ARTHRITIS & RHEUMATOLOGY
Vol. 00, No. 00, Month 2017, pp 00–00
DOI 10.1002/art.40064

SPECIAL ARTICLE

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

Rohit Aggarwal,[1] Lisa G. Rider,[2] Nicolino Ruperto,[3] Nastaran Bayat,[2] Brian Erman,[4] Brian M. Feldman,[5] Chester V. Oddis,[1] Anthony A. Amato,[6] Hector Chinoy,[7] Robert G. Cooper,[8] Maryam Dastmalchi,[9] David Fiorentino,[10] David Isenberg,[11] James D. Katz,[2] Andrew Mammen,[12] Marianne de Visser,[13] Steven R. Ytterberg,[14] Ingrid E. Lundberg,[9] Lorinda Chung,[10] Katalin Danko,[15] Ignacio García-De la Torre,[16] Yeong Wook Song,[17] Luca Villa,[3] Mariangela Rinaldi,[3] Howard Rockette,[1] Peter A. Lachenbruch,[2] Frederick W. Miller,[2] and Jiri Vencovsky,[18] for the International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation

This criteria set has been approved by the American College of Rheumatology (ACR) Board of Directors and the European League Against Rheumatism (EULAR) Executive Committee. This signifies that the criteria set has been quantitatively validated using patient data, and it has undergone validation based on an independent data set. All ACR/EULAR-approved criteria sets are expected to undergo intermittent updates.

The ACR is an independent, professional, medical and scientific society that does not guarantee, warrant, or endorse any commercial product or service.

This article is published simultaneously in the May 2017 issue of *Annals of the Rheumatic Diseases*.

Supported in part by the American College of Rheumatology, the European League Against Rheumatism, Cure JM Foundation, Myositis UK, Istituto G. Gaslini and the Paediatric Rheumatology International Trials Organisation (PRINTO), the Myositis Association, and the NIH (National Institute of Environmental Health Sciences [NIEHS], National Center for Advancing Translational Sciences, and National Institute of Arthritis and Musculoskeletal and Skin Diseases). Dr. García-De la Torre's work was supported in part by CONACYT (Programa Nacional de Posgrados de Calidad). Dr. Song's work was supported by the Korea Health Technology R & D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Republic of Korea (grant HI14C1277). Dr. Vencovsky's work was supported by the Ministry of Health, Czech Republic (Institute of Rheumatology project for conceptual development of a research organization, 00023728).

[1]Rohit Aggarwal, MD, MSc, Chester V. Oddis, MD, Howard Rockette, PhD: University of Pittsburgh, Pittsburgh, Pennsylvania; [2]Lisa G. Rider, MD, Nastaran Bayat, MD, James D. Katz, MD, Peter A. Lachenbruch, PhD, Frederick W. Miller, MD, PhD: NIEHS, NIH, Bethesda, Maryland; [3]Nicolino Ruperto, MD, MPH, Luca Villa, MA, Mariangela Rinaldi, MEng: Istituto Giannina Gaslini, Pediatria II - Reumatologia, PRINTO, Genoa, Italy; [4]Brian Erman, MS: Social and Scientific Systems, Inc., Durham, North Carolina; [5]Brian M. Feldman, MD, MSc, FRCPC: The Hospital for Sick Children, Toronto, Ontario, Canada; [6]Anthony A. Amato, MD: Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; [7]Hector Chinoy, PhD, MRCP: Central Manchester University Hospitals NHS Foundation Trust, University of Manchester, Manchester, UK; [8]Robert G.

Objective. **To develop response criteria for adult dermatomyositis (DM) and polymyositis (PM).**

Methods. **Expert surveys, logistic regression, and conjoint analysis were used to develop 287 definitions using core set measures. Myositis experts rated greater improvement among multiple pairwise scenarios in conjoint analysis surveys, where different levels of improvement in 2 core set measures were presented. The PAPRIKA (Potentially All Pairwise Rankings of All Possible Alternatives) method determined the relative weights of core set measures and conjoint analysis definitions. The performance characteristics of the definitions were evaluated on patient profiles using expert consensus (gold standard) and were validated using data from a clinical trial. The nominal group technique was used to reach consensus.**

Results. **Consensus was reached for a conjoint analysis–based continuous model using absolute percent change in core set measures (physician, patient, and extramuscular global activity, muscle strength, Health Assessment Questionnaire, and muscle enzyme levels). A total improvement score (range 0–100), determined by summing scores for each core set measure, was based on improvement in and relative weight of each core set measure. Thresholds for minimal, moderate, and major improvement were ≥20, ≥40, and ≥60 points in the total improvement score. The same criteria were chosen for juvenile DM, with different improvement thresholds. Sensitivity and specificity in DM/PM patient cohorts were 85% and 92%, 90% and 96%, and 92% and 98% for minimal, moderate, and major improvement, respectively. Definitions were validated in the clinical trial analysis for differentiating the physician rating of improvement ($P < 0.001$).**

Conclusion. **The response criteria for adult DM/PM consisted of the conjoint analysis model based on absolute percent change in 6 core set measures, with thresholds for minimal, moderate, and major improvement.**

Idiopathic inflammatory myopathies are a group of acquired, heterogeneous, systemic connective tissue diseases that include adult dermatomyositis (DM) and polymyositis (PM) and juvenile DM (1). Despite significant morbidity and mortality associated with DM/PM, there are currently no therapies approved for these syndromes by the Food and Drug Administration or the European Medicines Agency based on randomized controlled trials. However, with the advancement in novel therapeutic agents that target various biologic pathways implicated in the pathogenesis of DM/PM (2), there is a need for well-designed clinical trials using validated and universally accepted outcome measures. Recently completed clinical trials in adult DM/PM and juvenile DM have used varying response criteria (3–5), again highlighting the need for both data- and consensus-driven criteria to be used uniformly in future studies. Core set measures of myositis disease activity for adult DM/PM clinical trials have been established and validated by the International Myositis Assessment and Clinical Studies Group (IMACS) (6–8); these measures were used as the foundation for the current study. We undertook this study because there is a need for composite response criteria in myositis, given the heterogeneity of the disease and the fact that no single core set measure adequately covers all the domains in myositis. For example, muscle enzyme levels can be normal in active DM, and active muscle weakness in DM can occur without active rash.

Preliminary response criteria for adult DM/PM had been developed and partially validated by IMACS; these criteria were based on at least 20% improvement in 3 of 6 core set measures, with no more than 2 core set measures worsening by at least 25% (which cannot be muscle strength) (8,9). However, those criteria were considered preliminary, because they were not prospectively validated. Moreover, newer methodologies such as conjoint analysis and other continuous or hybrid approaches for developing response criteria had not been evaluated (10–14). The preliminary criteria had other potential limitations, including equal weights being applied to each core set measure and the lack of quantitative or continuous outcomes. With the growing repertoire of potential therapeutic agents, some of which may yield better results than only minimal clinical improvement, there is also a need to develop criteria for moderate and major clinical improvement.

Cooper, MD: University of Liverpool, Liverpool, UK; [9]Maryam Dastmalchi, MD, PhD, Ingrid E. Lundberg, MD, PhD: Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden; [10]David Fiorentino, MD, PhD, Lorinda Chung, MD: Stanford University, Redwood City, California; [11]David Isenberg, MD: University College London, London, UK; [12]Andrew Mammen, MD, PhD: Johns Hopkins University School of Medicine, Baltimore, Maryland; [13]Marianne de Visser, MD, PhD: Academic Medical Center, Amsterdam, The Netherlands; [14]Steven R. Ytterberg, MD: Mayo Clinic, Rochester, Minnesota; [15]Katalin Danko, MD, PhD, DSc: University of Debrecen, Debrecen, Hungary; [16]Ignacio Garcia-De la Torre, MD: Hospital General de Occidente de la Secretaría de Salud and University of Guadalajara, Guadalajara, México; [17]Yeong Wook Song, MD, PhD: Graduate School of Convergence Science and Technology and Seoul National University Hospital, Seoul, Korea; [18]Jiri Vencovsky, MD, PhD: Charles University, Prague, Czech Republic. See Appendix A for members of the International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation who contributed to developing the response criteria.

