Published in final edited form as:

Arthritis Rheumatol. 2017 May ; 69(5): 911–923. doi:10.1002/art.40060.

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

Lisa G. Rider, MD[*,1], **Rohit Aggarwal, MD, MSc**[*,2], **Angela Pistorio, MD, PhD**[3], **Nastaran Bayat, MD**[1], **Brian Erman**[4], **Brian M. Feldman, MD, MSc, FRCPC**[5], **Adam M. Huber, MD**[6], **Rolando Cimaz, MD**[7], **Rubén J. Cuttica, MD**[8], **Sheila Knupp de Oliveira, MD**[9], **Carol B. Lindsley, MD**[10], **Clarissa A. Pilkington, BSc, MBBS, MRCP**[11], **Marilyn Punaro, MD**[12], **Angelo Ravelli, MD**[13,14] **[Prof.]**, **Ann M. Reed, MD**[15], **Kelly Rouster-Stevens, MD**[16], **Annet van Royen, MD, PhD**[17], **Frank Dressler, MD**[18], **Claudia Saad Magalhaes, MD**[19], **Tamás Constantin, MD, PhD**[20], **Joyce E. Davidson, FRCP, FRCPCH**[21], **Bo Magnusson, MD**[22], **Ricardo Russo, MD**[23], **Luca Villa**[13], **Mariangela Rinaldi**[13], **Howard Rockette, PhD**[2], **Peter A. Lachenbruch, PhD**[1], **Frederick W. Miller, MD, PhD**[1], **Jiri Vencovsky, MD, PhD**[**,24], **Nicolino Ruperto, MD, MPH**[**,13] **for the International Myositis Assessment and Clinical Studies Group (IMACS) and the Paediatric Rheumatology International Trials Organisation (PRINTO)**

[1]Environmental Autoimmunity Group, Clinical Research Branch, NIEHS, NIH, Bethesda, MD [2]University of Pittsburgh, Pittsburgh, PA [3]Istituto Giannina Gaslini, Servizio di Epidemiologia e Biostatistica, Genoa, Italy [4]Social and Scientific Systems, Inc., Durham, NC [5]The Hospital for Sick Children, Toronto, Ontario, Canada [6]IWK Health Centre, Halifax, Nova Scotia, Canada [7]University of Firenze, Firenze, Italy [8]Hospital de Niños Pedro de Elizalde, University of Buenos Aires, Buenos Aires, Argentina [9]Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil [10]University of Kansas City Medical Center, Kansas City, KS [11]Great Ormond Street Hospital for Children NHS Trust, London, United Kingdom [12]University of Texas Southwestern Medical Center, Dallas, TX [13]Istituto Giannina Gaslini, Pediatria II, PRINTO, Genoa, Italy [14]Università degli Studi di Genova, Dipartimento di Pediatria, Genoa, Italy [15]Duke University, Durham, NC [16]Emory University School of Medicine, Atlanta, GA [17]University Medical Centre Utrecht – Wilhelmina Children's Hospital, Utrecht, Netherlands [18]Hannover Medical School, Hannover, Germany [19]Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, Saõ Paulo, Brazil [20]Semmelweis University, Budapest, Hungary [21]Royal Hospitals for Sick Children, Glasgow and Edinburgh, United Kingdom [22]Karolinska University Hospital, Stockholm, Sweden [23]Hospital de

Address correspondence to: Nicolino Ruperto, MD, MPH, Istituto Giannina Gaslini, Pediatria II, PRINTO, Via G. Gaslini 5, 16147 Genova Italy, Phone: +39-010-382854, Fax: +39-010-3933424, nicolaruperto@gaslini.org, www.printo.it or http://www.printo.it/pediatric-rheumatology/.

*co-first authors

**co-last authors

Pediatría Garrahan, Buenos Aires, Argentina [24]Institute of Rheumatology and Department of Rheumatology, 1st Medical Faculty, Charles University, Prague, Czech Republic

Abstract

Objective—Develop response criteria for juvenile dermatomyositis (JDM).

Methods—We analyzed the performance of 312 definitions that used core set measures (CSM) from either the International Myositis Assessment and Clinical Studies Group (IMACS) or the Pediatric Rheumatology International Trials Organization (PRINTO) and were derived from natural history data and a conjoint-analysis survey. They were further validated in the PRINTO trial of prednisone alone compared to prednisone with methotrexate or cyclosporine and the Rituximab in Myositis trial. Experts considered 14 top-performing candidate criteria based on their performance characteristics and clinical face validity using nominal group technique at a consensus conference.

Results—Consensus was reached for a conjoint analysis–based continuous model with a Total Improvement Score of 0-100, using absolute percent change in CSM with thresholds for minimal (≥30 points), moderate (≥45), and major improvement (≥70). The same criteria were chosen for adult dermatomyositis/polymyositis with differing thresholds for improvement. The sensitivity and specificity were 89% and 91-98% for minimal, 92-94% and 94-99% for moderate, and 91-98% and 85-85% for major improvement, respectively, in JDM patient cohorts using the IMACS and PRINTO CSM. These criteria were validated in the PRINTO trial for differentiating between treatment arms for minimal and moderate improvement (P=0.009–0.057) and in the Rituximab trial for significantly differentiating the physician rating of improvement (P<0.006).

Conclusion—The response criteria for JDM was a conjoint analysis–based model using a continuous improvement score based on absolute percent change in CSM, with thresholds for minimal, moderate, and major improvement.

