Seminars in Arthritis and Rheumatism 64 (2024) 152339

Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Responsiveness and meaningful thresholds of PROMIS pain interference, fatigue, and physical function forms in adults with idiopathic inflammatory myopathies: Report from the OMERACT Myositis Working Group

D Saygin [a,1], D DiRenzo [b,1], J Raaphorst [c], I de Groot [d], CO Bingham [e], IE Lundberg [f], M Regardt [g], C Sarver [h], M de Visser [c], LJ Maxwell [i], D Beaton [j], JY Kim [k], M Needham [l], H Alexanderson [m], L Christopher-Stine [e], CA Mecoli [e,2], JK Park [k,2,*], on behalf of OMERACT Myositis Working Group

[a] Division of Rheumatology, Department of Medicine, University of Pittsburgh, Pittsburgh, USA
[b] Division of Rheumatology, Department of Medicine, University of Pennsylvania, Philadelphia, USA
[c] Department of Neurology, Amsterdam UMC, University of Amsterdam, Amsterdam Neuroscience, Amsterdam, the Netherlands
[d] Patient Research Partner, Netherlands
[e] Division of Rheumatology, Department of Medicine, Johns Hopkins University, Baltimore, USA
[f] Division of Rheumatology, Department of Medicine, Solna, Karolinska Institutet, Karolinska University, Stockholm, Sweden
[g] Department of Neurobiology, Care Sciences and Society, Division of Occupational Therapy, Karolinska Institutet and Medical Unit Occupational Therapy and Physical Therapy, Karolinska University Hospital, Stockholm, Sweden
[h] Patient Research Partner, USA
[i] Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[j] Institute for Work & Health and Institute for Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada
[k] Division of Rheumatology, Department of Internal Medicine, Seoul National University Hospital and College of Medicine, Seoul, Korea
[l] Department of Neurology, Fiona Stanley Hospital, IIID Murdoch University and University of Notre Dame, Perth, Australia
[m] Medical Unit Occupational Therapy and Physical Therapy, Karolinska University Hospital and Division of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Key words:
Myositis
Myopathy
PROMIS
OMERACT
Outcome measures
Psychometrics

ABSTRACT

Background: A series of qualitative studies conducted by the OMERACT Myositis Working Group identified pain interference, fatigue, and physical function as highly important life impact domains for adults with idiopathic inflammatory myositis (IIM). In this study, our goal was to assess the responsiveness and minimal important difference of PROMIS pain interference (6a), fatigue (7a), and physical function (8b).

Methods: Adults with IIM from USA, Netherlands, Korea, Sweden, and Australia with two "clinical" visits were enrolled in this prospective study. Anchor questions on a Likert scale were collected at baseline, and manual muscle testing (MMT), physician and patient reported global disease activity, and PROMIS instruments were collected at both visits. Responsiveness was assessed with i) ANOVA, ii) paired *t*-test, effect size and standardized response mean, and iii) Pearson correlation. Minimal important difference (MID), minimal important change (MIC) and minimal detectable change (MDC) values were calculated.

Results: 114 patients with IIM (median age 60, 60 % female) completed both visits. Changes in PROMIS instruments were significantly different among anchor categories. Patients who reported improvement had a significant improvement in their PROMIS scores with at least medium effect size, while patients who reported worsening and stability did not show a significant change with weak effect size. PROMIS instruments had weak to moderate correlations with MMT, patient and physician global disease activity. MID was approximately 2–3 points for Pain Interference and 3–4 points for Fatigue and Physical Function

* Corresponding author at: Division of Rheumatology, Department of Internal Medicine, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080.
E-mail address: jinkyunpark@snu.ac.kr (J. Park).
[1] Co-first authors.
[2] Co-senior authors.

https://doi.org/10.1016/j.semarthrit.2023.152339

Available online 9 December 2023

forms based on the method used. MIC was approximately 4–5 for improvement of all the instruments, while MDC was 1.7–2 points for Pain Interference and Physical Function and 3.2–3.9 for Fatigue.
Conclusion: This study provides evidence towards the responsiveness of the PROMIS instruments in a large international prospective cohort of adults with IIM supporting their use as PROMs in adult myositis.

Introduction

The idiopathic inflammatory myopathies (IIM) are multisystem, autoimmune diseases with many different clinical presentations and subtypes and predominantly affecting muscles [1]. Subtypes include dermatomyositis, polymyositis, immune-mediated necrotizing myopathy, inclusion body myositis, overlap myositis, and antisynthetase syndrome [2,3]. Muscle weakness and impaired physical function are usually a hallmark feature of IIM, but extramuscular manifestations can be present including skin rash, interstitial lung disease, cardiac dysfunction, and arthritis [4].

Measuring patient-reported response to life impacts in clinical trials has been challenging as there are no fully validated patient reported outcome measures (PROs) that exist for adult patients with IIM. The Outcome Measures in Rheumatology (OMERACT) Myositis Working Group has been evaluating the validity of candidate life-impact instruments to suit the needs of people living with, caring for, and researching IIM for nearly a decade [5]. To date, the Health Assessment Questionnaire-Disability Index (HAQ-DI) has been substantially used in IIM clinical trials and cohorts to measure physical function but does have limitations [6], including ceiling and floor effects, lack of content validity, and limited responsiveness. These limitations prompted further exploration by the OMERACT Myositis Working Group to find suitable instruments to measure physical function among other domains [7].

