusion body myositis: a systematised review. Clin Exp Rheumatol. 2023 Mar; 41(2):399.

37. Garand KLF, Malandraki GA, Dimachkie MM. Update on the evaluation and management of dysphagia in sporadic inclusion body myositis. Curr Opin Otolaryngol Head Neck Surg. 2023 Sep 4.

38. Roy B, Zubair A, Petschke K, O'Connor KC, Paltiel AD, Nowak RJ. Reliability of patient self-reports to clinician-assigned functional scores of inclusion body myositis. J Neurol Sci. 2022 May 15; 436:120228.

39. Paltiel AD, Ingvarsson E, Lee DK, Leff RL, Nowak RJ, Petschke KD, et al. Demographic and clinical features of inclusion body myositis in North America. Muscle Nerve. 2015 Oct; 52(4):527-533.

40. Dimachkie MM, Barohn RJ. Inclusion body myositis. Neurol Clin. 2014 Aug; 32(3):629-646, vii.

FIGURES

Figure 1. Test-Retest reliability of the Inclusion Body Myositis Functional Rating Scale (IBMFRS). Bland-Altman Plot showing degree of agreement of the IBMFRS from (A) baseline to Month 4 (N=78), (B) month 4 to month 8 (N=77), and (C) month 8 to month 12 (N=78).

