Rheumatology, 2023, **62**, 3672–3679
https://doi.org/10.1093/rheumatology/kead110
Advance access publication 17 March 2023
Original Article

British Society for Rheumatology RHEUMATOLOGY OXFORD

Clinical science

Performance of the 2016 ACR-EULAR Myositis Response Criteria in adult dermatomyositis/polymyositis therapeutic trials and consensus profiles

Didem Saygin [1], Hanna Kim [2], Christian Douglas[3], Brian Erman[3], Jesse Wilkerson[3], John A. McGrath[3], Chester V. Oddis[1], Ingrid E. Lundberg [4], Anthony A. Amato[5], Ignacio García-De La Torre[6], Hector Chinoy [7], David Fiorentino[8], Lorinda Chung[8], Yeong-Wook Song [9], Frederick W. Miller[10], Nicolino Ruperto [11], Jiri Vencovsky[12], Rohit Aggarwal[1], Lisa G. Rider [10,*]; for the International Myositis Assessment and Clinical Studies Group (IMACS)[‡]

[1]Section of Rheumatology at University of Chicago and Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[2]Juvenile Myositis Pathogenesis and Therapeutics Unit, National Institute of Arthritis Musculoskeletal and Skin Diseases, National Institutes of Health (NIH), Bethesda, MD, USA
[3]Social & Scientific Systems, Inc, Durham, NC, USA
[4]Division of Rheumatology, Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden
[5]Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[6]Hospital General de Occidente de la Secretaría de Salud and Universidad de Guadalajara, Department of Immunology and Rheumatology, Mexico
[7]National Institute for Health Research Manchester Biomedical Research Centre, Division of Musculoskeletal and Dermatological Sciences, Manchester University NHS Foundation Trust, The University of Manchester, Manchester, UK
[8]Department of Dermatology, Stanford University School of Medicine, Redwood City, CA, USA
[9]Medical Research Center, Institute of Human-Environment Interface Biology, Department of Internal Medicine, Seoul National University
[10]Environmental Autoimmunity Group, Clinical Research Branch, National Institute of Environmental Health Sciences, NIH, Bethesda, MD, USA
[11]IRCCS Istituto Giannina Gaslini, UOSID Centro Trial, Reumatologia, Pediatria II, PRINTO, Genoa, Italy
[12]Department of Rheumatology, 1st Medical Faculty, Institute of Rheumatology; Charles University, Prague, Czech Republic
D.S. and H.K. are co-first authors.

*Correspondence to: Lisa G. Rider, Environmental Autoimmunity Group Clinical Research Branch, National Institute of Environmental Health Sciences, National Institute of Health, CRC Room 6-5700, 10 Center Drive, MSC 1301, Bethesda, MD 20892-1301, USA. E-mail: riderl@mail.nih.gov
[‡]See Acknowledgements section for a list of the International Myositis Assessment and Clinical Studies Group (IMACS).

Abstract

Objective: The ACR-EULAR Myositis Response Criteria (MRC) were developed as a composite measure using absolute percentage change in six core set measures (CSMs). We aimed to further validate the MRC by assessing the contribution of each CSM, frequency of strength *vs* extramuscular activity improvement, representation of patient-reported outcome measures (PROM), and frequency of CSM worsening.

Methods: Data from adult dermatomyositis/polymyositis patients in the rituximab ($n=147$), etanercept ($n=14$), and abatacept ($n=19$) trials, and consensus patient profiles ($n=232$) were evaluated. The Total Improvement Score (TIS), number of improving *vs* worsening CSMs, frequency of improvement with and without muscle-related CSMs, and contribution of PROM were evaluated by MRC category. Regression analysis was performed to assess contribution of each CSM to the MRC.

Results: Of 412 adults with dermatomyositis/polymyositis, there were 37%, 24%, 25%, and 14% with no, minimal, moderate, and major MRC improvement, respectively. The number of improving CSMs and absolute percentage change in all CSMs increased by improvement category. In minimal-moderate improvement, only physician-reported disease activity contributed significantly more than expected by MRC. Of patients with at least minimal improvement, 95% had improvement in muscle-related measures and a majority (84%) had improvement in PROM. Patients with minimal improvement had worsening in a median of 1 CSM, and most patients with moderate-major improvement had no worsening CSMs. Physician assessment of change generally agreed with MRC improvement categories.

Conclusion: The ACR-EULAR MRC performs consistently across multiple studies, further supporting its use as an efficacy end point in future myositis therapeutic trials.

Keywords: DM, PM, myositis, response criteria, outcome assessment, outcome measure, clinical trial

Received: 23 November 2022. **Accepted:** 1 March 2023

Rheumatology key messages

- Most dermatomyositis/polymyositis patients with improvement by Myositis Response Criteria improved in muscle-related and patient-reported outcome measures.
- Worsening in myositis core set measures was infrequent in patients with Myositis Response Criteria improvement.
- The Myositis Response Criteria performed consistently across multiple studies, validating their use in therapeutic trials.