Through exhaustive literature reviews [8], international patient focus groups [9], modified Delphi surveys [10], and in-person workshops, the core domains of pain interference, fatigue, and physical function were identified by the OMERACT Myositis Working Group. The content validity of candidate instruments was then evaluated through cognitive interviewing [11]. The Patient Reported Outcomes Measurement Information System (PROMIS) pain interference 6a, physical function 8b, and fatigue 7a instruments emerged as candidate instruments through this process. Further validation revealed excellent test-retest reliability and construct validity [12].

The purpose of this study was to determine the responsiveness and minimal important difference (MID) for the proposed PROMIS candidate instruments (pain interference 6a, physical function 8b, and fatigue 7a) in adults with IIM.

Methods

Study design and participants

The study was designed as a multisite prospective observational study with two visits [13,14]. Participants were included who fulfilled either ACR/EULAR 2017 Classification or the Bohan and Peter criteria for idiopathic inflammatory myositis [15–17]. Patients with IBM were excluded. When applicable, patients were further categorized into the IIM subtypes of anti-synthetase syndrome, overlap myositis, or immune-mediated necrotizing myopathy (IMNM) based on criteria followed at each individual site [3]. Patients were enrolled from the outpatient clinics in Australia (Perron Institute), Netherlands (Amsterdam University Medical Center), South Korea (Seoul National University Hospital), Sweden (Karolinska University Hospital), and USA (Johns Hopkins Myositis Center).

Outcome measures

Manual muscle testing (MMT8, scored 0–80), physician reported global disease activity (0–10), patient reported global disease activity (0–10), serum creatine kinase (CK) within 6-weeks of the study visit, and the PROMIS instruments for pain interference (v1.1, 6a), fatigue (v1.0, 7a), and physical function (v2.0, 8b) were collected at the baseline and follow-up visits. A subgroup of patients also had the HAQ-DI collected. CK fold change was calculated by dividing the CK value by the upper limit of normal range. PROMIS instruments had a recall period of 7 days. T-scores were calculated using the Health Measures Scoring Service (https://www.assessmentcenter.net/ac_scoringservice). The range of t-scores are 41.0–78.3 for Pain Interference with higher scores indicating more pain interference, 29.4–83.2 for Fatigue with higher scores indicating higher levels of fatigue, and 20.3–60.1 for Physical Function with higher scores indicating better physical function. PROMIS instruments that were already translated and adapted for the local cultures in English, Swedish, Korean, and Dutch were used.

At follow-up, participants answered separate anchor questions for each domain of pain, fatigue, and physical function: “Since your last appointment, has your [domain] gotten a lot better, a little better, stayed the same, got a little worse, or got a lot worse?”. The local ethics committees at each participating site approved the research protocols. All participants provided written informed consent for participation and the study procedures were conducted in accordance with the Helsinki Declaration.

Statistical analysis

Continuous variables were reported as median [IQR] while categorical variables were reported as n (%). Only participants with datapoints at both time-points and response to anchor questions were included in the analyses. Myositis working group members were distributed a set of questions to agree on *a priori* hypotheses of responsiveness. All statistical analyses were done using GraphPad Prism software (version 9.0.0, San Diego, CA, USA).

Responsiveness was assessed using three methods. The first method involved comparison of change in PROMIS instrument scores across the anchor groups using ANOVA (F statistic and p value). Our hypothesis for this method was that the PROMIS instrument score change is significant in those who improved and worsened, and not in those who reported unchanged symptoms. The second method was assessment of significance of within-person change over time in PROMIS instruments using paired *t*-test, effect size and standardized response mean (SRM). Effect size was interpreted as small for <0.2, medium for 0.2–0.5 and large >0.5 according to Cohen’s criteria [18]. Our hypothesis for this method was that the PROMIS instrument scores significantly changed in those who improved and worsened with at least medium effect size and did not change significantly in patients who stayed the same. The third method involved examining the correlations between changes in the PROMIS instrument scores and the myositis outcome measures using Pearson correlation. Pearson’s r was interpreted as weak for $r<0.4$, moderate 0.4–0.7, and strong >0.7 [19].

Meaningful thresholds included minimal important difference (MID), minimal important change (MIC) and minimal detectable change (MDC). MID and MIC both refer to the smallest score change which patients perceive as beneficial or worthwhile [20]. However, while MID is the difference between groups of patients with different levels of clinical improvement, MIC is the smallest score change over time in patients with improvement or worsening. MDC refers to the smallest change score in an instrument that falls outside the measurement error [21]. Two methods were utilized to calculate MID: i) anchor-based

method [22] and ii) Receiver Operating Characteristics (ROC) threshold. MIC was calculated as the median difference in PROMIS score in patients who reported "a little better" and "a little worse". MDC was calculated using the distribution-based method.

In the anchor-based method, MID was calculated as the mean change in PROMIS scores that corresponded to a one category shift to achieve a little better or a little worse in the domain-specific anchor question. In order to use anchor-based method, a Spearman's rank order correlation coefficient of $r \geq |0.3|$was necessary between anchor and the PROMIS instrument [23]. The anchor questions were external to the PROMIS instruments and obtained concurrently with the other instruments. MID was calculated separately for improvement and worsening based on studies showing that the perceived thresholds of patients used for improvement and worsening could be different [24–26]. We also generated cumulative distribution function curves to visually examine the change in PROMIS scores by the anchor category.

The Receiver Operating Characteristics (ROC) method takes both the external anchor as well as the variability of the score into consideration for MID [27]. The participants were divided into two groups based on their response to the anchor question: no change and those who had at least a little bit of improvement. ROC curve was created by plotting the sensitivity against the specificity for each change score. Area under the curve (AUC) was calculated for each ROC curve and shows the ability of the instrument to distinguish no change vs improvement. AUC >0.7 is interpreted as adequate [28]. MID was determined based on the change with the highest Youden's index (sensitivity + specificity – 1).