Keywords

juvenile dermatomyositis; response criteria; conjoint analysis; definitions of improvement; hybrid or continuous definition; outcome assessment; consensus

---

Juvenile dermatomyositis (JDM) is a systemic autoimmune disease characterized by chronic skeletal muscle inflammation and weakness. Core set measures (CSM) to assess JDM disease activity have been established and validated by the International Myositis Assessment and Clinical Studies Group (IMACS) and the Paediatric Rheumatology International Trials Organisation (PRINTO), with provisional endorsement by the American College of Rheumatology and the European League Against Rheumatism (1-6). Both core sets include physician and parent global activity, muscle strength, and physical function. IMACS also includes the most abnormal serum muscle enzyme and extramuscular global activity, whereas PRINTO includes instead a health-related quality-of-life measure, the Childhood Health Questionnaire, and a global activity score, the Disease Activity Score. IMACS measures muscle strength by manual muscle testing and PRINTO by the Childhood Myositis Assessment Scale (1;2;5). Combinations of these measures to determine clinical

improvement were developed to enhance the sensitivity of responses and decrease needed sample sizes, by using large prospective natural history data sets and expert clinician consensus as the gold standard. For both PRINTO and IMACS, at least 20% improvement in three of six CSM with no more than one or two worsening (muscle strength was not allowed to worsen) had been established as preliminary response criteria, and additional combinations of improvement in the CSM serve as secondary response criteria (7;8). PRINTO adapted their top criteria for minimal clinical improvement to moderate and major improvement by using cutoffs of 50% and 70%, akin to improvement criteria for juvenile idiopathic arthritis (9-11).

Although the preliminary response criteria for JDM advanced the assessment of patients and their responses to treatment, those criteria were limited by differences in the CSM and final consensus response criteria between IMACS and PRINTO, a lack of randomized controlled trial data for full validation, and inadequate exploration of more sensitive approaches using hybrid or continuous methods (12). The preliminary response criteria also considered each CSM equally, rather than differentially weighting them. However, most myositis experts agree that some CSM are more important, such as Physician Global Activity and muscle strength (3;13). For PRINTO studies, physician global evaluation of disease activity, muscle strength, and parent's global evaluation of the child's overall well-being were weighted as the most important CSM in a logistic regression analysis (3;8). Moreover, the preliminary response criteria did not validate criteria for moderate or major improvement. There is, therefore, a clear need to have standardized improvement criteria for all levels of improvement in future clinical trials, similar to what has been done for rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA).

For these reasons, IMACS and PRINTO conducted a joint effort to develop fully validated response criteria for JDM, including criteria for minimal, moderate, and major clinical response. The present report focuses on the consensus conference that considered the top candidate definitions of response leading to the final JDM response criteria.

Methods

In separate publications (14;15), we described the methodology used (a) to create patient profiles using natural history data and obtain expert consensus on minimal, moderate, and major improvement (14); (b) to determine differential weights of the CSM using conjoint analysis; and (c) to draft six types of candidate definitions for response criteria using the myositis expert survey on thresholds of improvement and data-driven methods, such as logistic regression and conjoint analysis (Table 1).

Conjoint analysis is a choice modeling or discrete choice experiment, which is a valid methodology for developing composite criteria and has been used recently in rheumatology (16-19). In the conjoint-analysis surveys administered using 1000Minds online software (20), 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, relative weights of CSMs and their levels of improvement were established

and used to develop a scoring system by mathematical methods based on linear programming (21) such that when all six CSM are considered together, the maximum score (Total Improvement Score) possible for representing a patient's improvement is 100 and the minimum score is 0.

We then compared the performance characteristics of the drafted definitions in the patient profiles using expert consensus ratings as a gold standard and externally validated the candidate response criteria by applying them to clinical trial data. This process led to the development of traditional categorical as well as continuous candidate definitions for response criteria, with thresholds for minimal, moderate, and major improvement (22). Continuous candidate definitions can also be considered hybrid definitions, because the same definition can be used either as a continuous outcome measure by using the Total Improvement Score or as categorical outcome measure by using the thresholds for minimal, moderate, and major improvement.

Candidate definitions were evaluated using consensus profile ratings as the gold standard, by 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 using IMACS or PRINTO CSM (1)] were externally validated. The PRINTO trial randomized patients with new-onset JDM to receive prednisone alone (n=47) or prednisone combined with methotrexate or cyclosporine (n=46 patients per arm) (11). Chi-square analysis was used to compare the percentages of patients meeting the candidate definitions for response at the primary endpoint (6 months) for the combined treatment arms versus the prednisone alone (placebo) arm. Definitions with a significant difference ($P < 0.05$) between treatment arms for minimal improvement were further considered. Both PRINTO and IMACS CSM were available in this trial. A second trial validation dataset included 48 JDM patients enrolled in the Rituximab in Myositis (RIM) trial for treatment-refractory patients. It had a randomized placebo-phase design where patients received either rituximab or placebo at weeks 0 and 1, and at weeks 8 and 9 their treatment assignment was blindly reversed (23). We used the Mann-Whitney U test to determine whether each candidate definition could differentiate between the treating physician's rating of improvement (score range, 1-7) at 6 months, a time point when most patients improved and that was also comparable to the PRINTO trial. For the RIM trial, only the IMACS CSM were available.

We then selected the top candidate definitions, up to four top-performing definitions from each of the six different types of candidate definitions (Table 1), for consideration at the final consensus conference, as a manageable number of definitions to discuss.