Introduction

The idiopathic inflammatory myopathies (IIMs) are heterogeneous autoimmune diseases, characterized by muscle weakness and systemic features, including rashes, arthritis, and interstitial lung disease [1]. In order to assess response to therapy, the 2016 ACR-EULAR Criteria for Minimal, Moderate, and Major Clinical Response for dermatomyositis (DM) and polymyositis (PM), known as the Myositis Response Criteria, were developed as a composite measure that reflects changes in six differentially weighted core set activity measures. These six myositis core set measures include Physician and Patient Global Disease Activity, Manual Muscle Testing to assess strength, Health Assessment Questionnaire (HAQ) for impact on physical function, Extramuscular Disease Activity, and muscle enzymes; a higher value for each measure is associated with greater disease activity except for Manual Muscle Testing, for which lower values are associated with greater weakness [2]. The Myositis Response Criteria provide a continuous Total Improvement Score (TIS) ranging from 0 to 100, with thresholds for minimal, moderate and major improvement of 20, 40 and 60, respectively [3].

A comprehensive process was used to develop the Myositis Response Criteria, including expert rating of patient profiles, validation of the candidate criteria in several studies, and a nominal group technique for achieving expert consensus [4]. The relative weight of each core set measure was determined using the Potentially All Pairwise RanKings of all possible Alternatives (PAPRIKA) methodology and a conjoint analysis survey, with physician ratings as the gold standard [3]. Absolute percentage changes in the six individual core set measures are summed to provide the Total Improvement Score. The core set measures are differentially weighted based on their relative importance: Manual Muscle Testing is weighted most heavily, followed by Physician Global Activity and Extramuscular Activity, with muscle enzyme having least relative importance.

Given the ability to use the Myositis Response Criteria as both a continuous and categorical measure, its good initial performance characteristics, and ease of computation with a web calculator, the Myositis Response Criteria has been widely adopted and utilized as an end point in several myositis therapeutic trials, including two demonstrating efficacy in DM, with one leading to regulatory approval [5, 6]. Notwithstanding its strengths, further validation of the Myositis Response Criteria is required, and concerns in several areas need to be addressed. The contribution of each core set measure to the Total Improvement Score and the various Myositis Response Criteria improvement categories remain to be examined in real-life studies. It is unclear whether patients can achieve response as per the Myositis Response Criteria without improvement in muscle strength, a concern of regulatory agencies. The degree of representation of patient perspectives in the Myositis Response Criteria is unknown. The Myositis Response Criteria also differs from the previous response criteria, known as the Definition of Improvement, which specified a maximum number of worsening core set measures in the setting of improvement [7]. It is unclear whether patients can achieve improvement as per the Myositis Response Criteria with simultaneous worsening in core set measures.

The objectives of the present study were to assess the performance of the Myositis Response Criteria in adult DM/PM therapeutic trials and natural history patient profiles. We examined the contribution of each core set measure to Total Improvement Score, the frequency of strength *vs* extramuscular disease activity (Extramuscular) improvement, representation of patient-reported outcome measures, frequency of worsening in core set measures, and agreement between physician assessment of change categories *vs* the Myositis Response Criteria categories.

Methods

Patients

Data from the rituximab in myositis ($n = 147$; NCT00106184), etanercept in DM ($n = 14$; NCT00282880), and abatacept treatment of DM/PM trials ($n = 19$; NCT01315938), and DM/PM consensus patient profiles from natural history and open-label treatment studies ($n = 232$) were included (Supplementary Table S1, available at *Rheumatology* online) [3, 8–14]. Patients had at least moderately active disease in the rituximab trial, while the other two treatment studies required active disease without defined core set measure thresholds, and the natural history studies had no disease activity enrolment criteria [3, 9]. The studies comply with the Declaration of Helsinki. The locally appointed ethics committees have approved the research protocols, and the present study was approved under a myositis natural history protocol (94-E-0165) by the National Institutes of Health (NIH) institutional review board. Written informed consent has been obtained from the subjects (or their legally authorized representative).

Statistical analyses

The Total Improvement Score, number of improving and worsening core set measures, and absolute percentage change in each core set measure by Myositis Response Criteria category were described. A core set measure was considered improving or worsening if the absolute percentage change was >5%, except for manual muscle testing which was >2% absolute percentage change, as per the Myositis Response Criteria definition [4]. The Wilcoxon test, with Bonferroni adjustment to the *P*-value thresholds, was performed for comparison among the Myositis Response Criteria categories. The expected contribution of each core set measure was calculated as the maximum contribution of each core set measure

to the maximum possible Total Improvement Score, based on the core set measure weights from the Myositis Response Criteria [4]. The observed *vs* expected percentage contribution of each core set measure to Total Improvement Score was compared by the Sign test. Generalized linear regression analysis was performed to assess the contribution of each core set measure to the Total Improvement Score. The frequency of improvement was calculated for muscle-related core set measures [Manual Muscle Testing, HAQ, and/or muscle enzyme (creatine kinase, CK)], non-muscle-related core set measures (Extramuscular), and patient-reported outcome measures (Patient Global Activity, HAQ). Agreement between categorical physician assessment of change and the Myositis Response Criteria categories was assessed by weighted Cohen's Kappa using data from the rituximab trial and consensus profiles.