The distribution-based method relies on the variability of the score and aims to determine the change associated with exceeding the measurement error. Mean differences associated with 0.2 SD, 0.5 SD and Standardized Error Mean (SEM) (SD x $\sqrt{(1-\alpha)}$) were calculated for baseline, follow-up, and mean change. Cronbach α values of the PROMIS instruments calculated in our previous work (0.97 for PROMIS Pain Interference, 0.89 for Fatigue, and 0.96 for Physical Function) were used as α in SEM calculation [12].

Results

Characteristics of the study participants

In total, 114 adults with IIM (DM [50 %], PM [10.5 %], IMNM [20.2 %], ASYS [13.2 %] and overlap myositis [6.1 %]) were enrolled from USA (32.5 %), Netherlands (23.7 %), South Korea (21.9 %), Sweden (18.4 %), and Australia (3.5 %) (Table 1). The participants had a median [IQR] age of 60 [50–70] and consisted of 60 % females. Approximately 51.7 % of the patients were in their first year of diagnosis. The median interval between baseline and follow-up visits was 150.5 days. The baseline scores of the myositis outcome measures and PROMIS instruments are shown in Table 1. Most participants reported that their pain interference, fatigue, and physical function stayed the same (37.9–41.7 %) or got better (39.5–47.2 %).

Table 1
Characteristics of the study participants.

		n (%) or median [IQR]
Number of participants		114
USA		37 (32.5 %)
Netherlands		27 (23.7 %)
South Korea		25 (21.9 %)
Sweden		21 (18.4 %)
Australia		4 (3.5 %)
Age		60.0 [50.0–70.0]
Sex (F)		67 (58.8 %)
Disease duration (days)		350.5 [4.5–1604.0]
Interval between visits (days)		150.5 [99.0–199.3]
Diagnosis		
Dermatomyositis		57 (50.0 %)
Polymyositis		12 (10.5 %)
Immune mediated necrotizing myopathy		23 (20.2 %)
Anti-synthetase syndrome		15 (13.2 %)
Overlap myositis		7 (6.1 %)
Myositis outcome measures		
Manual muscle testing (0–80)		74 [63–80]
Physician Global (0–10)		4 [2–6]
Patient Global (0–10)		4.6 [2–6.9]
Creatine kinase fold change*		1.8 [0.5–16.7]
Health Assessment Questionnaire (HAQ-DI)		0.9 [0.4–1.8]
PROMIS instruments		
PROMIS pain interference		55.7 [41.1–66.3]
PROMIS fatigue		57.9 [49.6–66.4]
PROMIS physical function		37.9 [31.8–45.1]
Anchor distribution		
Pain Interference	A lot better	19.4 %
	A little better	23.2 %
	Stayed the same	41.7 %
	A little worse	10.2 %
	A lot worse	5.6 %
Fatigue	A lot better	13.2 %
	A little better	26.3 %
	Stayed the same	41.2 %
	A little worse	13.2 %
	A lot worse	6.1 %
Physical Function	A lot better	17.6 %
	A little better	29.6 %
	Stayed the same	37.9 %
	A little worse	12.0 %
	A lot worse	2.8 %

IQR: Interquartile range, USA: United States of America, F: Female.
* Creatine kinase fold change = Creatine kinase level / Upper limit of normal.

Responsiveness

A list of *a priori* hypotheses was agreed upon within the Myositis Working Group for correlations. In summary, most of the group members (defined as >70 % agreement) agreed on moderate correlations between the PROMIS instruments and the myositis outcome measures except two correlations between fatigue and physician reported global disease activity and MMT for which the agreement was 67 %. Therefore, these two hypotheses were dropped.

I. PROMIS Pain Interference: Change in the PROMIS Pain Interference was significantly different among those who improved, stayed stable, and worsened ($p = 0.003$) (Fig. 1). Change in the PROMIS Pain Interference over time was significant in those who improved ($p<0.0001$) with effect size of −0.64 and SRM of −0.70 and not significant in those who worsened or remained stable with small effect sizes of 0.23 and −0.24 and SRM of 0.23 and −0.24, respectively (Fig. 2, Table 2). PROMIS Pain Interference had weak, statistically significant, correlations with MMT, Patient Global, Physician Global, and CK, and borderline moderate correlation with HAQ-DI (Table 3). Thus, only one of the five *a priori* hypotheses were met for the Pain Interference which was the weak correlation CK.

II. PROMIS Fatigue: Change in the PROMIS Fatigue was significantly different among those who improved, stayed stable, and worsened ($p<0.0001$ for Fatigue) (Fig. 1). Change in the PROMIS Fatigue over time was significant in those who improved ($p<0.0001$) with effect size of −0.83 and SRM of −0.93 and not significant in those who worsened or remained stable with small effect sizes of 0.17 and −0.21 and SRM of 0.28 and −0.25, respectively (Fig. 2, Table 2). PROMIS Fatigue had weak statistically significant correlations with MMT and CK, moderate correlations with patient reported global disease activity and HAQ-DI, and no significant correlation with physician reported global disease activity (Table 3). All three *a priori* hypotheses were met which were the moderate correlation with Patient Global and weak correlations with HAQ-DI and CK.

