Published in final edited form as:

Ann Rheum Dis. 2017 May ; 76(5): 782–791. doi:10.1136/annrheumdis-2017-211401.

2016 American College of Rheumatology (ACR) – European League Against Rheumatism (EULAR) Criteria for Minimal, Moderate and Major Clinical Response for Adult Dermatomyositis and Polymyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative

Rohit Aggarwal, MD, MSc[*,1], **Lisa G. Rider, MD**[*,2], **Nicolino Ruperto, MD, MPH**[3], **Nastaran Bayat, MD**[2], **Brian Erman, MS**[4], **Brian M. Feldman, MD, MSc, FRCPC**[5], **Chester V. Oddis, MD**[1], **Anthony A. Amato, MD**[6], **Hector Chinoy, PhD, MRCP**[7], **Robert G. Cooper, MD**[8], **Maryam Dastmalchi, MD, PhD**[9], **David Fiorentino, MD, PhD**[10], **David Isenberg, MD**[11], **James D. Katz, MD**[12], **Andrew Mammen, MD, PhD**[13], **Marianne de Visser, MD, PhD**[14], **Steven R. Ytterberg, MD**[15], **Ingrid E. Lundberg, MD, PhD**[9], **Lorinda Chung, MD**[10], **Katalin Danko, MD,**

Address correspondence to: Dr. Rohit Aggarwal, Rohit Aggarwal, M.D., M.Sc., Associate Professor of Medicine, Division of Rheumatology and Clinical Immunology, Department of Medicine, University of Pittsburgh, 3601 5th Avenue, Suite 2B Pittsburgh, PA 15261, USA, Phone: 412-383-8100 Fax: 412-383-8864, aggarwalr@upmc.edu.

*co-first

**co-last authors

[20]Appendix of contributors:

Steering Committee: Lisa G. Rider (Co-Principal Investigator), Nicolino Ruperto (Co-Principal Investigator), Rohit Aggarwal (Methodology Lead), Frederick 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, David Fiorentino, Ignacio Garcia De la Torre, Katalin Dankó, 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 Garcia 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, Ignacio Garcia-De la Torre, Mazen Dimachkie, Steve DiMartino, Lyubomir Dourmishev, Floranne Ernste, David Fiorentino, 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 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, Marianne de Visser, 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 Di Martino, David Fiorentino, Ignacio Garcia-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, David Fiorentino, David Isenberg, James D. Katz, Andrew Mammen, Chester V. Oddis, Jiri Vencovsky, Marianne de Visser, Steven R. Ytterberg

Participants in Consensus Conference, Pediatric Working Group: Rolando Cimaz, Rubén Cuttica, Brian M. Feldman, Adam M. Huber, Carol B. Lindsley, Sheila Knupp Feitosa de Oliveira, Clarissa Pilkington, Marilynn Punaro, Angelo Ravelli, Ann Reed, Kelly Rouster-Stevens, Annet van Royen-Kerkhof

PhD, DSc[16], **Ignacio Garcia-De la Torre, MD**[17], **Yeong Wook Song, MD, PhD**[18], **Luca Villa**[3], **Mariangela Rinaldi**[3], **Howard Rockette, PhD**[1], **Peter A. Lachenbruch, PhD**[2], **Frederick W. Miller, MD, PhD**[**,2], and **Jiri Vencovsky, MD, PhD**[**,19] **for the International Myositis Assessment and Clinical Studies Group (IMACS) and the Paediatric Rheumatology Collaborative Study Group (PRINTO)**[20]