Consensus conference

Nominal group technique was used at a consensus conference held in Paris, France on June 9-10, 2014, led by experienced moderators (Drs. Ruperto and Rider for the pediatric working group). The methodologies used to develop the new candidate response criteria and performance characteristics of each type of candidate definition were reviewed with the participants in a general session. The 12 pediatric working group participants first

independently and then as a group reviewed the performance characteristics of the 14 top candidate definitions of response criteria for JDM. Data for minimal, moderate, and major clinical response were presented for each definition, including a detailed spreadsheet that included the performance in the patient profiles using the IMACS and PRINTO CSM, including sensitivity, specificity, area under the curve (AUC), as well as kappa and odds ratio. AUC was defined as the average of the sensitivity and specificity for all categorical candidate definitions, as well as for thresholds of minimal, moderate, and major improvement in continuous candidate definitions. In addition, for continuous definitions, an AUC for the Total Improvement Score 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 (24-26). Results of the external validation for each candidate definition from the PRINTO and Rituximab clinical trial datasets were also presented.

Pediatric working group

After reviewing the performance of the 14 top-performing candidate definitions, the 12 pediatric working group participants developed consensus response criteria for minimal, moderate, and major improvement for JDM. Participants were informed of the secondary goal to reach consensus on response criteria for both JDM and adult dermatomyositis (DM)/polymyositis (PM). Participants were first asked to rank their top five choices, considering the data presented, based on face validity, feasibility, and generalizability, and to determine which response criteria were most clinically meaningful. The voting process was conducted in a systematic fashion with a predetermined format using nominal group technique (27;28) facilitated by an internet-based system developed by the PRINTO coordinating center (29;30). Voting was done anonymously and independently using the online voting software. After the initial round of voting, the results were shared with the group. Each participant was then asked to explain their top- and bottom-ranked choices to the group. The rounds of voting continued in the same manner until consensus was reached (≥80% of the votes) or until it was clear that consensus would not be reached. Between each round, after the participants were shown the results, the administrators were allowed to remove candidate definitions that decisively received a small proportion of the votes. In the final round, participants were asked to select their final top response criteria. The pediatric working group also voted on additional issues, including use of both IMACS and PRINTO CSM and response criteria for JDM that would interchange both the IMACS and PRINTO measures. Participants also voted on re-testing the performance of the top candidate response criteria in future trials.

Combined pediatric and adult working group

After consensus was attained for JDM response criteria, a combined working group of 22 pediatric and adult experts was formed to determine whether consensus could be reached on final, common response criteria for both JDM and adult DM/PM. Common response criteria that would include both JDM and adult DM/PM patients were considered for use in clinical trials, which might facilitate drug approvals for myositis. Experienced moderators (Drs. Ruperto, Rider, Aggarwal, and Miller) led the combined working group. For the first round of votes, the top adult and pediatric definitions from the final round of voting in each working group were considered. The online voting system was utilized again, and each

participant discussed their top-choice candidate definition using nominal group technique in a round-robin fashion. At each round, participants were asked to select only one candidate top response criterion; discussion was stopped once consensus ≥80% was reached. For determining the thresholds of improvement for the selected definition, the required consensus was ≥70%, which was done by post-conference voting.

Results

The performance characteristics of 101 of 312 candidate definitions were excellent (sensitivity and specificity ≥80%, AUC ≥0.90 for minimal improvement), and 30 candidate definitions also performed well in two clinical trials, where they differentiated between treatment arms ($P<0.05$ for minimal improvement) and differentiated treating physician's improvement score at week 24 ($P<0.001$) (13).

Top candidate definitions for response criteria

Fourteen top-performing candidate definitions were brought to the pediatric working group for consideration at the consensus conference (Table 2 and Supplementary Table 1). These candidate criteria included nine categorical definitions in which different criteria were set for minimal, moderate, and major improvement and five continuous definitions in which improvement points are given on a continuous scale that corresponds to the magnitude of improvement, with different thresholds for minimal, moderate, and major improvement. Among the nine categorical definitions, two were published IMACS and PRINTO response criteria (7-9), four were newly drafted definitions based on a survey of experts, and three were weighted definitions. Of the continuous definitions, two were developed by logistic regression and three were developed from the conjoint-analysis survey. Of the 14 candidate criteria considered, 11 were based on relative percent change, and 3 were based on absolute percent change in the CSM.

The performance characteristics of these 14 candidate definitions are provided in Table 2 and Supplementary Tables 1 and 2. In the patient profiles, with expert consensus as a gold standard, all definitions presented at the conference had sensitivity and specificity ≥87% and AUC ≥0.90 for minimal improvement (Table 2 and Supplementary Table 1). For moderate improvement, specificity decreased but was ≥80% and AUC ≥0.88, and for major improvement specificity was generally ≥75% and AUC ≥0.84. For continuous definitions, the AUCs (from receiver operating characteristic curves) for Total Improvement Score were generally better than AUCs (average of sensitivity and specificity) for the thresholds of minimal, moderate, and major improvement. Performance was similar among the IMACS and PRINTO CSM for each definition.

Almost all candidate criteria were validated using the PRINTO trial at 6 months, where they could differentiate between treatment arms, with $P<0.05$ for minimal improvement (Table 2 and Supplementary Table 1). All candidate criteria were also validated in 48 JDM patients in the RIM trial (23). All definitions could differentiate the median treating physician's improvement score at week 24 ($P\leq0.006$).

Consensus conference voting

Among the 14 candidate definitions, 13 and 11 candidate definitions of response were promoted in the first and second voting rounds, respectively. In round three, six candidate definitions were chosen, each receiving a similar number of votes. These six included the three conjoint analysis–based continuous definitions, a conjoint analysis–based weighted definition, a logistic regression absolute percent change definition, and the previously published PRINTO preliminary response criterion (8;9). In the fourth round of voting and discussion, participants reached consensus on a final top response criterion, a conjoint analysis–based continuous model using absolute percent change in the IMACS or PRINTO CSM (Table 3).