Drs. Aggarwal and Rider contributed equally to this work. Drs. Miller and Vencovsky contributed equally to this work.

Address correspondence to Rohit Aggarwal, MD, MSc, Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh, 3601 5th Avenue, Suite 2B, Pittsburgh, PA 15261. E-mail: aggarwalr@upmc.edu.

Submitted for publication February 11, 2016; accepted in revised form January 31, 2017.

For these reasons, and with support from the American College of Rheumatology, European League Against Rheumatism, IMACS, and the Paediatric Rheumatology International Trials Organisation (PRINTO) (15), a collaboration was established to develop a data- and consensus-driven process involving multiple clinical data sets and the international myositis community in order to develop and validate response criteria for adult DM/PM and juvenile DM. This effort involved a comprehensive approach to developing candidate definitions for the response criteria, including continuous or hybrid definitions, using conjoint analysis (13,14,16–19), and for developing criteria for minimal as well as greater degrees of improvement. This article focuses on the criteria for minimal and moderate improvement for adult DM/PM, whereas the threshold for major improvement is considered preliminary. A companion article focuses on the juvenile DM response criteria (20).

Methods

Core set measures and patient profile consensus. To develop patient profiles as well as candidate definitions for response criteria in adult PM and DM, we used previously validated IMACS myositis core set measures for patients with adult DM/PM, which include physician and patient global activity on a 10-cm visual analog scale (VAS), muscle strength measured by manual muscle testing (MMT), physical function measured by the Health Assessment Questionnaire (HAQ) (21), extramuscular global activity measured by the physician on a 10-cm VAS, and the most abnormal serum muscle enzyme (8,22). The entire process, from the development of these profiles and candidate definitions through final consensus voting, is shown in the flow diagram in Figure 1 (23,24). Details of the methodology used to develop patient profiles, candidate definitions, validation, and expert consensus will be described in a separate publication (24). Briefly, patient data from natural history studies and uncontrolled clinical trials were used to develop patient profiles, which were then rated by adult myositis experts to achieve consensus as to whether improvement was none, minimal, moderate, or major. The expert consensus of improvement was used as the gold standard to validate various candidate definitions. The Bohan and Peter classification was used to designate definite or probable adult DM/PM (25).

Candidate definitions of response criteria. Six different types of candidate definitions for minimal, moderate, and major response (Table 1) were developed (23,26): 3 types of definitions were traditional (categorical), and 3 were continuous (hybrid). Traditional definitions provide only categorical outcomes of minimal, moderate, and major improvement, or not improved, based on the criteria, whereas continuous definitions yield an improvement score as a continuous outcome measure, with thresholds of minimal, moderate, and major improvement serving as categorical outcomes. Continuous definitions are considered hybrid definitions, because the same definition can be used as a continuous or categorical outcome measure based on the study requirements. Definitions utilizing either absolute percent change (final minus baseline divided by range and multiplied by 100) or

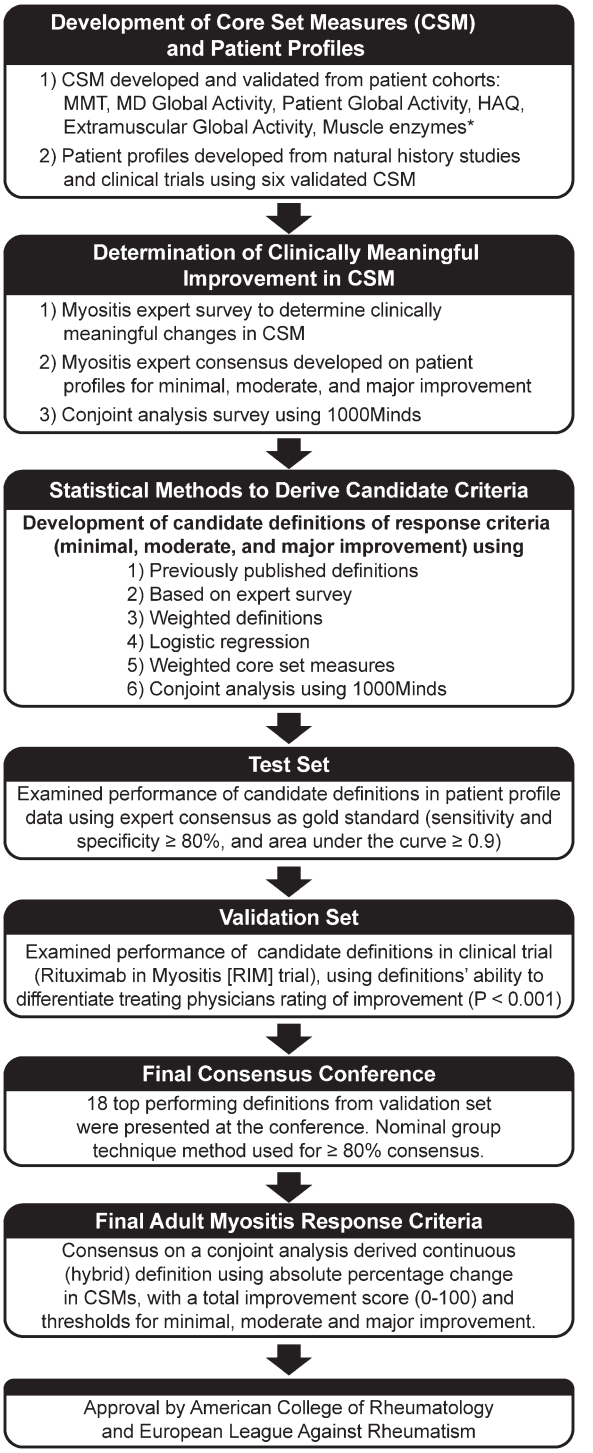

Figure 1. Flow diagram of the entire process used to develop and validate the approved response criteria for adult dermatomyositis and polymyositis.

Table 1. Types of candidate definitions for response criteria that were developed and tested*