[1]Department of Medicine, Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA [2]Environmental Autoimmunity Group, NIEHS, NIH, Bethesda, MD [3]Istituto Giannina Gaslini, Pediatria II, PRINTO, Genoa, Italy [4]Social and Scientific Systems, Inc., Durham, NC [5]The Hospital for Sick Children, Toronto, Ontario, Canada [6]Brigham and Women's Hospital and Harvard Medical School, Boston, MA [7]National Institute of Health Research Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, University of Manchester, Manchester, United Kingdom [8]Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, United Kingdom [9]Rheumatology Unit, Department of Medicine, Solna, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden [10]Stanford University, Redwood City, CA [11]University College London, London, United Kingdom [12]NIAMS, NIH, Bethesda, MD [13]Johns Hopkins University School of Medicine, Baltimore, MD [14]Academic Medical Center, Amsterdam, The Netherlands [15]Mayo Clinic, Rochester, MN [16]University of Debrecen, Debrecen, Hungary [17]Hospital General de Occidente de la Secretaría de Salud, and University of Guadalajara, Guadalajara, Jal, México [18]Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology, and College of Medicine, Medical Research Center, Seoul National University Hospital, Seoul, Korea [19]Institute of Rheumatology and Department of Rheumatology, 1st Medical Faculty, Charles University, Prague, Czech Republic

Abstract

Objective—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 (CSM). Myositis experts rated greater improvement among multiple pair-wise scenarios in conjoint analysis surveys, where different levels of improvement in two CSM were presented. The PAPRIKA (Potentially All Pairwise Rankings of All Possible Alternatives) method determined relative weights of CSM and conjoint analysis definitions. Performance characteristics of definitions were evaluated on patient profiles using expert consensus (gold standard) and were validated using a clinical trial. Nominal group technique was used for consensus.

Results—Consensus was reached for a conjoint analysis–based continuous model using absolute percentage change in CSMs (physician, patient, and extra-muscular global activity, muscle strength, health assessment questionnaire and muscle enzymes). A Total Improvement Score (0–100), determined by summing scores in each CSM, was based on the improvement and relative weight of each CSM. 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 90% and 96% for minimal, moderate, and major improvement,

respectively. Definitions were validated in trial analysis for differentiating the physician rating of improvement ($P<0.001$).

Conclusion—The response criteria for adult DM/PM was the conjoint analysis model based on absolute percentage change in six CSMs, with thresholds for minimal, moderate, and major improvement.

Keywords

adult; dermatomyositis; polymyositis; response criteria; conjoint analysis; definitions of improvement; hybrid or continuous definition; outcome criteria; consensus

Idiopathic inflammatory myopathies are a group of acquired, heterogeneous, systemic connective tissue diseases that include adult dermatomyositis (DM) and polymyositis (PM) and juvenile DM (JDM) (1). Despite significant morbidity and mortality associated with DM/PM, there are currently no therapies approved for these syndromes by the United States Food and Drug Administration or European Medicines Agency based on randomized controlled trials. However, with the advancement in novel therapeutics that target various biological 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. Recent clinical trials completed in adult DM/PM and JDM have utilized 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 (CSM) 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). They 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 CSM adequately covers all the domains in myositis. For example, muscle enzymes can be normal in active DM, and active muscle weakness in DM can occur without active rash.

Preliminary response criteria had been developed and partially validated by IMACS for adult DM/PM; they were based on at least 20% improvement in three of six CSM with no more than two CSM worse by at least 25%, with muscle strength not allowed to worsen (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, too, including equal weights being applied to each CSM 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. For these reasons, and with support from the American College of Rheumatology, European League Against Rheumatism, IMACS, and the Paediatric Rheumatology International Trials Organization (PRINTO)(15), a collaboration was established to develop a data- and consensus-driven process involving multiple clinical datasets 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

for 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 major improvement is considered preliminary. A companion article focuses on the JDM 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 CSM for patients with adult DM/PM, which include Physician and Patient Global Activity on a 10-cm Visual Analogue Scale (VAS), muscle strength measured by manual muscle testing (MMT), physical function measured by the Health Assessment Questionnaire (HAQ), Extramuscular Global Activity measured by the physician on a 10-cm VAS, and the most abnormal serum muscle enzyme (8;21). The entire process, from the development of these profiles and candidate definitions through final consensus voting, is represented in the flow diagram in Figure 1 (22;23). Detailed methodology used to develop patient profiles, candidate definitions, validation, and expert consensus will be described in a separate publication (23). Briefly, real patient data from natural history studies and uncontrolled clinical trials were utilized 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. Definite or probable criteria of Bohan and Peter classification were used to designate adult PM/DM (24).