Table 2 and Supplementary Table 1 provide the performance characteristics in the patient profiles and the trial validation for each of the top candidate response criteria presented at the conference. For the top conjoint analysis–based continuous response criteria using absolute percent change in each of the CSM, the sensitivity and specificity in the patient profiles was generally >90% and AUC >0.90 for both the IMACS and PRINTO measures. For the PRINTO trial, a difference in the treatment arms was detected for minimal and moderate improvement using the top response criteria, and in the RIM trial a difference in the physician's rating of improvement when the response criteria rated the patient as improved versus not improved was detected for minimal, moderate, and major improvement.

Pediatric experts favored the conjoint analysis–based continuous response criteria because of the continuous improvement score that corresponds to the magnitude of improvement and provides the ability to categorize a patient's degree of change into minimal, moderate, and major improvement. The continuous model definitions also differentially weight the various CSM, which experts thought were congruent with their assessment of the relative importance of each of the CSM. The top response criterion was based on absolute percent change in CSM, which was also favored by the participants because, given the various visual analogue scale measurements used in the CSM, the absolute percent changes were more congruent than relative percent changes with actual clinical changes that the myositis experts see in clinical practice.

Combined pediatric-adult working group

For this round of votes, the top two pediatric (Table 2) and adult definitions were considered (22). Two rounds of voting resulted in final consensus response criteria, with 91% of participants voting for the conjoint analysis–based continuous response criteria (Conjoint Analysis Model 3, see Table 2) based on absolute percent change in the CSM (Table 3). It was agreed that the top response criteria would be used in future clinical trials that combined JDM and adult DM/PM. Because the final response criteria were similar, participants favored using response criteria that would be common to JDM and adult DM/PM, and they favored combined studies when possible, as well as the possibility of comparing outcomes in separate studies using the same final response criteria.

Other votes

In a post-conference final vote by the Delphi method, 74% of the participants agreed to use the following pediatric threshold values for minimal, moderate, and major response for JDM patients: Total Improvement Score ≥30 (on a scale of 0 to 100) for minimal, ≥45 for moderate, and ≥70 for major improvement. In contrast, the final thresholds for minimal, moderate, and major response for adult DM/PM were ≥20, ≥40, and ≥60 points, respectively. The pediatric working group also reached consensus that, given the overall similarity between the IMACS and PRINTO response criteria, a joint IMACS-PRINTO response criteria for JDM is being proposed. The current development of the response criteria in parallel between the IMACS and PRINTO CSM necessitates that either all of the IMACS or all of the PRINTO CSM be used. The pediatric experts, however, committed to measure both IMACS and PRINTO CSM in future therapeutic trials, with 92% agreement, and to continue to test the interchangeability of the IMACS and PRINTO CSM. The group also unanimously agreed to retest the validity of the top five candidate definitions for response criteria and to utilize the other four definitions as secondary endpoints in future clinical trials. The top three of these criteria, the conjoint-analysis definitions, are the same for both JDM and adult DM/PM, with different thresholds of improvement (Table 3, Supplementary Table 3).

Discussion

Conjoint analysis–based continuous response criteria, based on absolute percent change in the CSM, were developed as the consensus- and data-driven response criteria for minimal, moderate, and major improvement for JDM. In the response criteria, either IMACS or PRINTO CSM could be used. In addition, it was also agreed that the same response criteria, using the IMACS CSM but with different thresholds for improvement, would be the consensus response criteria for adult and combined JDM and adult DM/PM trials in the future (22).

The comprehensive process used to develop final response criteria for minimal, moderate, and major improvement for JDM included the use of large prospective natural history datasets for JDM and two randomized controlled trials for validation, which included a wide range of disease activity and different stages of disease, from recently diagnosed to treatment-refractory patients (11;13;23). The involvement of many clinical experts who had experience using the CSM in JDM patients was also critical. They provided input at several points throughout the process, including determination of thresholds for improvement in CSM by which definitions of response were drafted, achievement of gold standard ratings of improvement by evaluating and developing consensus patient profiles, completion of the conjoint-analysis surveys to develop differential weights for the CSM, and participation in the final consensus conference to achieve consensus for common response criteria with greatest clinical face validity. The current response criteria (Table 3) also resolve the differences between PRINTO and IMACS CSM by having tested candidate definitions of response criteria in parallel using both sets of measures and learning that they are largely interchangeable and that their performance is comparable. Moreover, this project brought both IMACS and PRINTO consortia to work together for this rare disease.

The combined group of pediatric and adult experts selected the same top-choice definition but with differing thresholds for improvement, which had very similar performance characteristics and were thought to be more appropriate for use in clinical trials that would, in the future, combine adult and pediatric patients.

The final response criteria selected, conjoint analysis–based continuous response criteria using absolute percent change in the CSM, has many advantages. For each measure, improvement points are calculated based on the level of change in that measure, and each CSM is differentially weighted, such that changes in muscle strength and Physician Global Activity are weighted more heavily than changes in the most abnormal enzyme or quality of life. A Total Improvement Score can be obtained as a continuous measure, and the means or medians of Total Improvement Scores can be compared between treatment arms (31). A Total Improvement Score between 0 and 100 also corresponds to the degree of improvement, with higher scores corresponding to a greater magnitude of improvement. This score may be more sensitive to change, resulting in smaller trial sample sizes (31;32). Alternatively, thresholds for minimal, moderate, and major improvement have been established that allow dichotomous use of the response criteria as well. Therefore, this is truly a hybrid model that can be used as either a continuous or categorical outcome measure within the same response criteria depending on the trial design and needs of the study. The response criteria allow input from all the CSM, instead of relying only on a few measures to determine whether a patient has improved. However, although this response criterion was developed using all six CSM, the response criteria could still be used if fewer CSM were obtained, allowing for greater flexibility in the types of patients and improvements that can occur, but we caution that the response criteria are most accurate when all six CSM are used. As such, the response criteria signify a major advance in assessing improvement in treatment trials and other clinical research studies by providing data-driven response criteria, which were developed by consensus of major stakeholders in the field who come from all over the world.