Type of candidate definitions of response	Description	Example of candidate definition for the response criteria
Previously published (categorical definition)	Previously published definitions of improvement that were retested	Minimal. Three of any 6 improved by ≥20%, no more than 2 worse by >25% (which cannot be MMT) (9) Moderate. Three of any 6 improved by ≥50%, no more than 2 worse by >25% (which cannot be MMT) Major. Three of any 6 improved by ≥70%, no more than 2 worse by >25% (which cannot be MMT)
Newly drafted (categorical definition)	Drafted relative or absolute % change candidate definitions of response, based on recent CSM survey	Minimal. Two of any 6 improved by ≥30%, no more than 1 worse by >30% (which cannot be MMT) Moderate. Two of any 6 improved by ≥50%, no more than 1 worse by >30% (which cannot be MMT) Major. Two of any 6 improved by ≥75%, no more than 1 worse by >30% (which cannot be MMT)
Weighted (categorical definition)	Applied conjoint analysis relative weights to CSM in newly drafted definitions; each CSM receives improvement points (corresponding relative weights), when it reaches the threshold for minimal, moderate, or major improvement; worsening points are applied similarly; improvement is calculated based on a total score of improvement versus worsening	Improvement = at least 2.5 total improvement points of a maximum possible score of 8, and no more than 2.5 worsening points, where MD global = 1.5 points, patient global = 1 point, MMT = 2 points, HAQ = 1.5 points, extramusc = 1.5 points, enzyme = 0.5 point Minimal. Improvement points given when CSM ≥30%; worsening points given when CSM worse by >25% Moderate. Improvement points given when CSM ≥50%; worsening points given when CSM worse by >25% Major. Improvement points given when CSM ≥75%; worsening points given when CSM worse by >25%
Logistic regression (continuous definition)	Model of improvement using combination of CSM with different weights, as developed in the logistic regression model and rounded for better feasibility; total scores derived, with different cutoffs, for minimal, moderate, and major improvement	Improvement score = 5 × (MD global % change) + 3 × (patient global % change) + (MMT % change) + 2 × (HAQ % change) + 2 × (extramusc % change) + 2.5 × (enzyme % change) Minimal. Improvement score ≥250 Moderate. Improvement score ≥500 Major. Improvement score ≥750
Core set measure–weighted (continuous definition)	Multiply the % change in each CSM by the weights derived from conjoint analysis, then sum (% change in each CSM × conjoint analysis weights) to get final total improvement score; different thresholds for minimal, moderate, and major improvement established based on consensus profile ratings as gold standard	Improvement score = 2 × (MD global % change) + (patient global % change) + 3 × (MMT % change) + 1.5 × (HAQ % change) + 1.5 × (extramusc % change) + (enzyme % change) Minimal. Improvement score ≥100 Moderate. Improvement score ≥250 Major. Improvement score ≥400
Conjoint analysis (continuous definition)	For a given range in the level of improvement in each CSM, a score is assigned, as developed by the conjoint-analysis survey results and modeling; greater degrees of improvement receive higher scores; a patient is minimally improved if the improvement score is above the cutoff for minimal improvement; similarly, for moderate and major improvement	Cut points for the model are: Minimal. Improvement score ≥20 Moderate. Improvement score ≥40 Major. Improvement score ≥60†

* MMT = manual muscle testing; CSM = core set measure; MD global = physician global activity score; patient global = patient global activity score; HAQ = Health Assessment Questionnaire; extramusc = extramuscular global activity; enzyme = most abnormal serum muscle enzyme value among aldolase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, and creatine kinase.
† See Table 3 for cut points for the full model.

relative percent change (final minus baseline, divided by baseline and multiplied by 100) were evaluated as candidate definitions.

Conjoint analysis surveys. Conjoint analysis surveys were administered to myositis experts using 1000Minds online software (11). Experts were presented with pairs of hypothetical patient scenarios; each patient had different levels of improvement in the same 2 core set measures, assuming other core set measures remained the same. Experts rated which of the 2 scenarios had greater improvement. Based on the rater's response, all other hypothetical patients that could be pairwise ranked were eliminated via the property of transitivity, thereby significantly reducing the number of scenarios presented. The PAPRIKA (Potentially All Pairwise Rankings of All Possible Alternatives) method was used to determine the relative importance of the core set measures. Relative weights of core set measures and their levels of improvement were used to develop a scoring system by mathematical methods based on linear programming (13), such that when all 6 core set measures are considered together, the maximum score (total improvement score) possible for representing a patient's improvement is 100 and the minimum score is 0. The thresholds for minimal, moderate, and major improvement in the total improvement score were based on optimum sensitivity and specificity (using the Youden index [27]) in the subset of patient cohort data.

Validation of candidate response criteria. The performance characteristics of candidate criteria were evaluated using consensus profile ratings as the gold standard, assessing sensitivity, specificity, and area under the curve (AUC) to compare the performance of these candidate definitions. Those that performed well in the consensus profiles (sensitivity and specificity ≥80%, AUC ≥0.9 for minimal improvement, and AUC ≥0.8 for moderate and major improvement) were externally validated using data for adult DM/PM patients (n = 142) enrolled in the Rituximab in Myositis (RIM) trial (3). The treating physician's rating of improvement (0–7 scale) at 24 weeks in the RIM trial was used for validation, and a 1-point change in the physician's rating was considered clinically significant (3). We then selected the top candidate definitions (up to 4 top-performing definitions from each of the 6 different types of candidate definitions) for consideration at the final consensus conference, in order to discuss a manageable number of definitions at the conference.

Consensus conference. The nominal group technique (NGT) was applied to develop consensus among experts in adult DM/PM regarding the top-performing candidate definitions for minimal and moderate improvement in adult DM/PM (28–30). Experienced moderators (RA and FWM) led the NGT consensus-development process for the adult working group and the combined adult and pediatric working group (RA, LGR, NR, and FWM). Given the paucity of data on major improvement, we considered the major improvement thresholds as preliminary for the final consensus meeting. For each candidate definition, the methodologic details used to develop it and its performance characteristics in the consensus patient profiles and the RIM trial were presented to the adult working group. Each of the 12 participants in the adult working group independently reviewed the performance characteristics of all 18 top candidate definitions for adult DM/PM. Detailed data for each candidate definition, including sensitivity, specificity, and AUC as well as kappa values and odds ratios for minimal, moderate, and major improvement, were provided. The AUC was determined from the receiver operating characteristic curve as a plot of sensitivity versus (1 – specificity) for total improvement scores as well as for thresholds (27).

Adult working group. The primary goal for the adult working group was to develop consensus response criteria for minimal and moderate clinical improvement in adult DM/PM based on the data presented, as well as the face validity, feasibility, and generalizability of the proposed candidate criteria. The experts in the adult working group included internationally recognized rheumatologists, neurologists, and dermatologists who have considerable experience in myositis and with the core set measures. Voting was conducted in an independent, anonymous, and systematic manner via a web-based system developed by staff at the PRINTO coordinating center (31,32). In the initial rounds of voting, participants were asked to rank their top 5 choices. The results were compiled, and aggregate votes and rank of each candidate definition were shared with the group after each round of voting. Participants were then asked in a random manner to discuss their top-ranked and bottom-ranked choices. Candidate definitions receiving a small proportion of votes were eliminated. In subsequent voting rounds, participants were asked to re-rank their choices after reviewing the previous round's voting and discussion. When fewer than 5 candidate definitions remained, each participant selected one as the top response criteria. The objective was to continue the rounds of voting in the same manner until a single candidate definition reached consensus (≥80% of the votes) or until it was clear that consensus would not be reached.

Combined adult and pediatric working group. After consensus was achieved by each working group, both groups then came together to vote on common response criteria to be used for both adult DM/PM and juvenile DM (20) as the outcome measure for combined clinical trials. For this voting round, the top candidate definitions from the final round of voting in each working group were considered, and a similar online voting system and the NGT were used until consensus of ≥80% was reached (28–30). For determining the thresholds of improvement for the selected definition, the required consensus was ≥70%, which was done by post-conference voting.

Results

Candidate definitions. A total of 287 adult DM/PM candidate response criteria were drafted or derived using data-driven methods. Included were 10 previously published definitions, 134 newly drafted definitions based on expert survey results, 63 weighted definitions, 68 logistic regression definitions, 6 conjoint analysis definitions, and 6 definitions in which differential weights were applied to the improvement achieved in each core set measure. Among these definitions, 163 used relative percent change and 124 used absolute percent change in the core set measures.