Results

Distribution and improvement in core set measures by improvement category

By Myositis Response Criteria category, there was a significant monotonic increase in the Total Improvement Score and the number of core set measures that improved (Table 1). The median Total Improvement Score in all patients was 8 for no improvement, 28 for minimal improvement, 48 for moderate

Table 1. Distribution and change in core set measures by improvement categories for DM/PM studies[a]

	Myositis Response Criteria Categories[b]			
	No Improvement (*N* = 151, 36.7%)	**Minimal Improvement (*N* = 101, 24.5%)**	**Moderate Improvement (*N* = 102, 24.8%)**	**Major Improvement (*N* = 58, 14.1%)**
Median Total Improvement Score	7.5 [2.5–12.5]	27.5 [22.5–32.5][d]	47.5 [42.5–52.5][d,e]	70.0 [65.0–77.5][d,e,f]
Median number of core set measures improved	1.0 (0.0–3.0)	3.0 (1.0–5.0)[d]	5.0 (3.0–6.0)[d,e]	6.0 (4.0–6.0)[d,e,f]
Median absolute percentage change in each core set measure				
Physician Global Disease Activity	0.0 [–7.0 to 4.0]	9.0 [3.0–15.0][d]	19.5 [12.0–27.0][d,e]	30.0 [21.0–42.0][d,e,f]
Patient Global Disease Activity	–2.0 [–12.0 to 7.0]	9.0 [–3.0 to 21.0][d]	21.0 [6.0–35.0][d,e]	37.5 [29.0–49.0][d,e,f]
Manual Muscle Testing	0.0 [–4.0 to 3.0]	5.0 [1.0–9.0][d]	10.0 [6.0–15.0][d,e]	20.0 [16.0–25.0][d,e,f]
HAQ	0.0 [–8.0 to 4.0]	4.0 [0.0–13.0][d]	11.0 [0.0–21.0][d]	29.0 [17.0–42.0][d,e,f]
Extramuscular Disease Activity	–2.0 [–8.5 to 0.0]	5.0 [0.0–13.0][d]	13.0 [6.0–20.0][d,e]	21.0 [12.0–33.0][d,e,f]
Muscle Enzyme	1.0 [–8.0 to 8.0]	6.0 [–1.0 to 18.0][d]	14.5 [2.0–36.0][d,e]	39.0 [11.0–86.0][d,e,f]
Median percentage contribution to Total Improvement Score[c]				
Physician Global Disease Activity (20% Expected Contribution)	0.0 [0.0–20.0][g]	23.1 [0.0–33.3][g]	30.4 [17.7–35.0][g]	23.8 [21.2–26.7][g]
Patient Global Disease Activity (10% Expected Contribution)	0.0 [0.0–14.3][g]	7.1 [0.0–22.2]	10.5 [5.0–15.8]	10.6 [9.1–13.3][g]
Manual Muscle Testing (32% Expected Contribution)	0.0 [0.0–57.1][g]	33.3 [0.0–47.2]	25.0 [20.0–42.1]	32.2 [28.6–36.7]
HAQ (10% Expected Contribution)	0.0 [0.0–10.0][g]	0.0 [0.0–21.4]	10.8 [0.0–15]	10.7 [9.4–12.5][g]
Extramuscular Global Activity (20% Expected Contribution)	0.0 [0.0–20.0][g]	0.0 [0.0–30.0]	17.7 [13.6–27.3]	17.9 [12.5–20.8][g]
Muscle Enzyme (7.5% Expected Contribution)	7.5 [0.0–25.0][g]	6.7 [0.0–18.2]	6.3 [0.0–14.3]	8.7 [4.2–10.3]
Median number of core set measures worsening	2.0 (0.0–6.0)	1.0 (0.0–4.0)	0.0 (0.0–3.0)	0.0 (0.0–2.0)
Subjects with 1 worsening core set measure	29.0 (19.2%)	32.0 (31.7%)	24.0 (23.5%)	3.0 (5.2%)
Subjects with ≥2 worsening core set measures	89.0 (58.9%)	20.0 (19.8%)	7.0 (6.9%)	1.0 (1.7%)

[a] The results presented here are based on the combined data, while results for the individual studies are presented in Supplementary Tables S1, S3 and S5, available at *Rheumatology* online.
[b] Median values are shown with [interquartile range] or (range), or data is expressed as *n* (%). Threshold for the Myositis Response Criteria Improvement Categories of Minimal, Moderate and Major Improvement categories are ≥20, ≥40 and ≥60, respectively.
[c] The observed percentage contribution of each core set measure to the Total Improvement Score was calculated as: (core set measure Improvement Score/Total Improvement Score) × 100. The expected contribution of each core set measure was based on (maximum possible Total Improvement Score point contribution of core set measure/100).
[d] Statistically significant difference (*P*-value < 0.006) from the No Improvement category.
[e] Statistically significant difference (*P*-value < 0.006) from the Minimal Improvement category.
[f] Statistically significant difference (*P*-value < 0.006) from the Moderate Improvement category.
[g] Statistically significant difference (*P*-value < 0.005) from the Expected Contribution (Sign Test). Muscle Enzyme: most abnormal serum muscle enzyme value at baseline.

improvement, and 70 for major improvement. The median number of core set measures improved was 1 for patients with no improvement, 3 for minimal improvement, 5 for moderate improvement, and 6 for major improvement. These trends were similar in individual studies and were significant in the rituximab trial and consensus profiles (Supplementary Table S1, available at *Rheumatology* online).