Prior response criteria in rheumatic diseases have included relative percent change (33;34), whereas myositis response criteria are based on absolute percent change. The experts favored the use of absolute percent change for various reasons. In this study several CSM used the 10-cm Visual Analogue Scale, and the experts felt that absolute percent change better represents the degree of change they see in clinical practice. Moreover, absolute percent changes can be calculated when the baseline CSM is zero and give similar results for similar degrees of change at either end of the Visual Analogue Scale.

The participants also favored using the same response criteria for JDM and adult DM/PM, but with cut-points or thresholds for improvement specific to pediatric or adult patients. Having common response criteria facilitates the potential to conduct combined clinical trials, such as the RIM trial (23), and to compare the outcomes of trials and studies conducted separately. Participants agreed to include other top-performing definitions that were highly rated as secondary endpoints in future clinical trials. Among these were not only other conjoint analysis–based continuous models but also the published PRINTO preliminary response criteria (8;9). Future work should also evaluate whether a baseline composite score

threshold derived from the PRINTO or IMACS CSM could be used as inclusion criteria for future clinical trials.

Limitations of the present work include the lack of a placebo group in the RIM trial. For this reason, the physician's assessment of improvement at 6 months was used instead. We were fortunate to have another controlled clinical trial for JDM that had three treatment arms, for external validation (11), where we evaluated the ability of the candidate definitions to differentiate between treatment arms. Although thresholds for major improvement were developed and validated on fewer patients, we felt it was sufficient given that 29% of patients had major improvement in patient profiles and 17% had major improvement in the clinical trials used for validation. The final conjoint analysis–based continuous response criteria also do not address worsening in the CSM; however, this generally does not affect the outcome, as when patients are rated as improved, no more than one or two measures worsen in our clinical datasets. Also, although we tested the interchange of IMACS and PRINTO CSM, we tested these variations as two parallel CSM but did not examine intermixing the PRINTO and IMACS CSM. Further work to examine the interchangeability of the IMACS and PRINTO CSM will be needed. The datasets used to develop the new response criteria primarily contained recently diagnosed or flaring patients, and further work is needed to determine how the response criteria perform in patients with longstanding disease or those with significant disease damage. Finally, although the application of the criteria might seem cumbersome, as regularly done for JIA and RA, the evaluation of improvement will be facilitated by appropriate dedicated software or apps, or in the future, by simplification of the way the CSM are evaluated (e.g., similar to the Juvenile Arthritis Disease Activity Score for JIA)(35). The time required to apply these criteria is estimated to be 25-35 minutes to complete the CSMs at each visit (1) and 2-3 minutes to hand-calculate the Total Improvement Score and degree of response. Both IMACS and PRINTO are developing a web-based tool as well as a downloadable calculator that will allow easy administration of the response criteria and immediate calculation. The apparent complexity is, however, counterbalanced by the establishment of different validated levels of improvement, which constitute the real novelty of this project and which have never been validated as such either for RA or JIA, despite being regularly reported in clinical trials.

In sum, conjoint analysis–based continuous response criteria that establish different thresholds for minimal, moderate, and major improvement and utilize the absolute percent change in CSM was chosen as the consensus response criteria for JDM and underwent validation using both natural history and trial data. These response criteria should be highly acceptable and widely used given that they were developed with consensus among many myositis experts in the world. They should be sensitive in detecting differences in improvement and in quantitating the degree of improvement, as seen in the two clinical trials. Thus, clinical trials that test new therapies for JDM should be easier to design, conduct, and compare.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the American College of Rheumatology, the European League Against Rheumatism, the Intramural Research Program of the National Institutes of Health, 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; Istituto Giannina Gaslini Genova (Italy) and the Paediatric Rheumatology International Trials Organisation (PRINTO); Cure JM Foundation; Myositis UK; and The Myositis Association. Paul Hansen owns the 1000Minds software referred to in this article, which he co-invented with Franz Ombler. Jiri Vencovsky's work in myositis supported by the project (Ministry of Health, Czech Republic) for conceptual development of research organization 00023728 (Institute of Rheumatology).

We thank the following people for invaluable input and feedback on project development and support: Dr. Brian Feldman, American College of Rheumatology; Dr. Daniel Aletaha, European League Against Rheumatism; Drs. Suzette Peng and Sarah Yim, 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; and Ms. Irene Oakley, Myositis UK. We thank Drs. Michael Ward and Steven Pavletic for critical reading of the manuscript.