Validation. Candidate definitions with a sensitivity and specificity of ≥80%, AUC ≥0.9 for minimal, and AUC ≥0.8 for moderate and major improvement in the patient profile analysis using expert consensus rating as the gold standard were evaluated for external validation using RIM clinical trial data (3) (see Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at

Table 2. Detailed performance characteristics of patient profiles and clinical trial data for the top 5 candidate response criteria definitions presented at the consensus conference*

Candidate definitions for response criteria, improvement category, core set measure	Profiles (n = 270)†				RIM trial (n = 147)			
	Sensitivity, %	Specificity, %	Threshold AUC	Total AUC	Candidate definition, improved physician's rating‡	Candidate definition, not improved physician's rating‡	*P*	Rank
Conjoint analysis absolute % change (model 3)§								1
Minimal (improvement score ≥20)	85	92	0.89	0.96	2.0	4.0	<0.001	
Moderate (improvement score ≥40)	90	96	0.93	0.99	2.0	3.0	<0.001	
Major (total improvement score ≥60)	92	98	0.95	1.00	2.0	3.0	<0.001	
Conjoint analysis relative % change (model 1)¶								2
Minimal (improvement score ≥33)	94	90	0.92	0.98	2.0	4.0	<0.001	
Moderate (improvement score ≥55)	93	93	0.93	0.99	2.0	3.0	<0.001	
Major (improvement score ≥70)	100	95	0.97	0.99	2.0	3.0	<0.001	
Conjoint analysis relative % change (model 2)¶								3
Minimal (improvement score ≥30)	94	92	0.93	0.98	2.0	4.0	<0.001	
Moderate (total improvement score ≥45)	94	88	0.91	0.98	2.0	3.0	<0.001	
Major (improvement score ≥65)	100	98	0.99	1.00	2.0	3.0	<0.001	
Weighted core set measure relative % change#								4
Minimal (improvement score ≥100)	92	91	0.91	0.97	2.0	3.0	<0.001	
Moderate (improvement score ≥250)	94	91	0.93	0.98	2.0	3.0	<0.001	
Major (improvement score ≥400)	100	94	0.97	1.00	2.0	3.0	<0.001	
Logistic regression relative % change**								5
Minimal (improvement score ≥75)	89	93	0.91	0.97	2.0	3.0	<0.001	
Moderate (improvement score ≥150)	94	88	0.91	0.98	2.0	3.0	<0.001	
Major (improvement score ≥300)	100	96	0.98	1.00	2.0	3.0	<0.001	

* Supplementary Table 2 (available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract) shows definitions 6–18 from the consensus conference ratings. The threshold area under the curve (AUC) was calculated as the AUC from the receiver operating characteristic (ROC) curve for the total improvement score and the threshold for minimal, moderate, and major improvement. The total AUC was calculated as the AUC from the ROC curve, using the total improvement score and the threshold cutoffs for minimal, moderate, and major improvement, and applies only to continuous definitions.

† The reference standard for sensitivity and specificity was myositis expert consensus rating of improvement.

‡ Physician's rating is the treating physician's rating on a Likert scale of 1–7, where lower scores represent a greater degree of improvement, at week 24 of the Rituximab in Myositis (RIM) trial (3). A 1-point difference in the physician's rating of improvement from no improvement to minimal improvement was considered not only statistically significant but also clinically significant.

§ Conjoint analysis–based continuous candidate response criteria using absolute percent change in core set measures (absolute percent change model) is shown in Table 3. These criteria are also the top response criteria for juvenile dermatomyositis (DM), but with different thresholds in the total improvement score for minimal, moderate, and major improvement (20).

¶ Conjoint analysis–based continuous candidate response criteria using relative percent change in core set measures are shown in Supplementary Table 3 (available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). These criteria are also the second- and third-choice criteria for juvenile DM, but with different thresholds in the total improvement score for minimal, moderate, and major improvement (20).

The total improvement score is calculated as 2 × (MD global % change) + (patient global % change) + 3 × (MMT % change) + 1.5 × (HAQ % change) + 1.5 × (extramusc % change) + (enzyme % change). (MD global = physician global activity; patient global = patient global activity; MMT = manual muscle testing; HAQ = Health Assessment Questionnaire; extramusc = extramuscular; enzyme = most abnormal serum muscle enzyme value among aldolase, alanine aminotransferase, aspartate aminotransferase, lactate hydrogenase, and creatine kinase.)

** The total improvement score is calculated as (MD global % change) + (patient global % change) + (MMT % change) + (HAQ % change) + (extramusc % change) + (enzyme % change).

http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). Thus, of 122 adult DM/PM candidate definitions evaluated using the RIM trial data, 36 adult DM/PM candidate definitions, including 25 using relative and 11 using absolute percent change in core set measures, had AUC ≥0.7 and showed validation in the clinical trial analysis.

Top candidate definitions. Of 36 validated definitions, 17 top-performing adult candidate definitions and the top pediatric response criteria (20) were considered by the adult working group at the consensus conference so that, in total, 18 candidate definitions were evaluated (Table 2 and Supplementary Table 2, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). They included 9 categorical definitions and 9 continuous definitions, in which 14 used relative percent change and 4 used absolute percent change in core set measures. In each categorical definition, a patient would either meet or not meet the response criteria of minimal, moderate, or major improvement based on the degree of improvement or worsening in each core set measure. In

the continuous definitions, however, each subject generates a total improvement score on a continuous scale, such that a greater degree of improvement corresponds to a higher score. Furthermore, patients could be categorized as achieving minimal, moderate, or major clinical improvement based on reaching the pre-set threshold score on the continuous scale. Table 2 shows the performance characteristics of the top 5 candidate definitions for the response criteria selected at the consensus conference (see Supplementary Table 2 for definitions 6–18).

In the patient profiles, with expert consensus as the gold standard, all top candidate definitions presented at the conference had excellent performance characteristics, with median sensitivity of 87% (interquartile range [IQR] 84–90%) and specificity of 94% (IQR 92–95%) for minimal improvement with a median AUC of 0.91 (IQR 0.90–0.92) (Table 2 and Supplementary Tables 1 and 2, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). Sensitivity, specificity, and AUC were similarly high for moderate and major improvement criteria for these definitions (Table 2 and Supplementary Tables 1 and 2). All candidate definitions presented at the conference were validated using the RIM trial data at the 24-week time point and were shown to differentiate ($P < 0.001$) between the treating physician's improvement score at week 24 in patients rated as improved versus not improved (3) (Table 2 and Supplementary Tables 1 and 2).

Consensus conference voting. The top-choice definition for the adult working group, which received 80% of the votes, was the conjoint analysis–based continuous definition model 1, which includes relative percent change in core set measures, including physician and patient global activity, muscle strength, physical function, most abnormal serum enzyme level, and extramuscular activity (Supplementary Table 3, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). The second-choice definition, receiving 20% of the votes, was the conjoint analysis–based continuous model 2, which also includes relative percent change in core set measures (see Supplementary Table 3). Models 1 and 2 differ only in the scores associated with each level of improvement in each core set measure.

However, in the final round of voting and discussion, adult working group participants reached unanimous consensus that the response criteria for adult DM/PM would be identical to the top-choice response criteria for juvenile DM, which is a conjoint analysis–based continuous definition (model 3) using absolute percent change in core set measures (Table 3) (20). Participants favored using the same response criteria for adult DM/PM and juvenile DM so that data from different studies can be harmonized more effectively and to facilitate combined trials, especially given that the definitions were similar with similar performance characteristics. Moreover, the absolute percent change in core set measures (model 3 [Table 3]) was thought to be more representative of meaningful clinical change compared with relative percent change in core set measures (models 1 and 2 [Supplementary Table 3]). Participants also voted to evaluate all top 5 candidate definitions from the adult working group in future clinical trials, with the other 4 as secondary outcome measures. The top 3 of these criteria, the conjoint analysis definitions, are the same for both adult DM/PM and juvenile DM, with different thresholds of improvement.