III. PROMIS Physical Function: Change in the PROMIS Physical Function was significantly different among those who improved, stayed stable, or worsened ($p<0.0001$) (Fig. 1). Change in the PROMIS Physical Function over time was significant in those who improved ($p<0.0001$) with effect size of 0.75 and SRM of 0.70 and not significant in those who

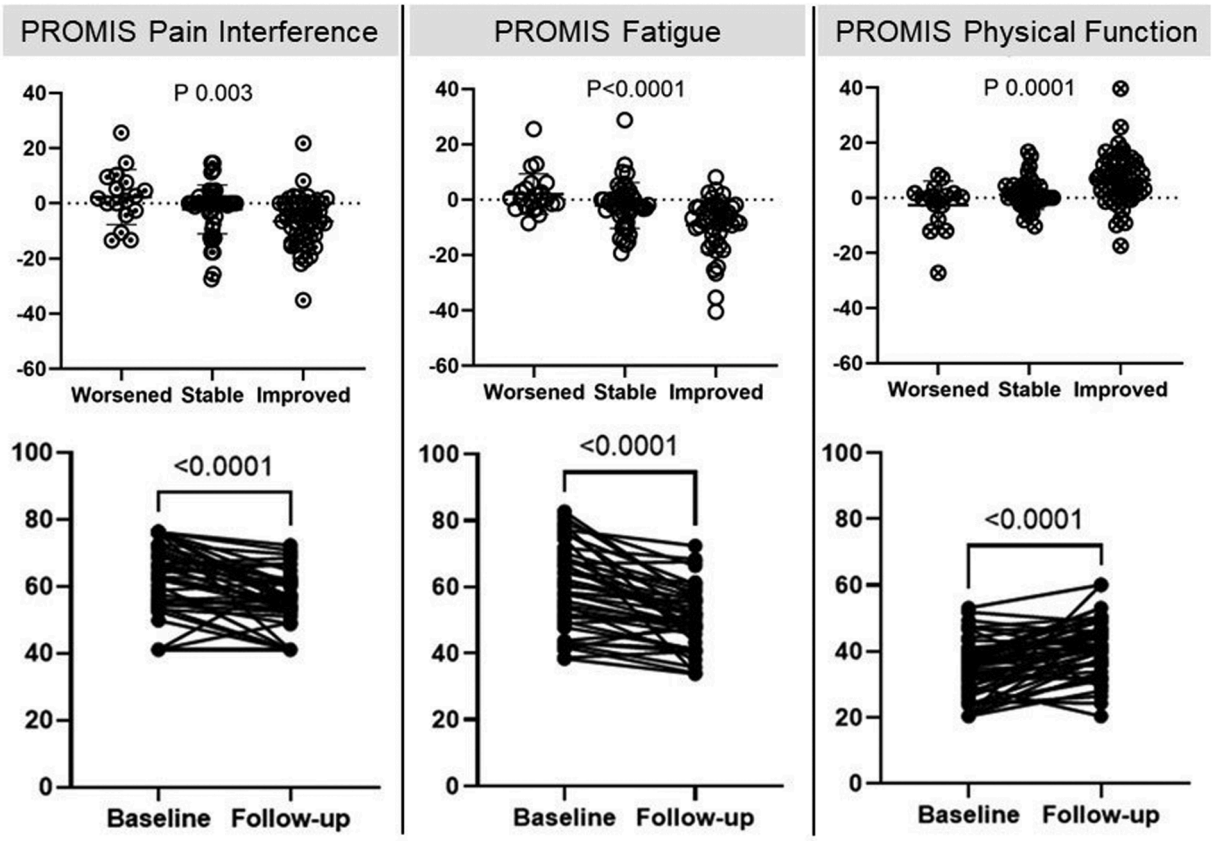

Fig. 1. Change in PROMIS instruments among the anchor categories.

worsened or remained stable with small effect sizes of −0.27 and 0.12 and SRM of −0.29 and 0.23, respectively (Fig. 2, Table 2). PROMIS Physical Function had moderate statistically significant correlations with MMT, Patient Global and HAQ-DI, and weak correlations with Physician Global and CK (Table 3). Three of the five *a priori* hypotheses were met which were moderate correlations with Patient Global and MMT, and weak correlation with CK.

Meaningful thresholds

I. PROMIS Pain Interference. The Spearman rank-order correlation coefficient was −0.37 for PROMIS Pain Interference and the pain anchor question response. The change in t-score for one category shift for Pain Interference was 2.3 for a little worse and 3.2 for a little better group representing the MID values according to the anchor-based method (Table 4, Fig. 3). AUC [CI] for ROC curve was 0.66 [0.55–0.78] (Supplementary Figure 1). Given that the AUC was <0.70 indicating poor discrimination, MID was not calculated using the ROC method. MIC was 5.2 for improvement and 1 for worsening. SEM, 0.2 SD and 0.5 SD were 1.68–2.04, 1.94–2.36 and 4.85–5.89, respectively.

II. PROMIS Fatigue. The Spearman rank-order correlation coefficient was −0.51 for PROMIS Fatigue and the fatigue anchor question response. The change in t-score for one category shift for Fatigue was 3.9 for a little worse and 4.1 for a little better group representing the MID values according to the anchor-based method (Table 4, Fig. 3). AUC [CI] for ROC curve was 0.72 [0.62–0.83] (Supplementary Figure 1). MID was calculated as 2.8 with a sensitivity of 80 % and specificity of 59.6 % based on the ROC method. MIC was 5 for improvement and 0.1 for worsening. SEM, 0.2 SD and 0.5 SD were 3.23–3.91, 1.95–2.36 and 4.87–5.90, respectively.