Appendix of contributors

Steering Committee: Lisa 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 Huber, Brian Feldman, Paul Hansen, Howard Rockette, Peter Lachenbruch, Nicolino Ruperto, Lisa Rider

Pediatric Core Set Survey Group: Maria Apaz, Suzanne Bowyer, Rolando Cimaz, Tamás Constantin, Megan Curran, Joyce Davidson, Brian Feldman, Thomas Griffin, Adam Huber, Olcay Jones, Susan Kim, Bianca Lang, Carol Lindsley, Daniel Lovell, Claudia Saad Magalhaes, Lauren M. Pachman, Clarissa Pilkington, Andrea Ponyi, Marilynn Punaro, Pierre Quartier, Athimalaipet V. Ramanan, Angelo Ravelli, Ann Reed, Robert Rennebohm, Annet Van Royen-Kerkhof, David D. Sherry, Clovis A. Silva, Elizabeth Stringer, Carol Wallace

Clinical trial or natural history study data set contributors: Frederick W. Miller, Chester Oddis, Ann Reed, Lisa G. Rider, Nicola Ruperto and PRINTO members

Pediatric Patient Profile Working Group: Maria Apaz, Tadej Avcin, Mara Becker, Michael W. Beresford, Rolando Cimaz, Tamás Constantin, Megan Curran, Ruben Cuttica, Joyce Davidson, Frank Dressler, Jeffrey Dvergsten, Brian Feldman, Virginia Paes Leme Ferriani, Berit Flato, Valeria Gerloni, Thomas Griffin, Michael Henrickson, Claas Hinze, Mark Hoeltzel, Adam Huber, Maria Ibarra, Norman Ilowite, Lisa Imundo, Olcay Jones, Susan Kim, Daniel Kingsbury, Bianca Lang, Carol Lindsley, Daniel Lovell, Alberto Martini, Claudia Saad Magalhaes, Bo Magnusson, Sheilagh Maguiness, Susan Maillard, Pernille Mathiesen, Liza McCann, Susan Nielsen, Sheila Knupp Feitosa de Oliveira, Lauren M. Pachman, Murray Passo, Clarissa Pilkington, Marilynn Punaro, Pierre Quartier, Egla Rabinovich, Athimalaipet V. Ramanan, Angelo Ravelli, Ann Reed, Robert Rennebohm, Lisa Rider, Rafael Rivas-Chacon, Angela Byun Robinson, Kelly Rouster-Stevens, Annet Van Royen-Kerkhof, Ricardo Russo, Lidia Rutkowska-Sak, Adriana Sallum, Helga Sanner, Heinrike Schmeling, Duygu Selcen, Bracha Shaham, David D. Sherry, Clovis A. Silva,

Charles H. Spencer, Robert Sundel, Marc Tardieu, Akaluck Thatayatikom, Janjaap Van Der Net, Dawn Wahezi, Carol Wallace, Francesco Zulian

Conjoint Analysis—Pediatric Group: Rolando Cimaz, Tamás Constantin, Ruben Cuttica, Joyce Davidson, Frank Dressler, Brian Feldman, Thomas Griffin, Michael Henrickson, Adam Huber, Lisa Imundo, Bianca Lang, Carol Lindsley, Claudia Saad Magalhaes, Bo Magnusson, Susan Maillard, Sheila Knupp Feitosa de Oliveira, Lauren M. Pachman, Murray Passo, Clarissa Pilkington, Marilynn Punaro, Angelo Ravelli, Ann Reed, Lisa Rider, Kelly Rouster-Stevens, Annet Van Royen-Kerkhof, Ricardo Russo, Bracha Shaham, Robert Sundel, Janjaap Van Der Net

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

Participants in Consensus Conference, Adult Working Group: Anthony Amato, Hector Chinoy, Robert G. Cooper, Maryam Dastmalchi, David Fiorentino, David Isenberg, James Katz, Andrew Mammen, Chester V. Oddis, Jiri Vencovsky, Marianne de Visser, Steven R. Ytterberg

Reference List

1. 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]
2. 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]
3. 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]
4. Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B, Feldman BM, et al. Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. Rheumatology (Oxford). 2003; 42(12):1452–9. [PubMed: 12832713]
5. 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]
6. 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]

7. 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]
8. Ruperto N, Pistorio A, Ravelli A, Rider LG, Pilkington C, Oliveira S, et al. The Pediatric Rheumatology International Trials Organization provisional criteria for the evaluation of response to therapy in juvenile dermatomyositis. Arthritis Care Res (Hoboken). 2010; 62:1533–41. [PubMed: 20583105]
9. Hasija R, Pistorio A, Ravelli A, Demirkaya E, Khubchandani R, Guseinova D, et al. Therapeutic approaches in the treatment of juvenile dermatomyositis in patients with recent-onset disease and in those experiencing disease flare: an international multicenter PRINTO study. Arthritis Rheum. 2011; 63(10):3142–52. [PubMed: 21647864]
10. Ruperto N, Pistorio A, Ravelli A, Hasija R, Guseinova D, Filocamo G, et al. Criteria to define response to therapy in paediatric rheumatic diseases. Eur J Clin Pharmacol. 2011; 67(Suppl 1):125–31. [PubMed: 21103985]
11. 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]
12. Felson DT, Furst DE, Boers M. Rationale and strategies for reevaluating the ACR20. J Rheumatol. 2007; 34(5):1184–7. [PubMed: 17477484]
13. Rider LG, Ruperto N, Pistorio A, Erman B, Bayat N, Lachenbruch PA, et al. 2016Development 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. :2016.
14. Rider LG, Aggarwal R, Bayat N, Erman B, Feldman BM, Huber AM, et al. A hybrid conjoint analysis model is proposed as the definition of minimal, moderate and major clinical improvement in juvenile dermatomyositis clinical trials. Arthritis Rheumatol. 66(S11):S578.2014;
15. 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.
16. 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]
17. 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]
18. 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.
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. 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]
21. 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.
22. Aggarwal R, Rider LG, Ruperto N, 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 Adult Dermatomyositis and Polymyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. Arthritis Rheumatol. 2016; 00(00):000.