The sensitivity and specificity of the top-choice criteria, the conjoint analysis absolute percent change (Table 3), were 85% and 92% for minimal improvement, 90% and 96% for moderate improvement, and 92% and 98% for major improvement, respectively (Table 2). The AUC was 0.96 for the total improvement score and 0.89, 0.93, and 0.95 for minimal, moderate, and major improvement thresholds, respectively (Table 2). In the RIM trial (3), these response criteria showed a significant difference in the physician's rating of improvement when the response criteria rated the patient as improved versus not improved for minimal, moderate, and major improvement ($P < 0.001$) (Table 2 and Supplementary Table 2, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40064/abstract). Myositis experts in the consensus conference favored the conjoint analysis–based continuous response criteria because the total improvement score is a continuous measure that corresponds to the magnitude of improvement in a patient and provides the ability to categorize a patient's degree of improvement as minimal, moderate, or major (making it truly a hybrid definition). Moreover, the differential weights for various core set measures were also thought to be congruent with an expert's assessment of the relative importance of each core set measure. An important consideration in the final selection was that the top-choice definition be based on absolute percent change in the core set measures, which was favored by the participants because, given the various VAS measurements used, the absolute percent change was thought to be more representative of meaningful clinical change.

Top candidate definitions considered by the combined pediatric/adult working group. Three candidate definitions were considered by the combined adult/pediatric working group; these included the top adult definitions (see Supplementary Table 3) and the top pediatric definitions (20), one of which was identical in both groups. Final consensus was reached for the combined adult DM/PM and juvenile DM response criteria, with 91% of

Table 3. Final myositis response criteria for minimal, moderate, and major improvement in adult dermatomyositis/polymyositis (DM/PM) and combined adult DM/PM and juvenile DM clinical trials and studies*

Core set measure, level of improvement based on absolute percent change	Improvement score
Physician global activity	
Worsening to 5% improvement	0
>5% to 15% improvement	7.5
>15% to 25% improvement	15
>25% to 40% improvement	17.5
>40% improvement	20
Patient global activity	
Worsening to 5% improvement	0
>5% to 15% improvement	2.5
>15% to 25% improvement	5
>25% to 40% improvement	7.5
>40% improvement	10
Manual muscle testing	
Worsening to 2% improvement	0
>2% to 10% improvement	10
>10% to 20% improvement	20
>20% to 30% improvement	27.5
>30% improvement	32.5
Health Assessment Questionnaire	
Worsening to 5% improvement	0
>5% to 15% improvement	5
>15% to 25% improvement	7.5
>25% to 40% improvement	7.5
>40% improvement	10
Enzyme (most abnormal)	
Worsening to 5% improvement	0
>5% to 15% improvement	2.5
>15% to 25% improvement	5
>25% to 40% improvement	7.5
>40% improvement	7.5
Extramuscular activity	
Worsening to 5% improvement	0
>5% to 15% improvement	7.5
>15% to 25% improvement	12.5
>25% to 40% improvement	15
>40% improvement	20

The **total improvement score** is the sum of all 6 improvement scores associated with the change in each core set measure. A total improvement score of **≥20** represents **minimal improvement**, a score of **≥40** represents **moderate improvement**, and a score of **≥60** represents **major improvement**.

* Note that these response criteria are also proposed for use in combined adult DM/PM and juvenile DM trials (20). For comparison, the thresholds of improvement in the total improvement score for juvenile DM are ≥30 for minimal improvement, ≥45 for moderate improvement, and ≥70 for major improvement. Also note that the criteria for major improvement for adult DM/PM are preliminary.
How to calculate the improvement score: The absolute percent change ([final value – baseline value]/range × 100) is calculated for each core set measure. For muscle enzymes, the most abnormal serum muscle enzyme level at baseline (creatine kinase, aldolase, alanine transaminase, aspartate aminotransferase, lactate dehydrogenase) is used. The enzyme range was calculated based on a 90% range of enzymes from natural history data (34,46), which for creatine kinase is 15 times the upper limit of normal (ULN), for aldolase is 6 times the ULN, and for lactate dehydrogenase, aspartate aminotransferase, and alanine transaminase is 3 times the ULN. The ULN is determined according to the individual laboratories in the participating centers. The ranges for physician global activity, patient global activity, manual muscle testing, Health Assessment Questionnaire, and extramuscular global activity are based on the instrument scale used (3,26). An improvement score is assigned for each core set measure based on the absolute percent change in the core set measure according to the definition. These individual core set measure improvement scores are then totaled among the 6 core set measures to give the total improvement score. The thresholds for minimal, moderate, and major improvement are provided. The total improvement score itself may also be compared among treatment arms in a trial. A total improvement score between 0 and 100 corresponds to the degree of improvement, with higher scores corresponding to a greater degree of improvement.

participants voting for the conjoint analysis–based continuous definition, based on absolute percent change in the core set measure (Table 3). The combined working group agreed that the same final response criteria will be used for clinical trials of both adult DM/PM and juvenile DM, but with different thresholds for improvement in adult versus

pediatric patients as well as different core set measures for adult patients (IMACS) and pediatric patients (IMACS and PRINTO). Participants favored using the same response criteria for adult DM/PM and juvenile DM, because the top definition from each working group was very similar (i.e., both being conjoint analysis–based continuous models, with excellent and similar performance characteristics) and because it would permit comparison of outcomes in separate studies. Although only the IMACS core set measures were used for adult DM/PM, for further congruence with pediatric core set measures, the experts in adult myositis agreed to include the Short Form-36 (33) as a health-related quality-of-life measure to correspond to the PRINTO quality-of-life core set measure, the parent form of the Child Health Questionnaire (34–36). In a post-conference final vote, consensus (74%) was reached on threshold values for minimal, moderate, and major response for adult DM/PM patients, which are ≥20 in the total improvement score for minimal improvement, ≥40 for moderate improvement, and ≥60 for major improvement. In contrast, consensus on the final threshold values for minimal, moderate, and major response for juvenile DM were ≥30, ≥45, and ≥70 points, respectively.

Discussion

After a systematic data- and consensus-driven process, a conjoint analysis–based continuous (i.e., hybrid) definition based on absolute percent change in core set measures was selected as the response criteria for adult DM/PM for minimal and moderate improvement in future clinical trials and studies (Figure 1). Because the total number of cases in the trial data sets and clinical profiles that achieved major improvement was small, it was decided that the thresholds for major improvement would be considered preliminary. The same continuous (or hybrid) definition, but with different thresholds for minimal, moderate, and major improvement in IMACS or PRINTO core set measures, will be used for juvenile DM clinical trials and studies, as well as for combined adult DM/PM and juvenile DM studies and clinical trials in the future (20,24).

The process for developing and validating the candidate definitions for the response criteria was extensive and comprehensive, as we used large prospective clinical cohort data sets to develop patient profiles, and myositis expert consensus was used as the gold standard for clinical response. Consequently, we derived 6 different types of candidate definitions, each with many variations, leading to a total of 287 candidate definitions tested, which were validated using natural history cohorts and data from a randomized clinical trial. Subsequently, a representative number of international myositis experts from various disciplines (rheumatology, neurology, and dermatology) agreed on an innovative continuous (or hybrid) model using absolute percent change in validated core set measures.

These response criteria were developed using a novel conjoint analysis methodology, the 1000Minds software (13). Conjoint analysis, or discrete choice experiment, is a statistical technique used to determine expert group decision-making around various measures (and multiple levels within each measure), providing the ability to develop differential weighting of measures and composite criteria using those measures. The 1000Minds software for conjoint analysis has been used recently to develop rheumatologic classification and/or outcome criteria for rheumatoid arthritis (RA), systemic sclerosis (12,13,37,38), and gout (11,16,17,39).