As the degree of Myositis Response Criteria improvement increased from minimal to major, the absolute percentage change of each core set measure increased (Table 1). The median absolute percentage change of each core set measure ranged between 4 and 9% for minimal improvement, between 10 and 21% for moderate improvement, and between 20 and 39% for major improvement. This trend of increasing percentage change in core set measures by Myositis Response Criteria category was consistent in the rituximab trial and consensus profiles, and variable in the small therapeutic studies (Supplementary Table S2, available at *Rheumatology* online).

Relationship between baseline core set measure values and level of improvement category

In the combined data, patients with minimal improvement had higher baseline Physician Global and Extramuscular Activity, compared with patients who had no improvement (Supplementary Table S3, available at *Rheumatology* online). In the consensus profiles, patients who had moderate or major improvement had higher baseline Physician and Patient Global Activity, Extramuscular Activity, HAQ scores, muscle enzyme levels, and lower Manual Muscle Testing scores, compared with patients who had no or minimal improvement. However, this trend was not seen in the three therapeutic trials.

Contribution of core set measures to the Total Improvement Score

In patients with minimal to moderate improvement, the percentage contribution of each core set measure to the Total Improvement Score was as expected, except for Physician Global Activity, which contributed more than expected in the combined data (Supplementary Table S4, available at *Rheumatology* online). Physician Global Activity contributed as expected for minimal improvement in the consensus profiles and the rituximab trial, but more than expected for moderate improvement in the consensus profiles. In patients with major improvement, there was variation in the contribution of core set measures to Total Improvement Score, with Physician and Patient Global Activity and HAQ contributing more than expected and Extramuscular Activity contributing less than expected. Similar trends present in the combined data were observed in individual studies, although they did not have adequate power to detect significance (Table 1, Supplementary Table S4, available at *Rheumatology* online). By multiple regression analyses, all core set measures, except muscle enzyme, contributed significantly to the Total Improvement Score in the combined data. Similar results were seen in the consensus profiles, whereas in the rituximab trial, all core set measures, including muscle enzyme, contributed significantly to Total Improvement Score (Supplementary Table S5, available at *Rheumatology* online).

Distribution and worsening in core set measures by improvement category

Worsening of a core set measure occurred infrequently when patients met improvement by the Myositis Response Criteria. Patients with minimal improvement had worsening in a median of 1 core set measure, whereas patients with moderate or major improvement had a median of zero core set measures worsen in both the combined data and individual studies (Table 1; Supplementary Table S6A, available at *Rheumatology* online). For those with minimal improvement and worsening in any core set measure, the median absolute percentage worsening ranged from 6% to 18% (Supplementary Table S6B, available at *Rheumatology* online). Among patients achieving moderate improvement by the Myositis Response Criteria, 24% had 1 core set measure worsening and 7% had worsening in $\geq$2 core set measures. Patients with major improvement infrequently had worsening in any core set measure: 5% had worsening in 1 core set measure, and 2% had worsening in $\geq$2 core set measures (Table 1). In patients with minimal improvement, the most frequent core set measure that worsened was Patient Global Activity, in 20%, while muscle-related measures worsened in 10–13% of patients (Supplementary Table S6B, available at *Rheumatology* online). Specifically, Manual Muscle Testing worsened in 10% of patients with minimal, 5% with moderate, and none with major improvement. The degree of Manual Muscle Testing worsening was 5–6% in patients with minimal and moderate improvement.

Improvement in muscle-related measures by improvement category

Regarding the frequency of muscle strength improvement within the Myositis Response Criteria categories, Manual Muscle Testing improved in 69% of patients with minimal improvement, 86% of patients with moderate improvement, and in all patients (100%) with major improvement. Ninety percent of patients with minimal improvement and 98–100% of patients with moderate or major improvement had improvement in at least one muscle-related core set measure (Table 2). Conversely, <20% of patients with minimal or moderate improvement by the Myositis Response Criteria had Extramuscular Activity improvement alone, without improvement in Manual Muscle Testing, and fewer than 9% achieved minimal or moderate improvement without improvement in any muscle-related measure. None of those with major improvement had Extramuscular Activity improvement in the absence of improvement in any of the muscle-related measures. These results were consistent in the rituximab trial and consensus profiles (Supplementary Table S7, available at *Rheumatology* online).