Candidate definitions of response criteria

Six different types of candidate definitions for minimal, moderate, and major response (Table 1) were developed (22;25): three types of definitions were traditional (categorical), and three 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 a continuous or categorical outcome measure based on the study requirements. Definitions utilizing either absolute percentage change (final minus baseline divided by range and multiplied by 100) or relative percentage 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 two CSM, assuming other CSM remained the same. Experts rated which of the two 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 determined the relative importance of the CSMs. Relative weights of CSMs and their levels of improvement were used to develop a scoring system by mathematical methods based on linear programming (13), such that when all six CSMs 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 (26)] in the subset of patient cohort data.

Validation of candidate response criteria

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%, area under the curve (AUC) ≥ 0.9 for minimal, and AUC ≥ 0.8 for moderate and major improvement) were externally validated using data from adult DM/PM subjects (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 physician rating was considered clinically significant (3). We then selected the top candidate definitions—up to four top-performing definitions from each of the six 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

Nominal group technique (NGT) was applied to develop consensus among adult DM/PM experts regarding the top-performing candidate definitions for minimal and moderate improvement in adult DM/PM (27–29). Experienced moderators (Drs. Aggarwal and Miller) led the NGT consensus for the adult working group and the combined adult and pediatric working group (Drs. Aggarwal, Miller, Ruperto, and Rider). 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 them and their performance characteristics in the consensus patient profiles and the RIM trial were presented to the adult working group (3). 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 and odds ratio for minimal, moderate, and major improvement, were provided. 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 (26).

Adult working group

The primary goal for the adult working group was to develop consensus response criteria for minimal and moderate clinical improvement for 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 CSM. Voting was conducted in an independent, anonymous, and systematic fashion via a web-based system developed by the PRINTO coordinating center (30;31). In initial rounds of voting, participants were asked to rank their top five 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 fashion to discuss their top- 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 five candidate definitions remained, each participant selected one as their 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 a common response criteria to be used for both adult DM/PM and JDM (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 was used until consensus ≥80% was reached (27–29). For determining the thresholds of improvement for the selected definition, the required consensus was ≥70%, which was done by post-conference voting.

RESULTS

Candidate definition for response criteria

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 CSM. Among these definitions, 163 used relative percent change and 124 used absolute percent change in CSM.

Validation

Candidate definitions with a sensitivity and specificity ≥ 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 the RIM clinical trial data (3) (Supplementary Table 1). Thus, of 122 adult DM/PM candidate definitions evaluated using the RIM trial data (3), 36 adult DM/PM candidate definitions, including 25 using relative and 11 using absolute percent change in CSM, 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) (20). They included nine categorical definitions and nine continuous definitions, in which 14 used relative percent change and four used absolute percent change in CSM. 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 CSM. 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 provides the performance characteristics of the top five 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 (interquartile range) sensitivity of 87% (84–90%) and specificity of 94% (92–95%) for minimal improvement with median AUC of 0.91 (0.90–0.92) (Table 2 and Supplementary Tables 1 and 2). 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 CSM, including physician and patient global activity, muscle strength, physical function, most abnormal serum enzyme level, and extra-muscular activity (Supplementary Table 3). The second-choice definition, receiving 20% of the votes, was the conjoint analysis–based continuous model 2, which also includes relative percent change in CSM (Supplementary Table 3). Models 1 and 2 differ only in the scores associated with each level of improvement in each CSM. 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 JDM, which is a conjoint analysis–based continuous definition (model 3) using absolute percent change in CSM (Table 3) (20). Participants favored using the same response criteria for adult DM/PM and JDM so that data from different studies can be harmonized more effectively and facilitate combined trials, especially given that the definitions were similar with similar performance characteristics. Moreover, the absolute percent change in CSMs (model 3) was thought to be more representative of meaningful clinical change than relative percent change in CSMs (models 1 and 2). Participants also voted to evaluate all top five candidate definitions from the adult working group in future clinical trials, with the other four as secondary outcome

measures. The top three of these criteria, the conjoint analysis definitions, are the same for both adult DM/PM and JDM, with different thresholds of improvement.