III. PROMIS Physical Function. The Spearman rank-order correlation coefficient was 0.43 for PROMIS Physical Function and the physical function anchor question response. The change in t-score for one category shift for Physical Function was 3.0 for a little worse and 3.6 for a little better group representing the MID values according to the anchor-based method (Table 4, Fig. 3). AUC [CI] for ROC curve was 0.71 [0.61–0.82] (Supplementary Figure 1). MID was calculated as 4.5 with a sensitivity of 60.8 % and specificity of 80.5 % based on the ROC method. MIC was 4.1 for improvement and 0.3 for worsening. SEM, 0.2 SD and 0.5 SD were 1.73–2.06, 1.73–2.06 and 4.31–5.15, respectively.

Discussion

Building on a decade of work of the OMERACT Myositis Working Group, this study provides evidence towards responsiveness of the PROMIS Pain Interference, Fatigue and Physical Function forms in a longitudinal, international cohort of adults with IIM. We also identified MID, MIC and MDC values for these instruments that correspond to the smallest meaningful difference for patients with IIM. This represents an important step forward in establishing PROs with adequate psychometric properties for use in clinical studies in adult patients with IIM.

All the PROMIS instruments showed statistically significant improvement in those who reported improvement (large effect size) and no significant change (weak-medium effect size) in those who reported staying the same. However, in those with worsening, scores did not change significantly, and effect size remained weak. This observed asymmetric directionality for improvement vs worsening could be due to several possible reasons. The first possible reason is the small number of patients who had worsening in our cohort (<20 %) compared to those who improved. It could also be due to already high scores (meaning worse functioning) observed at baseline in our patients which may leave small room to show further deterioration. Similar results were seen in a previous study [24]. In this study with patients with rheumatoid arthritis assessing the responsiveness of PROMIS physical function (20a,

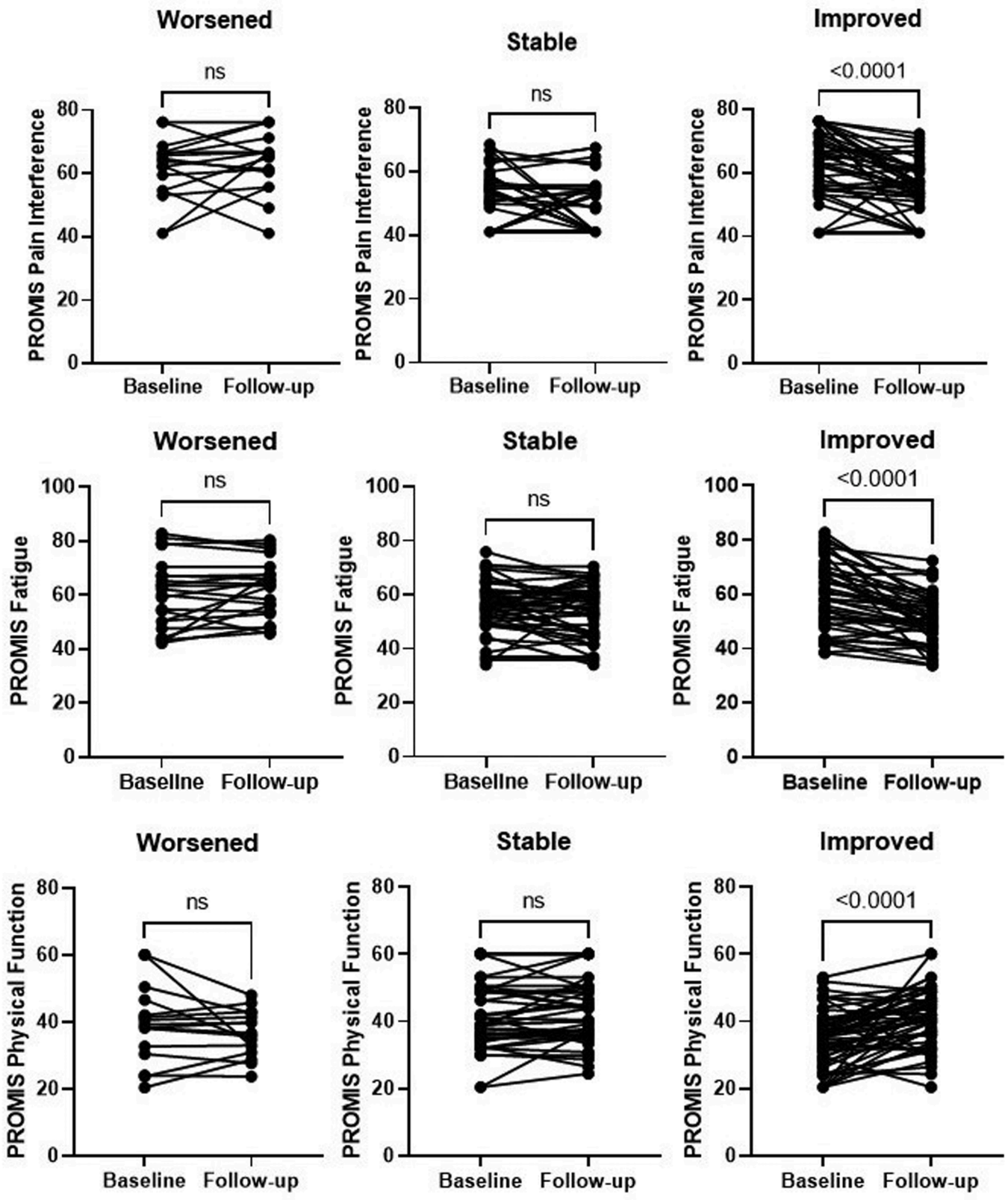

Fig. 2. Change in the PROMIS instruments over time by the domain-specific anchor category.

v1.0), pain interference (8a, v1.0), fatigue (7a, 8a, v1.0), participation (8a, v2.0), and adult profile 29 (v2.0), the patients who reported worsening had 6–7 points change in PROMIS scores while patients who reported improvement had 3–6 points change [24]. These similar results may suggest different thresholds perceived for improvement and worsening by patients. The lower MIDs that were seen for worsening compared to improvement across different domains also support this finding. Even though the ability of a PRO to demonstrate stability and improvement may be more important than its ability to show worsening, particularly in the clinical trial setting where patients with high disease

Table 2
Effect size and standardized response mean of the PROMIS instruments across the anchor categories.