Contribution of patient-reported outcome measures to the improvement categories

Regarding contribution of patient-reported outcome measures to Myositis Response Criteria improvement, 84% of patients with at least minimal improvement had improvement in Patient Global Activity or HAQ (Table 2). As Myositis Response Criteria improvement increased from minimal to major, the frequency of improvement in Patient Global Activity increased from 56% to 95%, and improvement in HAQ increased from 46% to 93% (Table 2). These trends held for individual studies, except the abatacept trial, where a smaller number of patients had improvement in patient-

Table 2. Distribution of muscle-related, extramuscular, and patient-reported measures by improvement categories in dermatomyositis/polymyositis studies[a]

Measures improved	Myositis Response Criteria Categories		
	Minimal Improvement ($N = 101$) *n* (%)	Moderate Improvement ($N = 102$) *n* (%)	Major Improvement ($N = 58$) *n* (%)
Frequency of Muscle-Related[b] *vs* Extramuscular Measure Contribution to Total Improvement Score			
Manual Muscle Testing contributing to Total Improvement Score	70 (69.3%)	88 (86.3%)[c]	58 (100.0%)[c,d]
Any muscle-related core set measure[b] contributing to Total Improvement Score	91 (90.1%)	100 (98.0%)[c]	58 (100.0%)[c]
Extramuscular Activity contributing to Total Improvement Score, without contribution of Manual Muscle Testing	19 (18.8%)	14 (13.7%)	0 (0.0%)[c,d]
Extramuscular Activity contributing to Total Improvement Score, without contribution of any muscle-related core set measure[b]	9 (8.9%)	2 (2.0%)	0 (0.0%)
Frequency of Patient-Reported Outcome Measure Contribution to Total Improvement Score			
Patient Global Activity contributing to Total Improvement Score	57 (56.4%)	81 (79.4%)[c]	55 (94.8%)[c,d]
HAQ contributing to Total Improvement Score	46 (45.5%)	60 (58.8%)	54 (93.1%)[c,d]
Patient Global Activity or HAQ contributing to Total Improvement Score	71 (70.3%)	91 (89.2%)[c]	58 (100.0%)[c,d]

[a] Results presented here are based on the combined data, while results for the individual studies are presented in the Supplementary Table S6, available at *Rheumatology* online.
[b] Any muscle-related core set measure included Manual Muscle Testing, HAQ, or Muscle Enzyme.
[c] Statistically significant difference (P-value < 0.017) from the Minimal Improvement category.
[d] Statistically significant difference (P-value < 0.017) from the Moderate Improvement category.

reported outcome measures (Supplementary Table S7, available at *Rheumatology* online).

Agreement between physician-assessed improvement categories and the Myositis Response Criteria categories

There were significant differences in the Total Improvement Score between physician-assessed improvement categories (Table 3). In the rituximab trial, the median Total Improvement Score was 38 for physician-rated slight improvement, 55 for moderate improvement, and 69 for marked improvement. For the consensus profiles, the median Total Improvement Score was 25 for minimal improvement, 48 for moderate improvement, and 75 for major improvement. Physician-assessed change categories had significant agreement with the Myositis Response Criteria categories. The agreement between these categories was 0.5 (SE 0.05) in the rituximab trial and 0.8 (SE 0.03) in the consensus profiles.

Discussion

We assessed the performance of the ACR-EULAR Myositis Response Criteria in three adult DM/PM randomized therapeutic trials and a large consensus profile dataset. We demonstrated an increasing contribution of core set measures across the improvement categories. Patients who meet Myositis Response Criteria improvement rarely experience worsening in core set measures. The majority of DM/PM patients who improve by the Myositis Response Criteria show improvement in muscle disease, while patient-reported outcome measures also contribute to improvement.

In our study, contribution of Physician Global Activity to the Total Improvement Score was significantly more than expected in patients with minimal improvement. This could be related to it being assigned the second-highest relative weight in the Myositis Response Criteria [4]. Although Manual Muscle Testing has the highest relative weight in the Myositis Response Criteria, it is usually less sensitive to change, making Physician Global Activity an important contributor to the Total Improvement Score. As the level of Myositis Response Criteria improvement increased from minimal to major, patients had an increasing number of improving core set measures and an increase in the absolute percentage change in all core set measures. Those who had moderate to major improvement showed improvement in almost all core set measures and $\geq$10% improvement in all six core set measures. Thus, despite the greater than expected contribution of Physician Global Activity to improvement, typically there was a significant degree of improvement in multiple core set measures, rather than improvement being driven by a single core set measure. When all core set measures are taken together, each core set measure contributed significantly to Total Improvement Score, except muscle enzyme. This could be related to the small relative weight assigned to muscle enzyme [4]. Furthermore, muscle enzymes may not change significantly in patients with active muscle disease or may not correlate with disease activity, particularly in DM [15]. In the original Myositis Response Criteria study, muscle enzyme was ranked as the least important core set measure in determining improvement [4].