For the top conjoint-analysis, absolute percent change continuous definition (Table 3), the sensitivity and specificity in the patient profiles 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), this response criteria showed a significant difference in the physician 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). Myositis experts 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 CSM were also thought to be congruent with an expert's assessment of the relative importance of each CSM. An important consideration in the final selection was that the top-choice definition be based on absolute percent change in the CSM, which was favored by the participants because, given the various visual analogue scale measurements used, the absolute percent change was thought to be more representative of meaningful clinical change.

Combined pediatric-adult working group

Three candidate definitions were considered by the combined adult/pediatric working group; they included the top adult definitions (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 JDM response criteria, with 91% of participants voting for the conjoint analysis–based continuous definition, based on absolute percent change in the CSM (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 JDM, but with different thresholds for improvement in adult versus pediatric patients as well as different CSM for adult (IMACS) and pediatric patients (IMACS and PRINTO). Participants favored using the same response criteria for adult DM/PM and JDM 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 CSM were used for adult DM/PM, for further congruence with pediatric CSM, the adult myositis experts agreed to include the Short Form-36 as a health-related quality-of-life measure to correspond to the PRINTO quality-of-life CSM, the parent form of the Child Health Questionnaire (32–34). 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 thresholds for minimal, moderate and major response for JDM was ≥30, ≥45, and ≥70 points, respectively.

DISCUSSION

After a systematic data- and consensus-driven process, a conjoint analysis–based continuous (i.e., hybrid) definition with absolute percent change in CSM 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 datasets 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 CSM will be used for JDM clinical trials and studies, as well as for combined adult DM/PM and JDM studies and clinical trials in the future (20;23).

The process for developing and validating the candidate definitions for the response criteria was extensive and comprehensive, as we used large prospective clinical cohort datasets to develop patient profiles, and myositis expert consensus was used as the gold standard for clinical response. Consequently, we derived six 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 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 CSM.

This response criteria was developed using a novel conjoint-analysis methodology—the 1000Minds software (13). Conjoint analysis, or discrete choice experiment, is a statistical technique 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. 1000Minds software for conjoint analysis has been used recently to develop rheumatologic classification and/or outcome criteria for rheumatoid arthritis, systemic sclerosis (12;13;35;36), and gout (11;16;17;37).

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 subject, rather than a dichotomous or categorical assessment of improvement. The Total Improvement Score is the sum of the improvement reflected in each of the six CSM, but the individual CSM 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 HAQ. These weights were consistent with our myositis expert survey (25), 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, 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, as this difference will be considered *clinically* significant. Similarly, 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 this criteria. Another advantage of continuous response criteria over the previous IMACS response criteria is that inclusion criteria for clinical trials will not require a minimal severity in any CSM, because all levels of improvement in each CSM contribute more or less to the response. However, each trial will have to determine its own entry criteria of baseline CSM abnormality, but those will depend on the effect size, disease or organ target, recruitment, and feasibility—not 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 CSM to enable reaching the threshold of ≥20% improvement in CSMs after treatment.

Another important aspect of this response criteria is that it is based on an absolute percent change in CSM rather than relative percent change, as used for scoring other rheumatologic diseases, such as rheumatoid arthritis (38;39) and prior myositis response criteria (9). The panelists felt strongly that absolute percent change rather than relative percent change in CSM more accurately reflects the degree of change. For example, in a subject with improvement of disease activity from 2 cm to 1 cm on a 10-cm VAS, this was interpreted by experts as more consistent with a 10% improvement (absolute percent change) and not as 50% improvement reflected by relative percent change. Also, because many of the myositis CSM arbitrarily have zero as the lower limit of normal, using 10-cm VAS scales, 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 a common response criteria for adult DM/PM and JDM, to facilitate combined clinical trials, such as the RIM trial (3). Another advantage of the response criteria is that although it is the same for adult DM/PM and JDM, it addresses the unique differences in the CSM responsiveness between the two disease entities by specifying higher thresholds for JDM than for adult DM/PM, which reflect the fact that more responsiveness is seen in JDM patients in clinical trials (3;5). Additionally, the JDM response criteria allows for the possibility of using the IMACS or PRINTO CSM and provides a more definitive threshold for major improvement (20).