Anchor categories	T-score		Δ in T-score	SD	Effect size	Standardized response mean	n
PROMIS Pain Interference							
Worsened	Baseline	61.4	2.3	9.9	0.23	0.23	17
	Follow-up	63.6					
Stable	Baseline	49.4	−2.2	8.9	−0.24	−0.24	45
	Follow-up	47.2					
Improved	Baseline	61.7	−6.6	9.3	−0.64*	−0.70*	46
	Follow-up	55.1					
PROMIS Fatigue							
Worsened	Baseline	60.3	2.0	7.3	0.17	0.28	22
	Follow-up	62.4					
Stable	Baseline	55.1	−2.1	8.3	−0.21	−0.25	47
	Follow-up	53.0					
Improved	Baseline	59.7	−9.2	9.8	−0.83*	−0.93*	45
	Follow-up	50.5					
PROMIS Physical Function							
Worsened	Baseline	39.1	−2.6	8.7	−0.27	−0.29	16
	Follow-up	36.5					
Stable	Baseline	42.0	1.2	5.4	0.12	0.23	41
	Follow-up	43.2					
Improved	Baseline	34.6	6.5	9.4	0.75*	0.70*	51
	Follow-up	41.1					

SD: Standard deviation.
* Large effect size and standardized response mean (interpreted as small <0.2, medium 0.2–0.5, large >0.5).

activity tend to be enrolled, further studies should be performed in cohorts enriched for patients with worsening to better understand these findings.

The results of this study support adequate responsiveness of the PROMIS instruments based on significance of change among the anchor categories, effect size, and SRM. However, only one of the *a priori* hypotheses for correlations between Pain Interference and the myositis outcome measures were met. The majority of the Myositis Working Group members estimated a moderate correlation between Pain Interference and myositis outcome measures of Physician Global, Patient Global, and Manual Muscle Testing. However, the results demonstrated a weak correlation between these measures. Pain is a relatively understudied symptom in IIM as it was believed to be a painless disease until recently [29]. Therefore, limited knowledge and scarce studies focusing on pain in IIM may potentially explain the discrepancy between *a priori* hypotheses and the results of the observed correlations. Another potential explanation could be the cutoffs that were used to define weak and moderate correlations in this study (<0.4 vs 0.4–0.7). There is a wide variability in thresholds and interpretation of correlation coefficients in the literature [30]. Even though these thresholds of correlation coefficients could still be appropriate for construct validity assessment, a slightly lower threshold may have been better for responsiveness assessment due to differences in variables used for correlation [31]. Unlike construct validity where the instrument scores are correlated with each other, responsiveness assessment requires correlation between longitudinal changes in the instrument scores. This may lead to attenuation of correlations by increasing the chance of introducing random error at two time points [32].

Meaningful thresholds were described to assist in interpretation of the PROs and include a wide variety of concepts including MID, MIC, and MDC. Differentiation of these concepts could be confusing with similar names; however, significant differences exist in their definition. While MID and MIC focus on the patient's perception of the smallest meaningful difference in the domain of interest [20], MDC does not take patient perspective into consideration and simply implies the minimal change that can be detected statistically based on SD and/or SEM of the instrument [33]. MID of an instrument can vary widely depending on the method used with no standard single test; thus, we calculated MID using two different methods to report a range of MID values. ½ SD of t-score (which corresponds to 5) is generally accepted as MID for PROMIS instruments; however, MID can show small differences between different diseases and is encouraged to be assessed separately for each disease group. Anchor-based and ROC-curve method showed similar results for MID with 2–3 points for Pain Interference and 3–4 points for Fatigue and Physical Function forms. MIC was calculated as 4–5 for improvement of these instruments. Similar to our results, a systematic review of MIC values of PROMIS instruments showed a t-score points between 2 and 6 as the most commonly reported MIC values [33].

Strengths of this study include its large size for a rare disease such as IIM, a diverse cohort from several countries across the world, using the rigorous OMERACT methodology, involving patient participation in the design and conduct of the study, and use of external, patient-reported anchors for assessment of responsiveness and meaningful thresholds.

Table 3
A priori and observed correlations between PROMIS instruments and the myositis outcome measures.