More than 90% of patients with minimal to moderate improvement and all patients with major improvement had improvement in muscle-related measures. As DM/PM is characteristically a muscle disease, the ability of the Myositis Response Criteria to capture improvement in muscle strength is encouraging. Only a small number of patients with at least minimal improvement had isolated improvement in Extramuscular Activity without corresponding improvement

Table 3. Distribution of improvement by physician-assessed change categories in DM/PM studies

Study	Physician-assessed change categories	Median Total Improvement Score [IQR]	Myositis Response Criteria Categories			
			No Improvement *n* (%)	Minimal Improvement *n* (%)	Moderate Improvement *n* (%)	Major Improvement *n* (%)
Rituximab trial[ab] ($N = 147$)	No Improvement/Worsening ($n = 46$)	12.5 [5.0–25.0]	31 (67.4)	9 (19.6)	5 (10.9)	1 (2.2)
	Slight Improvement ($n = 45$)	37.5 [25.0–47.5][c]	3 (6.7)	22 (48.9)	17 (37.8)	3 (6.7)
	Moderate Improvement ($n = 46$)	55.0 [42.5–62.5][c,d]	1 (2.2)	6 (13.0)	23 (50.0)	16 (34.8)
	Marked Improvement ($n = 10$)	68.7 [47.5–75.0][c,d]	0 (0.0)	1 (10.0)	3 (30.0)	6 (60.0)
Consensus profiles[ab] ($N = 228$)	No Improvement/Worsening ($n = 87$)	7.5 [0.0–10.0]	80 (92.0)	7 (8.0)	0 (0.0)	0 (0.0)
	Minimal Improvement ($n = 62$)	25.0 [17.5–35.0][c]	17 (27.4)	32 (51.6)	13 (21.0)	0 (0.0)
	Moderate Improvement ($n = 61$)	47.5 [37.5–55.0][c,d]	2 (3.3)	14 (23.0)	32 (52.5)	13 (21.3)
	Major Improvement ($n = 18$)	75.0 [70.0–90.0][c,d,e]	0 (0.0)	0 (0.0)	2 (11.1)	16 (88.9)

[a] Physician-assessed change categories were named differently in the rituximab trial: Slight Improvement rather than Minimal Improvement, and Marked Improvement rather than Major Improvement. Consensus was defined as >50% consensus for each specific improvement category.
[b] Statistically significant (P-value < 0.05) agreement between physician-assessed change categories and Myositis Response Criteria categories.
[c] Statistically significant difference ($P < 0.008$) in Total Improvement Score distribution from the No Improvement category.
[d] Statistically significant difference ($P < 0.008$) in Total Improvement Score distribution from the Minimal Improvement category.
[e] Statistically significant difference ($P < 0.008$) in Total Improvement Score distribution from the Moderate Improvement category. IQR: interquartile range.

in muscle-related measures. The infrequent improvement in Extramuscular Activity suggests difficulty achieving response without improvement in muscle-related measures in patients with active muscle disease. However, we could not evaluate Myositis Response Criteria performance in DM patients with skin-predominant disease with minimal or no muscle involvement, due to the small number of patients with isolated active skin disease enrolled in these studies. In an open-label trial of tofacitinib in patients with refractory DM with active skin disease and minimal muscle disease, all patients achieved at least minimal improvement by the Myositis Response Criteria, providing encouraging results regarding validity of the Total Improvement Score in patients with primarily active cutaneous disease [6]. Further studies are required to assess the performance of the Total Improvement Score in DM patients with predominantly active skin or other target organ disease, including pulmonary, and inactive or mildly active muscle disease.

The majority of patients achieving minimal to moderate improvement and all patients with major improvement had improvement in either Patient Global Activity or HAQ. Adequate representation of patient perspectives, including physical function/disability, is important for the Myositis Response Criteria to comprehensively assess the impact of therapeutic interventions [16].

The previous Definition of Improvement ensured that the criteria captured patients who improve, and who do not worsen significantly in some core set measures. Thus, it allowed no >2 core set measures to worsen by ≥25%, which could not include Manual Muscle Testing [7]. In this validation study, few core set measures worsened in patients who achieved improvement by the Myositis Response Criteria, with a median of only 1 worsening core set measure in patients with minimal improvement and zero worsening core set measures in patients with moderate to major improvement. Worsening in Manual Muscle Testing was seen in <10% of patients who achieved minimal to moderate improvement. These results suggest that patients who achieve improvement by the Myositis Response Criteria rarely experience worsening in core set measures, including Manual Muscle Testing. Thus, the Myositis Response Criteria likely does not need an additional requirement to limit worsening in core set measures, although further studies on the degree and frequency of worsening are needed.

Physician-assessed change categories and Myositis Response Criteria categories were in significant agreement in both the rituximab trial and the consensus profiles. The degree of agreement was noted to be higher in the consensus profiles, which were based on agreement among clinician experts pertaining to whether or not patients were at least minimally improved [4]. Therefore, the higher estimation of agreement between the Myositis Response Criteria and physician-assessed change categories in the consensus profiles may not be surprising. When the Myositis Response Criteria were developed, insufficient data were available to finalize the threshold for major improvement [4]. Here, the Total Improvement Score distribution of the physician-assessed moderate and marked improvement categories were not significantly different. The number of patients who had major improvement was small; thus, further studies are needed to assess the optimal threshold for major improvement.