Some limitations of the new response criteria should be noted. First, most of the CSM, although proven to have good reliability and validity, are subjective and evaluator dependent. However, similar metrics have been used successfully in rheumatoid arthritis trials, which used a physician global measure similar to that employed for myositis. Second, only one major clinical trial was available for validation, and it failed to meet its primary endpoint and was not truly placebo controlled. Thus, we validated the results using treating physician 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, future efforts will develop flare criteria for myositis. Fifth, the response criteria were developed using a PM diagnosis based on Bohan and Peter's classification criteria, but experts now recognize that PM, by those criteria, may include different syndromes, such as necrotizing myopathy, the anti-synthetase syndrome, and others (40;41). We believe that these response criteria will still be applicable to these newer entities given that the data- and consensus-driven processes used herein were inclusive of those syndromes. In the future, with changes in classification criteria terminology (42), 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 administration of the response criteria. The time required to apply these criteria is estimated to be 25 minutes to complete the CSMs 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 immediate. Moreover, although the criteria may appear to be complicated, the CSM 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 CSMs, with the exception of 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 (43).

In conclusion, the development of a 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. This response criteria is sensitive and specific and provides a way to determine clinically meaningful change corresponding to degree of clinical improvement. This response criteria was 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 percentage change in CSM with thresholds for minimal, moderate, and major improvement was selected as the response criteria to be used for adult clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the following people for 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, United States Food and Drug Administration; Drs. Thorsten Vetter

and Richard Vesely, European Medicines Agency; Mr. Bob Goldberg and Ms. Theresa Curry, The Myositis Association; Ms. Rhonda McKeever and Ms. Patti Lawler, Cure JM Foundation; Ms. Irene Oakley, Myositis UK; and Drs. Michael Ward, Steven Pavletic, and Adam Schiffenbauer for their critical review of the manuscript.

This work was supported in part by the American College of Rheumatology, the European League Against Rheumatism, Cure JM Foundation, Myositis UK, Istituto G. Gaslini (Genoa, Italy) and the Paediatric Rheumatology International Trials Organisation (PRINTO), The Myositis Association, and the National Institute of Environmental Health Sciences, the National Center for Advancing Translational Sciences, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health. Paul Hansen owns the 1000Minds software referred to in this article, which he co-invented with Franz Ombler. Jiri Vencovsky's work in myositis was supported by the project (Ministry of Health, Czech Republic) for conceptual development of research organization 00023728 (Institute of Rheumatology). Ignacio Garcia-De la Torre's work is supported in part by the PNPC from CONACYT (Mexico City). Yeong Wook Song's work was supported by grant (# HI14C1277) from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea.

Reference List

1. Rider LG, Miller FW. Deciphering the clinical presentations, pathogenesis, and treatment of the idiopathic inflammatory myopathies. JAMA. 2011; 305(2):183–90. [PubMed: 21224460]
2. Moghadam-Kia S, Aggarwal R, Oddis CV. Treatment of inflammatory myopathy: emerging therapies and therapeutic targets. Expert Rev Clin Immunol. 2015; 11(11):1265–75. [PubMed: 26313852]
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(2):314–24. [PubMed: 23124935]
4. Muscle Study Group. A randomized, pilot trial of etanercept in dermatomyositis. Ann Neurol. 2011; 70(3):427–36. [PubMed: 21688301]
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(10019):671–8. [PubMed: 26645190]
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–S157. [PubMed: 22588740]
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(11):1262–73. [PubMed: 11709610]
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(3):603–17. [PubMed: 12610824]
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(7):2281–90. [PubMed: 15248228]
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(2):193–202. [PubMed: 17330293]
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(R) Consensus Exercises. Arthritis Care Res (Hoboken). 2016; 68(5):667–72. [PubMed: 26414176]