PROMIS instruments	Myositis outcome measures	Consensus exercise,% agreement voting members ($n = 12$)	A priori hypotheses with expected correlation	Observed correlation Strength* (r)	Hypothesis met?
Pain Interference	Physician global	75 %	Moderate	Weak (0.25)	No
	Patient global	92 %	Moderate	Weak (0.36)	No
	MMT	83 %	Moderate	Weak (−0.28)	No
	HAQ-DI	83 %	Weak	Moderate (0.40)	No
	Creatine kinase	100 %	Weak	Weak (0.26)	Yes
Fatigue	Physician global	67 %	Dropped	Weak (0.18)	N/A
	Patient global	100 %	Moderate	Moderate (0.42)	Yes
	MMT	67 %	Dropped	Weak (−0.33)	N/A
	HAQ-DI	100 %	Moderate	Moderate (0.55)	Yes
	Creatine kinase	100 %	Weak	Weak (0.35)	Yes
Physical Function	Physician global	92 %	Moderate	Weak (−0.29)	No
	Patient global	100 %	Moderate	Moderate (−0.62)	Yes
	MMT	100 %	Moderate	Moderate (0.56)	Yes
	HAQ-DI	92 %	Strong	Moderate (−0.60)	No
	Creatine kinase	100 %	Weak	Weak (−0.36)	Yes

HAQ-DI: Health Assessment Questionnaire Disability Index. MMT: Manual muscle testing.
* Cutoffs for interpretation were <0.4, 0.4–0.7, and >0.7 for weak, moderate, and strong correlations.

Table 4
Minimal important difference (MID), minimal important change (MIC) and minimal detectable change (MDC) estimates of the PROMIS Pain Interference, Fatigue, and Physical Function forms.

PROMIS Instruments	MDC									Anchor-based MID		ROC curve based MID	MIC	
	Baseline			Follow-up			Change							
	SEM	0.2 SD	0.5 SD	SEM	0.2 SD	0.5 SD	SEM	0.2 SD	0.5 SD	Imp	Wor		Imp	Wor
Pain Interference	2.0	2.4	5.9	1.8	2.1	5.3	1.7	1.9	4.9	3.2	2.3	N/A	5.2	1
Fatigue	3.9	2.4	5.9	3.6	2.2	5.4	3.2	1.9	4.9	4.1	3.9	2.8	5.0	0.1
Physical Function	2.1	2.1	5.2	1.8	1.8	4.6	1.7	1.7	4.3	3.6	3.0	4.5	4.1	0.3

MDC: Minimal detectable change, MID: Minimal important difference, MIC: Minimal Important Difference, SEM: Standardized error mean, SD: Standard deviation, Imp: Improvement, Wor: Worsening.

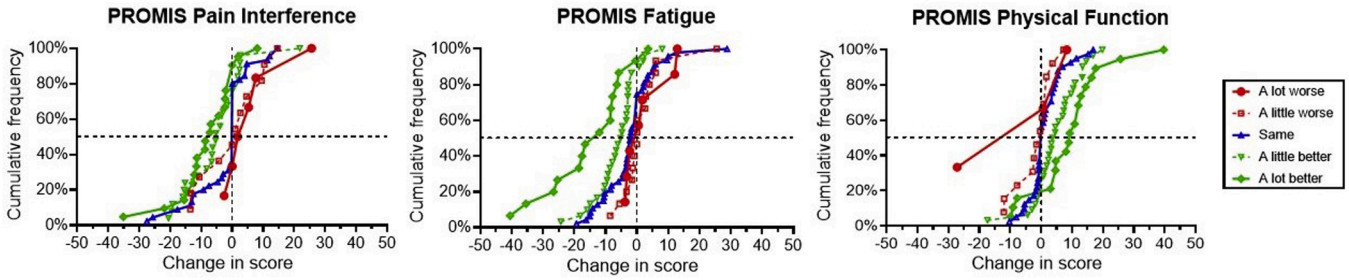

Fig. 3. Cumulative distribution function curves showing change in PROMIS instrument scores by domain-specific change categories. A clear separation between worsening (red) and improving (green) is observed across the score continuum. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

One of the limitations of this study is including a small number of patients who worsened, limiting the conclusions that can be drawn for the ability of the instruments to detect worsening. Further, the study participants had established disease under routine clinical care. Therefore, they likely had a lower disease activity and were less likely to exhibit a clinical change compared to the patients that are recruited into clinical trials. This may make it challenging to assess the responsiveness of the instruments. Our findings should be further tested in clinical trials where patients with active disease are enrolled who are more likely to demonstrate a clinical change due to a higher baseline disease activity and a therapeutic intervention. Lastly, our study sample consisted of prevalent cases with IIM who were slightly older than expected. This may limit the generalizability of our findings to younger individuals with IIM.

Conclusion

This study provides evidence toward the responsiveness of the PROMIS Pain Interference 6a, Fatigue 7a, and Physical Function 8b in a large international prospective cohort of adults with IIM supporting their use as PROs as outcome measures in adult myositis. Studies assessing the performance of these instruments in a clinical trial setting are currently underway [34].

Statement of clinical significance

PROMIS Pain Interference, Fatigue, and Physical Function forms showed adequate test-retest reliability, internal consistency, and construct validity in adults with idiopathic inflammatory myopathy. However, the responsiveness and meaningful thresholds of these instruments were unknown. This study by the OMERACT Myositis Working Group demonstrated evidence towards the responsiveness of these instruments in a large international prospective cohort and established the meaningful thresholds for each instrument.

CRediT authorship contribution statement

D Saygin: Formal analysis. **D DiRenzo:** Formal analysis. **J Raaphorst: . I de Groot:** Conceptualization. **CO Bingham:** Conceptualization, Data curation, Formal analysis. **IE Lundberg:** Conceptualization, Data curation. **M Regardt:** Conceptualization, Data curation. **C Sarver:** Conceptualization. **M de Visser:** Conceptualization, Data curation. **LJ Maxwell:** Conceptualization, Data curation, Formal analysis. **D Beaton:** Conceptualization, Data curation, Formal analysis. **JY Kim:** Conceptualization, Data curation. **M Needham:** Conceptualization, Data curation. **H Alexanderson:** Conceptualization, Data curation. **L Christopher-Stine:** Conceptualization, Data curation. **CA Mecoli:** Formal analysis. **JK Park:** Conceptualization, Data curation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The authors declare no relevant competing interests except that L.J.M. and D.B. are paid staff members of OMERACT.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.semarthrit.2023.152339.