This is the first large-scale study to assess the performance of the 2016 ACR-EULAR Myositis Response Criteria using comprehensive data from previous myositis therapeutic trials and natural history studies. One of the limitations is the small size of some clinical trials, limiting conclusions from these individual studies. However, the pooled data sample sizes were satisfactory. A major limitation is that the consensus profiles and the rituximab trial included in this study were also used in the development of the Myositis Response Criteria. Furthermore, there was overlap in experts who contributed to this manuscript, provided clinical data, and participated in the development of the Myositis Response Criteria. Given these overlaps and limitations in breadth of ideas across these data sources, generalizability of the results may be limited. Thus, larger and new data sources with broader inputs are needed to further test the performance of the Myositis Response Criteria. Despite overlap of contributing studies with the original Myositis Response Criteria development, the results presented here have not been previously assessed. Another limitation is that the patients included in our study had a diagnosis of DM or PM, and we were only able to

analyse the performance of the Myositis Response Criteria for these subgroups, although the PM group included patients with anti-synthetase syndrome and immune-mediated necrotizing myopathy. Further studies are required to assess the performance of the Myositis Response Criteria in other IIM subgroups, including anti-synthetase syndrome and immune-mediated necrotizing myopathy, although their inclusion in the PM subgroup suggests that performance of the Myositis Response Criteria will likely be adequate. In addition, given the Myositis Response Criteria is weighted more towards muscle-related core set measures, it may be less sensitive in reflecting changes in patients with clinically amyopathic or skin-predominant DM.

In conclusion, this study advances our understanding of the Myositis Response Criteria and their performance characteristics. It addresses several concerns regarding the contribution from and changes in individual core set measures, including demonstration of inclusion of patient perspectives, and agreement of Myositis Response Criteria with physician assessment of change. The ACR-EULAR Myositis Response Criteria perform consistently across multiple studies, supporting their use as an important clinical end point in future myositis clinical trials.

Supplementary material

Supplementary material is available at *Rheumatology* online.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding

This research was supported by the Intramural Research Programs of the National Institute of Environmental Health Sciences [ZIA ES101081], and National Institute of Arthritis and Musculoskeletal and Skin Diseases [ZIA AR041215-02] of the National Institutes of Health, by the NIHR Biomedical Research Centre Funding Scheme, European Union (contract QLG1-CT-2000–00514), European League Against Rheumatism, and American College of Rheumatology. Social & Scientific Systems was supported under a contract with NIEHS (HHSN273201600011C). Ingrid E. Lundberg received support from the Swedish Research council, and Stockholm Region Council (ALF project). Jiri Vencovsky also received support from the Czech Ministry of Health, Conceptual Development of Research Organization 00023728 (Institute of Rheumatology).

Disclosure statement: Authors disclosed the following additional sources of funding within the past 3 years or that were used to support the clinical trial data in this study. H.K. received other financial/material support, including the study drug (baricitinib) from Eli Lilly and Company. C.V.O. received clinical trial support for the rituximab trial NIAMS from the NIH (contract N01-AR-4–2273), and the study drug was provided by Genentech. I.E.L. received grants from the Swedish Rheumatism Association, the King Gustaf V 80 Year Foundation, and the King Gustaf V and Queen Victoria's Freemason Foundation; and support for the abatacept trial in the form of grants from Bristol-Myers Squibb (BMS), the Börje Dahlin Foundation, the Swedish Research Council (Grant K2014-52X-14045–14-3), the Swedish Rheumatism Association, the King Gustaf V 80-Year Foundation, and the regional agreement on medical training and clinical research between Stockholm County Council and the Karolinska Institutet; BMS provided the study drug. C.V.O. also participates in the Data Safety Monitoring Boards or Advisory Boards of Orphazyme, Kezar, and Janssen. A.A. discloses that the etanercept trial was supported by the NIH National Institute of Neurological Disorders and Stroke (R01 NS049639), and that Amgen provided start-up funding, etanercept, and matching placebo. He also discloses that he is a Medical Advisory Board Member/Consultant for Horizon Pharmaceuticals, Ono Pharm, Argenx, and Abcuro. H.C. has received grants from the Medical Research Council, Eli Lilly, and UCB; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from UCB, Biogen, and Astra Zeneca; support for attending meetings and/or travel from Abbvie; and he discloses participation on a Data Safety Monitoring Board or Advisory Board for Orphazyme, Kezar, and Argenx. L.C. has received grants from Boerhinger Ingelheim; consulting fees from Eicos Sciences, Genentech, Kyverna, Jasper, and Mitsubishi. F.W.M. has received clinical trial support for the rituximab trial from NIEHS, the NIH intramural research program (ZIA ES101081) and NIAMS (contract N01-AR-4–2273). N.R. has received consulting fees, payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Bridge, Amgen, AstraZeneca, Aurinia, Bayer, Bristol Myers Squibb, Celgene, inMed, Cambridge Healthcare Research, Domain Therapeutic, EMD Serono, Glaxo Smith Kline, Idorsia, Janssen, Eli Lilly, Novartis, Pfizer, Sobi, and UCB; N.R. also discloses participation on a Data Safety Monitoring Board or Advisory Board for Pfizer and Eli Lilly. J.V. has received consulting fees from Argenx; payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Werfen and Octapharma; J.V. discloses participation on a Data Safety Monitoring Board or Advisory Board for Horizon, Boehringer, and Octapharma. R.A. has received a Research Grant from Mallinckrodt, Pfizer, Bristol Myers-Squibb, Q32, and EMD Serono; R.A. is a Consultant for Mallinckrodt, Octapharma, CSL Behring, Bristol Myers-Squibb, EMD Serono, Kezar, Pfizer, AstraZeneca, Alexion, Argenx, Boehringer Ingelheim, Corbus, Janssen, Kyverna, Roivant, Merck, Galapagos, Actigraph, Abbvie, Scipher, Horizontal Therapeutics, Teva, Biogen, Beigene, ANI Pharmaceutical, Bioscience, and Capella. L.G.R. has received support for the rituximab trial from NIEHS and the NIH intramural research program (ZIA ES101081) and for NIAMS from the NIH (contract N01-AR-4–2273); L.G.R. has received other clinical study support from BMS and Hope Pharmaceuticals; L.G.R. is an unpaid consultant for Alexion, Argenx, Astra Zeneca, Boehringer Ingelheim, CSL Behring, Horizon Therapeutics, and Pfizer.