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(6):706–14. [PubMed: 24721558]
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(6):596–601. [PubMed: 21317432]
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(7):1467–70. [PubMed: 21724718]
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(8):1259–64. [PubMed: 23335569]
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(6):263–75. [PubMed: 25371708]
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. [PubMed: 22223558]
20. Rider LG, Aggarwal R, Pistorio A, Bayat N, Erman B, Feldman BM, et al. 2016 American College of Rheumatology (ACR) – European League Against Rheumatism (EULAR) 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. 2016; 00(00):000.
21. Rider LG. Outcome assessment in the adult and juvenile idiopathic inflammatory myopathies. Rheum Dis Clin North Am. 2002; 28(4):935–77. [PubMed: 12506779]
22. 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. Arthritis Rheumatol. 2014; 66(S11):S404.
23. Rider LG, Ruperto N, Pistorio A, Erman B, Bayat N, Lachenbruch PA, et al. 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). 2016; 2016 Submitted.
24. Bohan A, Peter JB. Polymyositis and dermatomyositis. Parts 1 and 2. N Engl J Med. 1975; 292:344–7. 3403–407. [PubMed: 1090839]
25. 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). Arthritis Rheum. 2011; 63(Suppl):S89.
26. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978; 8(4):283–98. [PubMed: 112681]
27. 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(2):142–6. [PubMed: 18460270]
28. 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.
29. Nair R, Aggarwal R, Khanna D. Methods of formal consensus in classification/diagnostic criteria and guideline development. Semin Arthritis Rheum. 2011; 41(2):95–105. [PubMed: 21420149]
30. 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(5):790–7. [PubMed: 20388738]

31. 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(12):2168–73. [PubMed: 24026675]
32. 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(1):4–13. [PubMed: 18163404]
33. 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–S9.
34. Apaz MT, Saad-Magalhaes C, Pistorio A, Ravelli A, de Oliveira SJ, Marcantoni MB, et al. Health-related quality of life of patients with juvenile dermatomyositis: results from the Pediatric Rheumatology International Trials Organisation multinational quality of life cohort study. Arthritis Rheum. 2009; 61(4):509–17. [PubMed: 19333974]
35. 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(11):2737–47. [PubMed: 24122180]
36. 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(9):2582–91. [PubMed: 20872596]
37. 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(10):2557–68. [PubMed: 26352873]
38. 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(6):727–35. [PubMed: 7779114]
39. 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(3):573–86. [PubMed: 21294106]
40. 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(1):227–32. [PubMed: 23422076]
41. 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(12):1635–8. [PubMed: 16679430]
42. Tjarnlund A, Bottai M, Rider LG, Werth VP, Pilkington C, de Visser M, et al. Progress report on development of classification criteria for adult and juvenile idiopathic inflammatory myopathies. Arthritis & Rheumatism. 2012; 64(Suppl):S323–S324.
43. Alexanderson H, Del GM, 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(3):581–92. [PubMed: 24429182]
44. 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. [PubMed: 22723809]