The criteria developed are continuous in nature and generate a total improvement score (on a scale of 0–100), which can provide a quantitative degree of improvement for each patient rather than a dichotomous or categorical assessment of improvement. The total improvement score is the sum of the improvement reflected in each of the 6 core set measures, but the individual core set measures are weighted, such that those deemed more important provide a greater contribution to the final score. For example, changes in the MMT and physician global disease activity scores are weighted more heavily than changes in the most abnormal enzyme or the HAQ. These weights were consistent with our myositis expert survey (26), which was independent of the process used to develop and validate our response criteria.

There are significant advantages of using continuous response criteria (especially in pilot studies). For example, it might be possible to enroll fewer subjects and still have sufficient statistical power to differentiate between treatment groups by using the mean or median total improvement score. Moreover, continuous measures have the best sensitivity to change, the use of which allows modest treatment differences to be detected as statistically significant, which in turn leads to better clinical trials (10). Moreover, the criteria developed provide thresholds for both minimal and moderate improvement, with a preliminary threshold for major improvement. Therefore, larger, adequately powered clinical trials and studies can use the threshold of minimal clinically significant improvement to differentiate the treatment groups, because this difference will be considered *clinically* significant. Similarly, the proportions of patients achieving minimal or moderate improvement can be determined and compared between treatment arms. The ability of the same response criteria to be used not only as a continuous measure, where a higher score implies greater improvement, but also as a categorical response of minimal and moderate improvement, results in a unique hybrid aspect to these criteria.

Another advantage of continuous response criteria over the previous IMACS response criteria is that inclusion criteria for clinical trials will not require minimal severity in any core set measure, because all levels of improvement in each core set measure contribute more or less to the response. However, for each trial the investigators will have to determine the entry criteria for baseline core set measure abnormality, but those will depend on the effect size, disease or organ target, recruitment, and feasibility rather than on the response criteria alone. This is an improvement over the previous IMACS preliminary response criteria, where the clinical trial inclusion criteria required a baseline deficit of at least 20% in each core set measure to enable reaching the threshold of ≥20% improvement in core set measures after treatment.

Another important aspect of these response criteria is that they are based on an absolute percent change in core set measures rather than relative percent change, as used for scoring other rheumatologic diseases such as RA (40,41) and prior myositis response criteria (9). The panelists strongly believed that absolute percent change rather than relative percent change in core set measures more accurately reflects the degree of change. For example, for a patient in whom disease activity improved from 2 cm to 1 cm on a 10-cm VAS, this was interpreted by experts as more consistent with 10% improvement (absolute percent change) and not as 50% improvement reflected by relative percent change. Also, because many of the myositis core set measures arbitrarily have 0 as the lower limit of normal, using 10-cm VAS, the relative percent change is difficult to calculate if there is a change from 0 to a higher value.

The myositis experts decided to use similar response criteria for adult DM/PM and juvenile DM, to facilitate combined clinical trials, such as the RIM trial (3). Another advantage of the response criteria is that although they are the same for adult DM/PM and juvenile DM, they address the unique differences in the core set measure responsiveness between the 2 disease entities by specifying higher thresholds for juvenile DM than for adult DM/PM, which reflects the fact that more responsiveness is seen in juvenile DM patients in clinical trials (3,5). Additionally, the juvenile DM response criteria allow for the possibility of using the IMACS or PRINTO core set measures and provide a more definitive threshold for major improvement (20).

Some limitations of the new response criteria should be noted. First, most of the core set measures, although proven to have good reliability and validity, are subjective and evaluator dependent. However, similar metrics have been used successfully in RA trials that used a physician global measure similar to that used for myositis.

Second, only one major clinical trial was available for validation, and it failed to meet its primary end point and was not truly placebo controlled. Thus, we validated the results using the treating physician's improvement scores in the clinical trial.

Third, the threshold for major improvement in the response criteria is considered preliminary due to an insufficient number of adult DM/PM cases showing major improvement. We believe that future studies using therapeutic agents that have a greater impact on myositis disease activity will lead to better clinical responses, thus allowing investigators to determine a final threshold for major improvement. We plan to validate major improvement in future studies.

Fourth, given that the criteria are focused on improvement and thus fail to differentiate between no change and worsening, these criteria might not be applicable in studies of worsening disease activity (i.e., disease flare designs) in myositis. However, in the future, it will be necessary to develop criteria for flare in myositis.

Fifth, the response criteria were developed using a PM diagnosis based on the Bohan and Peter classification criteria, but experts now recognize that PM, according to those criteria, may include different syndromes, such as necrotizing myopathy, the antisynthetase syndrome, and others (42,43). We believe that these response criteria will still be applicable to these newer entities given that the data- and consensus-driven processes described herein were inclusive of those syndromes. In the future, with changes in classification criteria terminology (44), the response criteria terminology will need to be modified accordingly.

Sixth, because the criteria are complex and might be difficult to apply in research studies, we are developing a web-based tool as well as a downloadable calculator that will allow easy application of the response criteria. The time required to apply these criteria is estimated to be 25 minutes to complete the core set measures at each visit (6) and 3 minutes to hand-calculate the total improvement score and degree of response, while with a computer-based system the calculation time is negligible. Moreover, although the criteria may appear to be complicated, the core set measures to be collected by any study or investigators are simple and are essentially the same as those in previous myositis studies and trials.

Finally, patient-reported outcomes as core set measures, with the exception of the HAQ and patient global assessment, were not part of the response criteria, perhaps due to the paucity of sensitive and responsive patient-reported outcomes for DM/PM (45).

In conclusion, the development of data- and consensus-driven conjoint analysis–based continuous

response criteria with quantitative assessment of improvement on a scale of 0–100 and with thresholds for minimal, moderate, and major (preliminary threshold) improvement marks a major advancement in assessing response in myositis clinical trials and studies. These response criteria are sensitive and specific and provide a way to determine clinically meaningful change corresponding to degree of clinical improvement. These response criteria were valid in a clinical trial and had excellent face validity and acceptance among myositis experts from various specialties who care for adult DM/PM patients in different parts of the world. A conjoint analysis–based definition with a continuous improvement score using absolute percent change in core set measures with thresholds for minimal, moderate, and major improvement was selected as the response criteria to be used for adult clinical trials.

ACKNOWLEDGMENTS

We thank the following individuals for providing invaluable input and feedback on project development and support: members of the American College of Rheumatology Criteria Committee; Dr. Daniel Aletaha (European League Against Rheumatism), Drs. Suzette Peng and Sarah Yim (US Food and Drug Administration), Drs. Thorsten Vetter and Richard Vesely (European Medicines Agency), Bob Goldberg and Theresa Curry (The Myositis Association), Rhonda McKeever and Patti Lawler (Cure JM Foundation), and Irene Oakley (Myositis UK). We also thank Drs. Michael Ward, Steven Pavletic, and Adam Schiffenbauer for their critical review of the manuscript. Paul Hansen, who with Franz Ombler owns and co-invented the 1000Minds software referred to in the article, provided intellectual and logistic support for this project.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Aggarwal and Rider had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Aggarwal, Rider, Ruperto, Bayat, Erman, Feldman, Oddis, Amato, Chinoy, Cooper, Dastmalchi, Fiorentino, Isenberg, Katz, Mammen, de Visser, Ytterberg, Lundberg, Chung, Danko, Garcia-De la Torre, Song, Villa, Rinaldi, Rockette, Lachenbruch, Miller, Vencovsky.