References

[1] Limaye VS, Lester S, Blumbergs P, Roberts-Thomson PJ. Idiopathic inflammatory myositis is associated with a high incidence of hypertension and diabetes mellitus. Int J Rheum Dis 2010;13(2):132–7.
[2] Oldroyd A, Lilleker J, Chinoy H. Idiopathic inflammatory myopathies - a guide to subtypes, diagnostic approach and treatment. Clin Med 2017;17(4):322–8.
[3] Leclair V, Lundberg IE. New myositis classification criteria-what we have learned since Bohan and Peter. Curr Rheumatol Rep 2018;20(4):18.
[4] Lundberg IE, Fujimoto M, Vencovsky J, et al. Idiopathic inflammatory myopathies. Nat Rev Dis Primers 2021;7(1):86.
[5] OMERACT Myositis Working Group. 2023. https://omeract.org/working-groups/myositis/. In.

[6] DiRenzo D, Bingham 3rd CO, Mecoli CA. Patient-reported outcomes in adult idiopathic inflammatory myopathies. Curr Rheumatol Rep 2019;21(11):62.
[7] Saygin D, Oddis CV, Dzanko S, et al. Utility of patient-reported outcomes measurement information system (PROMIS) physical function form in inflammatory myopathy. Semin Arthritis Rheum 2021;51(3):539–46.
[8] Alexanderson H, Del Grande M, Bingham 3rd CO, et al. Patient-reported outcomes and adult patients' disease experience in the idiopathic inflammatory myopathies. report from the OMERACT 11 Myositis Special Interest Group. J Rheumatol 2014;41(3):581–92.
[9] Mecoli CA, Park JK, Alexanderson H, et al. Perceptions of patients, caregivers, and healthcare providers of idiopathic inflammatory myopathies: an International OMERACT Study. J Rheumatol 2019;46(1):106–11.
[10] Park JK, Mecoli CA, Alexanderson H, et al. Advancing the development of patient-reported outcomes for adult myositis at OMERACT 2016: an International Delphi Study. J Rheumatol 2017;44(11):1683–7.
[11] Esfandiary T, Park JK, Alexanderson H, et al. Assessing the content validity of patient-reported outcome measures in adult myositis: a report from the OMERACT myositis working group. Semin Arthritis Rheum 2020;50(5):943–8.
[12] DiRenzo D, Saygin D, de Groot I, et al. Reliability and validity of PROMIS physical function, pain interference, and fatigue as patient reported outcome measures in adult idiopathic inflammatory myopathies: international study from the OMERACT myositis working group. Semin Arthritis Rheum 2023;58:152111.
[13] The OMERACT handbook, version 2.1. 2021. Updated April 1stAvailable from: https://omeracthandbook.org.
[14] Maxwell LJ, Beaton DE, Boers M, et al. The evolution of instrument selection for inclusion in core outcome sets at OMERACT: Filter 2.2. Semin Arthritis Rheum 2021;51(6):1320–30.
[15] Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). N Engl J Med 1975;292(8):403–7.
[16] Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). N Engl J Med 1975;292(7):344–7.
[17] Lundberg IE, Tjarnlund A, Bottai M, et al. European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Arthritis Rheumatol 2017;69(12):2271–82.
[18] Cohen J. Statistical power analysis for the behavioral sciences. Academic Press; 2013.
[19] Dancey CP, Reidy J. Statistics without maths for psychology. Pearson Education; 2007.
[20] Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10(4):407–15.
[21] Seamon BA, Kautz SA, Bowden MG, Velozo CA. Revisiting the concept of minimal detectable change for patient-reported outcome measures. Phys Ther 2022;102(8).
[22] Lydick E, Epstein RS. Interpretation of quality of life changes. Qual Life Res 1993;2(3):221–6.
[23] Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol 2008;61(2):102–9.
[24] Bartlett SJ, Gutierrez AK, Andersen KM, et al. Identifying minimal and meaningful change in a patient-reported outcomes measurement information system for rheumatoid arthritis: use of multiple methods and perspectives. Arthritis Care Res 2022;74(4):588–97.
[25] Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. Qual Life Res 2002;11(3):207–21.
[26] Hendrikx J, Fransen J, Kievit W, van Riel PL. Individual patient monitoring in daily clinical practice: a critical evaluation of minimal important change. Qual Life Res 2015;24(3):607–16.
[27] de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16(1):131–42.
[28] Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression. John Wiley & Sons; 2013.
[29] Leclair V, Tsui H, Hudson M. Pain in autoimmune inflammatory myopathies: a scoping review. RMD Open 2023;9(1).
[30] Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018;18(3):91–3.
[31] Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol 1997;50(8):869–79.
[32] de Vet H, Mokkink L, Knol D. Practical guides to biostatistics and epidemiology. Cambridge: Measurement in Medicine; 2011.
[33] Terwee CB, Peipert JD, Chapman R, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021;30(10):2729–54.
[34] Kamperman RG, Bogaards JA, Evers SW, et al. Treatment with add-on IVIg in Myositis Early In the diSease course May be sUperior to Steroids alone for reaching CLinical improvEment (TIME IS MUSCLE): study protocol of a phase-2 double-blind placebo-controlled randomised trial. BMJ Open 2023;13(7):e067435.