The views expressed in this publication are those of the authors and not necessarily those of the National Institutes of Health, the Department of Health and Human Services, the NHS, the National Institute for Health Research, or the Department of Health.

Acknowledgements

We thank Drs Susan Kim, Iago Pinal-Fernandez, Dana Ascherman, and Adam Schiffenbauer, and the IMACS Scientific Review Committee for critical reading of the manuscript. We thank members of IMACS, including participants in the Myositis Response Criteria consensus conference, the adult core set survey group, the patient profile working group, the conjoint analysis group, and the statistical team, as noted previously [4].

References

1. Lundberg IE, Fujimoto M, Vencovsky J *et al.* Idiopathic inflammatory myopathies. Nat Rev Dis Primers 2021;7:86.
2. Miller FW, Rider LG, Chung YL *et al.*; International Myositis Outcome Assessment Collaborative Study Group. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. Rheumatology (Oxford) 2001;40:1262–73.
3. Aggarwal R, Rider LG, Ruperto N *et al.*, International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: an International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. Ann Rheum Dis 2017;76:792–801.
4. Rider LG, Ruperto N, Pistorio A *et al.*; International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 ACR-EULAR adult dermatomyositis and polymyositis and juvenile dermatomyositis response criteria-methodological aspects. Rheumatology (Oxford) 2017;56:1884–93.
5. Aggarwal R, Charles-Schoeman C, Schessl J *et al.*; ProDERM Trial Group. Trial of intravenous immune globulin in dermatomyositis. N Engl J Med 2022;387:1264–78.
6. Paik JJ, Casciola-Rosen L, Shin JY *et al.* Study of tofacitinib in refractory dermatomyositis: an open-label pilot study of ten patients. Arthritis Rheumatol 2021;73:858–65.
7. Rider LG, Giannini EH, Brunner HI *et al.*; International Myositis Assessment and Clinical Studies Group. International consensus on preliminary definitions of improvement in adult and juvenile myositis. Arthritis Rheum 2004;50:2281–90.
8. Study MG. A randomized, pilot trial of etanercept in dermatomyositis. Ann Neurol 2011;70:427–36.
9. Oddis CV, Reed AM, Aggarwal R *et al.*; RIM Study Group. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 2013;65:314–24.
10. Tjarnlund A, Tang Q, Wick C *et al.* Abatacept in the treatment of adult dermatomyositis and polymyositis: a randomised, phase IIb treatment delayed-start trial. Ann Rheum Dis 2018;77:55–62.
11. Zong M, Dorph C, Dastmalchi M *et al.* Anakinra treatment in patients with refractory inflammatory myopathies and possible predictive response biomarkers: a mechanistic study with 12 months follow-up. Ann Rheum Dis 2014;73:913–20.
12. Chung L, Genovese MC, Fiorentino DF. A pilot trial of rituximab in the treatment of patients with dermatomyositis. Arch Dermatol 2007;143:763–7.
13. Dastmalchi M, Grundtman C, Alexanderson H *et al.* A high incidence of disease flares in an open pilot study of infliximab in patients with refractory inflammatory myopathies. Ann Rheum Dis 2008;67:1670–7.
14. Lundberg IE, Svensson J. Registries in idiopathic inflammatory myopathies. Curr Opin Rheumatol 2013;25:729–34.
15. Kroll M, Otis J, Kagen L. Serum enzyme, myoglobin and muscle strength relationships in polymyositis and dermatomyositis. J Rheumatol 1986;13:349–55.
16. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas 2018;9:353–67.