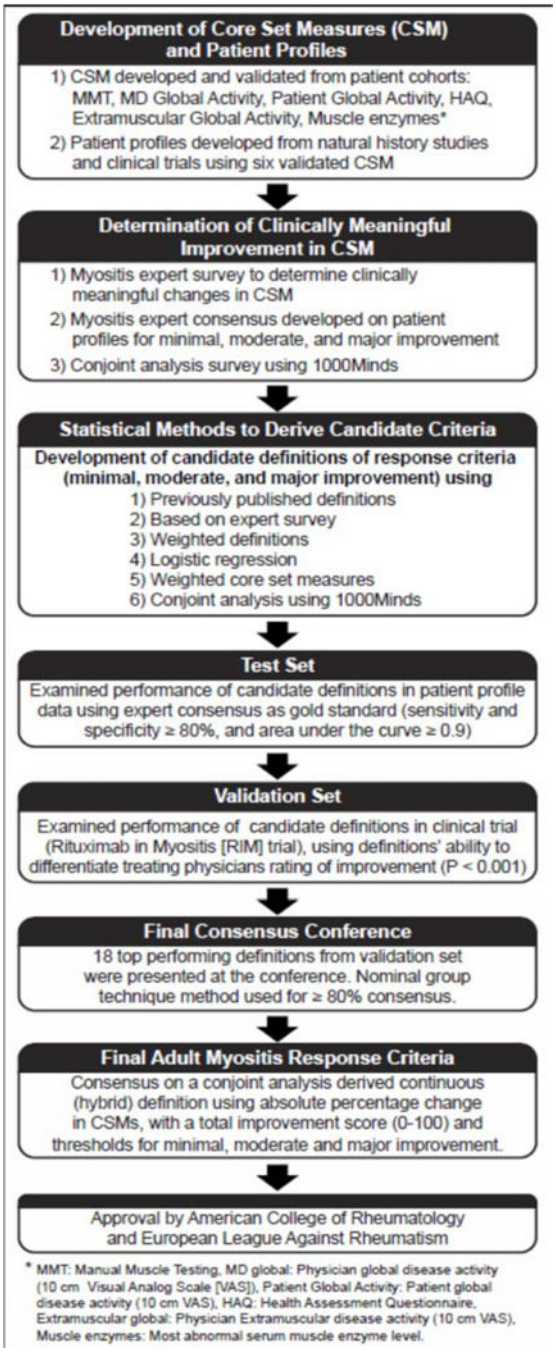

Figure 1.

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

Types of candidate definitions of response	Description	Example of candidate definition for response criteria
Previously published (categorical definition)	Previously published definitions of improvement that were re-tested.	MINIMAL: 3 of any 6 improved by ≥20%; no more than 2 worse by > 25%; which cannot be MMT (9) MODERATE: 3 of any 6 improved by ≥50%; no more than 2 worse by > 25%; which cannot be MMT MAJOR: 3 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 percent change candidate definitions of response, based on recent CSM survey.	MINIMAL: 2 of any 6 improved by ≥30%; no more than 1 worse by > 30%; which cannot be MMT MODERATE: 2 of any 6 improved by ≥50%; no more than 1 worse by > 30%; which cannot be MMT MAJOR: 2 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 out 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 percentage change in each CSM by the weights derived from conjoint analysis. Then sum (percent 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 their Improvement Score is above the cutoff for minimal improvement; similarly, for moderate and major improvement.	The full model is shown in Table 3, but here are the cut points for the adult DM/PM model: MINIMAL: Improvement Score ≥20 MODERATE: Improvement Score ≥40 MAJOR: Improvement Score ≥60

Abbreviations: MINIMAL, minimal improvement; MMT, manual muscle testing; MODERATE, moderate improvement; MAJOR, major improvement; CSM, core set measure; MD Global, Physician Global Activity; Patient Global, Patient's 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; DM, dermatomyositis; PM, polymyositis.

Table 2

Detailed performance characteristics of patient profiles and clinical trial data of the top 5 candidate response criteria definition presented at the consensus conference*

Candidate definition for response criteria	Improvement	Profiles (N=270)†				RIM Trial (N=147)‡			
		Sensitivity (%)	Specificity (%)	Threshold AUC	Total AUC	Candidate definition, improved physician rating§	Candidate definition, not improved physician rating§	*P* value	Rank
Conjoint analysis absolute % change (model 3)‖	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	1
	Major: Total Improvement Score ≥60	92	98	0.95	1.00	2.0	3.0	<0.001	
Conjoint analysis relative % change (model 1)¶	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	2
	Major: Improvement Score ≥70	100	95	0.97	0.99	2.0	3.0	<0.001	
Conjoint analysis relative % change (model 2)¶	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	3
	Major: Improvement Score ≥65	100	98	0.99	1.00	2.0	3.0	<0.001	
Weighted core set measure relative % change#	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	4
	Major: Improvement Score ≥400	100	94	0.97	1.00	2.0	3.0	<0.001	
Logistic regression relative % change**	Minimal: Improvement Score ≥75	89	93	0.91	0.97	2.0	3.0	<0.001	5

Candidate definition for response criteria	Improvement	Profiles (N=270)†				RIM Trial (N=147)‡			
		Sensitivity (%)	Specificity (%)	Threshold AUC	Total AUC	Candidate definition, improved physician rating§	Candidate definition, not improved physician rating§	*P* value	Rank
	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	

Abbreviations: %, percentage; Threshold AUC, area under the curve, calculated as the AUC from the receiver operating characteristic curve for the Total Improvement Score and the threshold for minimal, moderate, and major improvement; Total AUC, calculated as the AUC from the receiver operating characteristic curve, using the Total Improvement Score and the threshold cutoffs for minimal, moderate, and major improvement, which applies only to continuous definitions; Minimal, minimal improvement; Moderate, moderate improvement; Major, major improvement; N/A, not applicable.