23. 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]
24. Metz CE. Basic principles of ROC analysis. Semin Nucl Med. 1978; 8(4):283–98. [PubMed: 112681]
25. Finney, DJ. Statistical method in biological assay. 3. London: Charles Griffin; 1978.
26. Liengme, BV. A Guide to Microsoft Excel 2002 for Scientists and Engineers. Butterworth-Heinemann; 2002.
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. A guide to nominal group and delphi processes. Glenview, IL: Scott, Foresman and Company; 1975. Group techniques for program planning.
29. 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]
30. 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]
31. 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]
32. Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. Can J Psychiatry. 2002; 47(3):262–6. [PubMed: 11987478]
33. 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]
34. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum. 1997; 40(7):1202–9. [PubMed: 9214419]
35. Consolaro A, Ruperto N, Bazso A, Pistorio A, Magni-Manzoni S, Filocamo G, et al. Development and validation of a composite disease activity score for juvenile idiopathic arthritis. Arthritis Rheum. 2009; 61(5):658–66. [PubMed: 19405003]
36. 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]

Table 1

Types of candidate definitions for response criteria that were developed and tested

Type of candidate definitions of response	Description	Example of the candidate definition for the response criteria
Previously published (categorical definition)	Previously published response criteria that were retested.	MINIMAL: 3 of any 6 improved by ≥ 20%; no more than 1 worse by > 30%; which cannot be CMAS (8) MODERATE: 3 of any 6 improved by ≥ 50%; no more than 1 worse by > 30%; which cannot be CMAS (9) MAJOR: 3 of any 6 improved by ≥ 70%; no more than 1 worse by > 30%; which cannot be CMAS (9)
Newly drafted (categorical definition)	Drafted relative or absolute percent change candidate definitions of response, based on recent CSM survey.	MINIMAL: MD Global, muscle strength (MMT or CMAS), and one other CSM improved by ≥20% MODERATE: MD Global, muscle strength (MMT or CMAS), and one other CSM improved by ≥30% MAJOR: MD Global, muscle strength (MMT or CMAS), and one other CSM improved by ≥50%
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 3.5 Improvement Points out of 10 Total Improvement Points, and no more than 1.5 Worsening Points, where MD Global =2 points; Parent Global = 1 point; MMT/CMAS = 3 points; CHAQ = 1.5 points, ExtraMusc/DAS = 1.5 points, Enzyme/ CHQ-PhS = 1 point MINIMAL: Improvement Points given when CSM ≥20%; Worsening Points given when CSM worse by >30% MODERATE: Improvement Points given when CSM ≥50%; Worsening Points given when CSM worse by >30% MAJOR: Improvement Points given when CSM ≥75%; Worsening Points given when CSM worse by >30%
Logistic regression (continuous definition)	Model of improvement using a combination of CSM with different weights, as developed in the logistic regression model. Total scores derived, with different cutoffs for minimal, moderate, and major improvement. Relative percent change.	Improvement Score = (MD Global % change) + 0.5 × (Parent Global Activity % change) + 0.5 × (ExtraMusc Activity or DAS % change) MINIMAL: Improvement Score ≥ 15 MODERATE: Improvement Score ≥ 30 MAJOR: Improvement Score ≥ 60
Core set measure-weighted * (continuous definition)	Multiply the percent 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) + (Parent Global % change) + 3× (MMT or CMAS % change) + 1.5× (CHAQ % change) + 1.5× (ExtraMusc or DAS % change) + (Enzyme or CHQ-PhS % change) MINIMAL: Improvement Score ≥100 MODERATE: Improvement Score ≥250

Type of candidate definitions of response	Description	Example of the candidate definition for the response criteria
		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 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 absolute percent change model is shown in Table 3 and in Supplementary Table 2, but the cut points for the model for JDM are: MINIMAL: Improvement Score ≥ 30 MODERATE: Improvement Score ≥45 MAJOR: Improvement Score ≥70

Abbreviations: CMAS, Childhood Myositis Assessment Scale; CSM, core set measure; MD Global, Physician Global Activity; MMT, manual muscle testing; Parent Global, Parent's Global Activity Score; CHAQ, Childhood Health Assessment Questionnaire; ExtraMusc, Extramuscular Global Activity; DAS, Disease Activity Score; Enzyme, most abnormal serum muscle enzyme value among aldolase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, and creatine kinase; CHQ-PhS, Physical Summary Score of the Child Health Questionnaire-Parent Form 50.

*This type of definition was not brought to the final consensus conference.

Table 2

Detailed performance characteristics of patient profiles for the top five candidate definition presented at the consensus conference