Acquisition of data. Aggarwal, Rider, Ruperto, Bayat, Erman, Feldman, Oddis, Amato, Chinoy, Cooper, Dastmalchi, Fiorentino, Isenberg, Katz, Mammen, de Visser, Ytterberg, Lundberg, Chung, Danko, Garcia-De la Torre, Song, Villa, Rinaldi, Rockette, Lachenbruch, Miller, Vencovsky.

Analysis and interpretation of data. Aggarwal, Rider, Ruperto, Bayat, Erman, Feldman, Oddis, Amato, Chinoy, Cooper, Dastmalchi, Fiorentino, Isenberg, Katz, Mammen, de Visser, Ytterberg, Lundberg, Chung, Danko, Garcia-De la Torre, Song, Villa, Rinaldi, Rockette, Lachenbruch, Miller, Vencovsky.

REFERENCES

1. Rider LG, Miller FW. Deciphering the clinical presentations, pathogenesis, and treatment of the idiopathic inflammatory myopathies. JAMA 2011;305:183–90.
2. Moghadam-Kia S, Aggarwal R, Oddis CV. Treatment of inflammatory myopathy: emerging therapies and therapeutic targets. Expert Rev Clin Immunol 2015;11:1265–75.
3. Oddis CV, Reed AM, Aggarwal R, Rider LG, Ascherman DP, Levesque MC, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. Arthritis Rheum 2013;65:314–24.
4. Muscle Study Group. A randomized, pilot trial of etanercept in dermatomyositis. Ann Neurol 2011;70:427–36.
5. Ruperto N, Pistorio A, Oliveira S, Zulian F, Cuttica R, Ravelli A, et al. Prednisone versus prednisone plus ciclosporin versus prednisone plus methotrexate in new-onset juvenile dermatomyositis: a randomised trial. Lancet 2016;387:671–8.
6. Rider LG, Werth VP, Huber AM, Alexanderson H, Rao AP, Ruperto N, et al. Measures of adult and juvenile dermatomyositis, polymyositis, and inclusion body myositis: Physician and Patient/Parent Global Activity, Manual Muscle Testing (MMT), Health Assessment Questionnaire (HAQ)/Childhood Health Assessment Questionnaire (C-HAQ), Childhood Myositis Assessment Scale (CMAS), Myositis Disease Activity Assessment Tool (MDAAT), Disease Activity Score (DAS), Short Form 36 (SF-36), Child Health Questionnaire (CHQ), Physician Global Damage, Myositis Damage Index (MDI), Quantitative Muscle Testing (QMT), Myositis Functional Index-2 (FI-2), Myositis Activities Profile (MAP), Inclusion Body Myositis Functional Rating Scale (IBMFRS), Cutaneous Dermatomyositis Disease Area and Severity Index (CDASI), Cutaneous Assessment Tool (CAT), Dermatomyositis Skin Severity Index (DSSI), Skindex, and Dermatology Life Quality Index (DLQI). Arthritis Care Res (Hoboken) 2011;63 Suppl 11:S118–57.
7. Miller FW, Rider LG, Chung YL, Cooper R, Danko K, Farewell V, et al. Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. Rheumatology (Oxford) 2001;40:1262–73.
8. Rider LG, Giannini EH, Harris-Love M, Joe G, Isenberg D, Pilkington C, et al. Defining clinical improvement in adult and juvenile myositis. J Rheumatol 2003;30:603–17.
9. Rider LG, Giannini EH, Brunner HI, Ruperto N, James-Newton L, Reed AM, et al. International consensus on preliminary definitions of improvement in adult and juvenile myositis. Arthritis Rheum 2004;50:2281–90.
10. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. Arthritis Rheum 2007;57:193–202.
11. De Lautour H, Taylor WJ, Adebajo A, Alten R, Burgos-Vargas R, Chapman P, et al. Development of preliminary remission criteria for gout using Delphi and 1000Minds consensus exercises. Arthritis Care Res (Hoboken) 2016;68:667–72.
12. Johnson SR, Naden RP, Fransen J, van Den Hoogen F, Pope JE, Baron M, et al. Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. J Clin Epidemiol 2014;67:706–14.
13. Hansen P, Ombler F. A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives. J Multi-Crit Decis Anal 2008;15:87–107.
14. Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using discrete choice experiments to value health and health care. Dordrecht: Springer; 2008.
15. Ruperto N, Martini A. Networking in paediatrics: the example of the Paediatric Rheumatology International Trials Organisation (PRINTO). Arch Dis Child 2011;96:596–601.
16. Taylor WJ, Singh JA, Saag KG, Dalbeth N, MacDonald PA, Edwards NL, et al. Bringing it all together: a novel approach to the development of response criteria for chronic gout clinical trials. J Rheumatol 2011;38:1467–70.
17. Taylor WJ, Brown M, Aati O, Weatherall M, Dalbeth N. Do patient preferences for core outcome domains for chronic gout