*Supplementary Table 2 presents definitions 6–18 from the consensus conference ratings.

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

‡RIM trial, Rituximab in Myositis clinical trial (3).

§Physician rating: 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 clinical trial (3). Comparison of median rating when the candidate response criteria definition was improved versus not improved. A 1-point difference in physician rating of improvement from no improvement to minimal improvement was considered not just statistically significant, but also was clinically significant.

‖Conjoint analysis–based continuous candidate response criteria using absolute percentage change in core set measures (absolute percentage change model) is presented in Table 3. This criteria is also the top response criteria for juvenile dermatomyositis, but with different thresholds in the Total Improvement Score for minimal, moderate and major improvement (20).

¶Conjoint analysis–based continuous candidate definitions using relative percentage change in core set measures are presented in Supplementary Table 3. These criteria are also the second and third choice criteria for juvenile dermatomyositis, but with different thresholds in the Total Improvement Score for minimal, moderate and major improvement (20).

#Total Improvement Score = 2×(MD Global % change) + (Patient Global % change) + 3×(MMT % change) + 1.5×(HAQ % change) + 1.5×(ExtraMusc % change) + (Enzyme % change)

**Total Improvement Score = (MD Global % change) + (Patient Global % change) + (MMT % change) + (HAQ % change) + (ExtraMusc % change) + (Enzyme % change)

Table 3

Final myositis response criteria (conjoint analysis model 3) for minimal, moderate, and major improvement in adult DM/PM and combined adult DM/PM and JDM clinical trials and studies

Conjoint analysis–based continuous response criteria using absolute percentage change in core set measures		
Core set measure	**Level of improvement**	**Level 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 (MMT)	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 (HAQ)	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
Extra muscular 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
	Improvement category	**Total Improvement Score** *
Adult DM/PM thresholds	Minimal	≥ 20
	Moderate	≥ 40

Conjoint analysis–based continuous response criteria using absolute percentage change in core set measures

Core set measure	Level of improvement	Level score
JDM thresholds	Minimal	≥ 30
	Moderate	≥ 45
	Major	≥ 70

Abbreviations: DM, dermatomyositis; PM, polymyositis; JDM, juvenile dermatomyositis; Enzyme, most abnormal serum muscle enzyme level among creatine kinase, aldolase, alanine transaminase, aspartate aminotransferase, and lactate dehydrogenase.

*Note that this response criteria is also proposed for use in combined adult DM/PM and JDM clinical trials (20). For comparison, the thresholds of improvement in the Total Improvement Score for JDM are ≥ 30 for minimal, ≥ 45 for moderate, 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 percentage change (final value − baseline value / range) × 100 is calculated for each core set measure. For muscle enzymes, the most abnormal enzyme at baseline is used. The enzyme range was calculated based on 90% range of enzymes from natural history data (32;44), and for creatine kinase is 15 times the upper limit of normal; for aldolase is 6 times the upper limit of normal, and for lactate dehydrogenase, aspartate aminotransferase, and alanine transaminase is 3 times the upper limit of normal. Upper limit of normal is as per the individual laboratory participating in the center. Range for Physician Global, Patient Global, MMT, HAQ, and Extramuscular Global Activity are based on the instrument scale used in the trial. An improvement score is assigned for each core set measure based on the absolute percentage change in core set measure as per the definition. These individual core set measure improvement scores are then totaled among the six 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 0–100 also corresponds to the degree of improvement, with higher improvement scores corresponding to a greater degree of improvement.