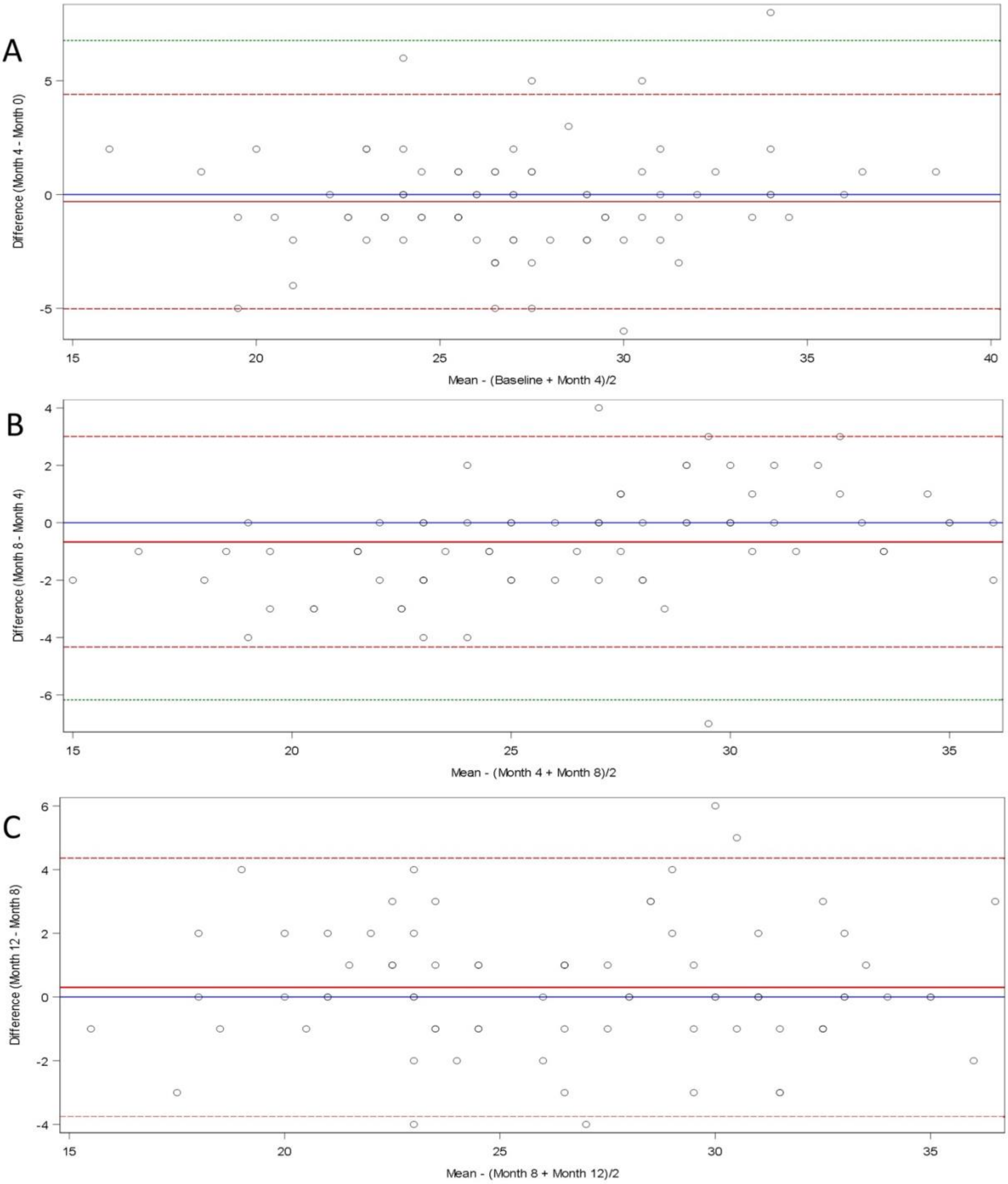
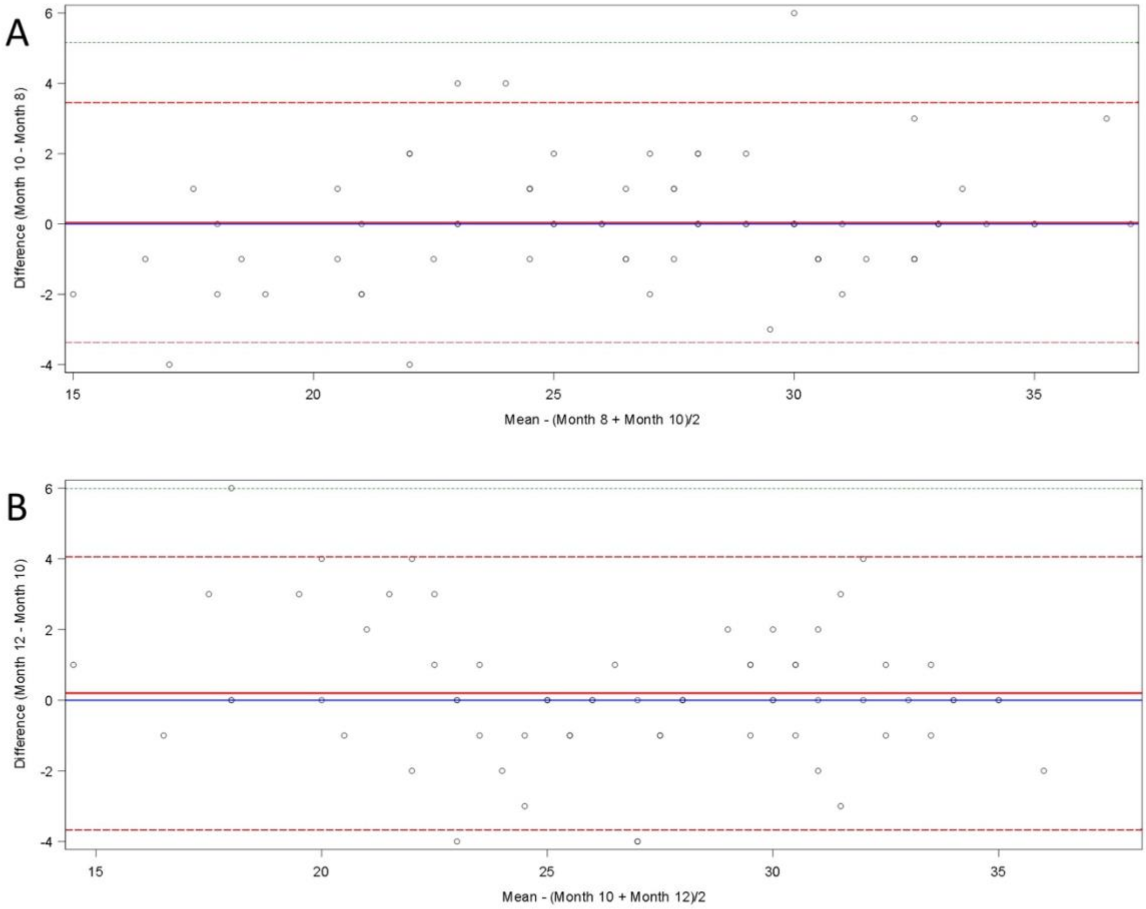


Figure 2. Equivalence of Inclusion Body Myositis Functional Rating Scale (IBMFRS) scoring in clinic versus via over the phone. Bland-Altman Plot showing degree of agreement of the IBMFRS from (A) month 8 (in-clinic) to Month 10 (phone) (N=79), and from month 10 (phone) to month 12 (in-clinic) (N=79).


A
Difference (Month 10 - Month 8)
Mean - (Month 8 + Month 10)/2
B
Difference (Month 12 - Month 10)
Mean - (Month 10 + Month 12)/2