studies support the validity of composite response criteria? Arthritis Care Res (Hoboken) 2013;65:1259–64.
18. Utz KS, Hoog J, Wentrup A, Berg S, Lammer A, Jainsch B, et al. Patient preferences for disease-modifying drugs in multiple sclerosis therapy: a choice-based conjoint analysis. Ther Adv Neurol Disord 2014;7:263–75.
19. De Bekker-Grob E, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ 2012;21:145–72.
20. Rider LG, Aggarwal R, Pistorio A, Bayat N, Erman B, Feldman BM, et al. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate and major clinical response for juvenile dermatomyositis: an International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation collaborative initiative. Arthritis Rheumatol doi: http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract. E-pub ahead of print.
21. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
22. Rider LG. Outcome assessment in the adult and juvenile idiopathic inflammatory myopathies. Rheum Dis Clin North Am 2002;28:935–77.
23. Aggarwal R, Rider LG, Ruperto N, Bayat N, Erman B, Feldman BM, et al. A consensus hybrid definition using a conjoint analysis is the proposed response criteria for minimal and moderate improvement for adult polymyositis and dermatomyositis clinical trials [abstract]. Arthritis Rheumatol 2014;66 Suppl:S404.
24. Rider LG, Ruperto N, Pistorio A, Erman B, Bayat N, Lachenbruch PA, et al. 2016 development of adult dermatomyositis and polymyositis and juvenile dermatomyositis response criteria: methodological aspects: an American College of Rheumatology/European League Against Rheumatism/International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation collaborative initiative. Rheumatology (Oxford). In press.
25. Bohan A, Peter JB. Polymyositis and dermatomyositis. Parts 1 and 2. N Engl J Med 1975;292:344–7, 403–7.
26. Rider LG, Lee J, Jansen A, Ruperto N, Huber AM, Oddis CV, et al. Defining clinically relevant changes in core set activity measures for adult and juvenile idiopathic inflammatory myopathies (IIM) [abstract]. Arthritis Rheum 2011;63 (Suppl):S89.
27. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283–98.
28. Ruperto N, Meiorin S, Iusan SM, Ravelli A, Pistorio A, Martini A. Consensus procedures and their role in pediatric rheumatology. Curr Rheumatol Rep 2008;10:142–6.
29. Delbecq A, van de Ven A, Gustafson D. Group techniques for program planning: a guide to nominal group and Delphi processes. Glenview (IL): Scott, Foresman and Company; 1975.
30. Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. Semin Arthritis Rheum 2011;41:95–105.
31. Ruperto N, Ozen S, Pistorio A, Dolezalova P, Brogan P, Cabral DA, et al. EULAR/PRINTO/PRES criteria for Henoch-Schonlein purpura, childhood polyarteritis nodosa, childhood Wegener granulomatosis and childhood Takayasu arteritis: Ankara 2008. Part I: overall methodology and clinical characterisation. Ann Rheum Dis 2010;69:790–7.
32. Piram M, Kone-Paut I, Lachmann HJ, Frenkel J, Ozen S, Kuemmerle-Deschner J, et al. Validation of the auto-inflammatory diseases activity index (AIDAI) for hereditary recurrent fever syndromes. Ann Rheum Dis 2014;73:2168–73.
33. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 health survey: manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
34. Ruperto N, Ravelli A, Pistorio A, Ferriani V, Calvo I, Ganser G, et al. The provisional Paediatric Rheumatology International Trials Organisation/American College of Rheumatology/European League Against Rheumatism Disease Activity Core Set for the Evaluation of Response to Therapy in Juvenile Dermatomyositis: a prospective validation study. Arthritis Rheum 2008;59:4–13.
35. Ruperto N, Ravelli A, Pistorio A, Malattia C, Cavuto S, Gado-West L, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries: review of the general methodology. Clin Exp Rheumatol 2001;19 Suppl 23:S1–9.
36. Apaz MT, Saad-Magalhaes C, Pistorio A, Ravelli A, de Oliveira Sato J, Marcantoni MB, et al. Health-related quality of life of patients with juvenile dermatomyositis: results from the Paediatric Rheumatology International Trials Organisation Multinational Quality of Life cohort study. Arthritis Rheum 2009;61:509–17.
37. Van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2013;65:2737–47.
38. Neogi T, Aletaha D, Silman AJ, Naden RL, Felson DT, Aggarwal R, et al. The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: phase 2 methodological report. Arthritis Rheum 2010;62:2582–91.
39. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheumatol 2015;67:2557–68.
40. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727–35.
41. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum 2011;63:573–86.
42. Aggarwal R, Cassidy E, Fertig N, Koontz DC, Lucas M, Ascherman DP, et al. Patients with non-Jo-1 anti-tRNA-synthetase autoantibodies have worse survival than Jo-1 positive patients. Ann Rheum Dis 2014;73:227–32.
43. Hengstman GJ, ter Laak HJ, Vree Egberts WT, Lundberg IE, Moutsopoulos HM, Vencovsky J, et al. Anti-signal recognition particle autoantibodies: marker of a necrotising myopathy. Ann Rheum Dis 2006;65:1635–8.
44. Tjarnlund A, Bottai M, Rider LG, Werth VP, Pilkington CA, de Visser M, et al. Progress report on development of classification criteria for adult and juvenile idiopathic inflammatory myopathies [abstract]. Arthritis Rheum 2012;64 Suppl:S323–4.
45. Alexanderson H, del Grande M, Bingham CO III, Orbai AM, Sarver C, Clegg-Smith K, 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;41:581–92.
46. Volochayev R, Csako G, Wesley R, Rider LG, Miller FW. Laboratory test abnormalities are common in polymyositis and dermatomyositis and differ among clinical and demographic groups. Open Rheumatol J 2012;6:54–63.

APPENDIX A: MEMBERS OF THE INTERNATIONAL MYOSITIS ASSESSMENT AND CLINICAL STUDIES GROUP AND THE PAEDIATRIC RHEUMATOLOGY INTERNATIONAL TRIALS ORGANISATION WHO CONTRIBUTED TO DEVELOPING THE RESPONSE CRITERIA

Steering committee: Lisa G. Rider (co–principal investigator), Nicolino Ruperto (co–principal investigator), Rohit Aggarwal (methodology lead), Frederick W. Miller, Jiri Vencovsky.

Statistical team: Rohit Aggarwal, Brian Erman, Nastaran Bayat, Angela Pistorio, Adam M. Huber, Brian M. Feldman, Paul

Hansen, Howard Rockette, Peter A. Lachenbruch, Nicolino Ruperto, Lisa G. Rider.

Adult core set survey group: Anthony A. Amato, Hector Chinoy, Lisa Christopher-Stine, Lorinda Chung, Robert G. Cooper, Lisa Criscione-Schreiber, Leslie Crofford, Mary E. Cronin, Katalin Dankó, David Fiorentino, Ignacio García-De la Torre, Patrick Gordon, Gerald Hengstman, James D. Katz, Andrew Mammen, Galina Marder, Neil McHugh, Chester V. Oddis, Elena Schiopu, Albert Selva-O'Callaghan, Yeong Wook Song, Jiri Vencovsky, Gil Wolfe, Robert Wortmann.

Clinical trial or natural history study data set contributions: Anthony A. Amato, Hector Chinoy, Lorinda Chung, Robert G. Cooper, Katalin Dankó, David Fiorentino, Ignacio García-De la Torre, Mark Gourley, Ingrid Lundberg, Frederick W. Miller, Chester V. Oddis, Paul Plotz, Lisa G. Rider, Yeong Wook Song, Jiri Vencovsky.

Adult patient profile working group: Rohit Aggarwal, Anthony A. Amato, Dana Ascherman, Richard Barohn, Olivier Benveniste, Jan De Bleecker, Jeffrey Callen, Christina Charles-Schoeman, Hector Chinoy, Lisa Christopher-Stine, Lorinda Chung, Robert G. Cooper, Leslie Crofford, Mary E. Cronin, Katalin Dankó, Sonye Danoff, Maryam Dastmalchi, Marianne de Visser, Mazen Dimachkie, Steve DiMartino, Lyubomir Dourmishev, Floranne Ernste, David Fiorentino, Ignacio García-De la Torre, Takahisa Gono, Patrick Gordon, Mark Gourley, David Isenberg, Yasuhiro Katsumata, James D. Katz, John Kissel, Richard L. Leff, Todd Levine, Ingrid Lundberg, Andrew Mammen, Herman Mann, Galina Marder, Isabelle Marie, Neil McHugh, Joseph Merola, Frederick W. Miller, Chester V. Oddis, Marzena Olesinska, Nancy Olsen, Nicolo Pipitone, Sindhu Ramchandren, Seward Rutkove, Lesley Ann Saketkoo, Adam Schiffenbauer, Albert Selva-O'Callaghan, Samuel Katsuyuki Shinjo, Rachel Shupak, Yeong Wook Song, Katarzyna Swierkocka, Jiri Vencovsky, Julia Wanschitz, Victoria Werth, Irene Whitt, Robert Wortmann, Steven R. Ytterberg.

Conjoint analysis, adult group: Rohit Aggarwal, Anthony A. Amato, Hector Chinoy, Lisa Christopher-Stine, Lorinda Chung, Robert G. Cooper, Mary E. Cronin, Katalin Dankó, Mazen Dimachkie, Steve DiMartino, David Fiorentino, Ignacio García-De la Torre, Patrick Gordon, Ingrid Lundberg, Herman Mann, Frederick W. Miller, Chester V. Oddis, Albert Selva-O'Callaghan, Jiri Vencovsky, Victoria Werth, Robert Wortmann, Steven R. Ytterberg.

Participants in consensus conference, adult working group: Anthony A. Amato, Hector Chinoy, Robert G. Cooper, Maryam Dastmalchi, Marianne de Visser, David Fiorentino, David Isenberg, James D. Katz, Andrew Mammen, Chester V. Oddis, Jiri Vencovsky, Steven R. Ytterberg.

Participants in consensus conference, pediatric working group: Rolando Cimaz, Rubén Cuttica, Sheila Knupp Feitosa de Oliveira, Brian M. Feldman, Adam M. Huber, Carol B. Lindsley, Clarissa Pilkington, Marilynn Punaro, Angelo Ravelli, Ann Reed, Kelly Rouster-Stevens, Annet van Royen-Kerkhof.