Improvement category	Candidate definition for response criteria	Candidate definition type based on final consensus rank order	Core set measures	Sensitivity (%)	Specificity (%)	Threshold AUC‡	Total Improvement Score AUC§	PRINTO trial* Tx (%)	PRINTO trial* Ctrl (%)	PRINTO trial* *P* value	RIM trial† Response criteria, improved‖	RIM trial† Response criteria, not improved‖	RIM trial† *P* value	Rank
Minimal	Improvement Score ≥ 30	Conjoint analysis, absolute percent change Model 3 (Table 3)¶	IMACS	89	91	0.90	0.98	75	53	0.009	2.0	3.0	<0.001	1
			PRINTO	89	98	0.93	0.99	73	55	0.038				
Moderate	Improvement Score ≥ 45		IMACS	92	99	0.95	0.99	70	53	0.057	2.0	3.0	<0.001	
			PRINTO	94	94	0.94	0.99	71	51	0.023				
Major	Improvement Score ≥ 70		IMACS	91	86	0.89	0.96	51	43	0.341	2.0	3.0	0.006	
			PRINTO	98	85	0.91	0.98	58	49	0.331				
Minimal	Improvement Score ≥ 33	Conjoint analysis, relative percent change# Model 1 (Supplementary Table 3)	IMACS	99	87	0.93	0.98	75	55	0.018	2.0	4.0	<0.001	2
			PRINTO	96	98	0.97	1.00	74	55	0.027				
Moderate	Improvement Score ≥ 60		IMACS	97	93	0.95	0.99	73	51	0.011	2.0	3.0	<0.001	
			PRINTO	97	96	0.96	1.00	70	51	0.032				
Major	Improvement Score ≥ 80		IMACS	91	87	0.89	0.96	57	49	0.396	1.5	3.0	<0.001	
			PRINTO	98	86	0.92	0.97	61	49	0.179				
Minimal	Improvement Score ≥ 33	Conjoint analysis, relative percent change# Model 2 (Supplementary Table 3)	IMACS	95	94	0.94	0.98	75	53	0.009	2.0	4.0	<0.001	3
			PRINTO	94	98	0.96	0.99	74	55	0.027				
Moderate	Improvement Score ≥ 55		IMACS	95	95	0.95	1.00	70	51	0.032	2.0	3.0	<0.001	
			PRINTO	97	98	0.98	1.00	70	51	0.032				
Major	Improvement Score ≥ 77		IMACS	93	86	0.90	0.97	49	47	0.814	1.0	2.0	0.011	
			PRINTO	96	90	0.93	0.99	59	49	0.273				
Minimal	Improvement Points given when CS≥20; Worsening Points given when CSM worse by > 30	Weighted definition, relative percent change**	IMACS	95	100	0.97	NA	70	51	0.032	2.0	3.0	<0.001	4
			PRINTO	92	98	0.95	NA	73	53	0.021				
Moderate	Improvement Points given when CSM ≥ 50%; Worsening Points given when CSM worse by >30%		IMACS	95	91	0.93	NA	68	51	0.045	2.0	3.0	<0.001	

Improve-ment category	Candidate definition for response criteria	Candidate definition type based on final consensus rank order	Core set measures	Sensitivity (%)	Specificity (%)	Threshold AUC‡	Total Improvement Score AUC§	PRINTO trial* Tx (%)	PRINTO trial* Ctrl (%)	PRINTO trial* *P* value	RIM trial† Response criteria, improved‖	RIM trial† Response criteria, not improved‖	RIM trial† *P* value	Rank
								71	51	0.023				
			PRINTO	95	92	0.94	NA							
			IMACS	100	81	0.91	NA	64	47	0.050	1.5	3.0	<0.001	
Major	Improvement Points given when CSM≥75%; Worsening Points given when CSM worse by >30%							62	49	0.142				
			PRINTO	98	85	0.91	NA							
			IMACS	93	100	0.97	NA	70	51	0.032	2.0	3.0	<0.001	
Minimal	3 of any 6improved by ≥20%; no more than 1 worse by > 30%; which cannot be MMT/CMAS (8)							71	51	0.023				
			PRINTO	88	100	0.94	NA							5
			IMACS	90	95	0.93	NA	66	51	0.081	2.0	3.0	<0.001	
Moderate	3 of any 6 improved by ≥ 50%; no more than 1 worse by > 30%; which cannot be MMT/CMAS (9)	Previously published definition (8;9), relative percent change						68	51	0.045				
			PRINTO	90	96	0.93	NA							
			IMACS	99	83	0.91	NA	63	49	0.111	2.0	3.0	<0.001	
Major	3 of any 6 improved by ≥ 70%; no more than 1 worse by > 30%; which cannot be MMT/CMAS (9)							60	49	0.223				
			PRINTO	99	89	0.94	NA							

The performance characteristics of patient profiles for definitions ranked 6-14 are presented in Supplementary Table 1.

Abbreviations: AUC, area under the curve; Minimal, minimal improvement; Moderate, moderate improvement; Major, major improvement; IMACS, International Myositis Assessment and Clinical Studies Group; PRINTO, Paediatric Rheumatology International Trials Organization; CSM, core set measure; NA, not applicable; MMT, manual muscle testing; CMAS, Childhood Myositis Assessment Scale.

Note that either IMACS or PRINTO CSM may be used in these candidate definitions of response; the candidate definitions were developed in parallel with IMACS or PRINTO CSM.

*PRINTO juvenile dermatomyositis trial of prednisone alone versus prednisone with methotrexate or cyclosporine (n = 139) (11).

†Rituximab in Myositis (RIM) Trial, juvenile dermatomyositis arm (n =48). Comparison of the treating physician's rating of improvement if the improvement criteria are met versus not at week 24 (19). 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.

‡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 Improvement Score AUC, calculated as the AUC from the receiver operator characteristic curve, using the Total Improvement Score and the threshold cutoffs for minimal, moderate, and major improvement, which applies only to continuous definitions.

Median Physician Improvement Score.

¶ Conjoint analysis–based continuous candidate response criteria using absolute percent change in core set measures (absolute percent change model) is presented in Table 3. These criteria are also the top response criteria for adult dermatomyositis/polymyositis, but with different thresholds in the Total Improvement Score for minimal, moderate, and major improvement (22).

Conjoint analysis–based continuous candidate definitions using relative percent change in core set measures are presented in Supplementary Table 3. These criteria are also the second and third choice criteria for adult dermatomyositis/polymyositis, but with different thresholds in the Total Improvement Score for minimal, moderate, and major improvement (22).

** Improvement = at least 3.5 Improvement Points out of 10 Total Improvement Points, and no more than 1.5 Worsening Points, where Physician Global Activity = 2 points; Parent Global Activity = 1 point; MMT or CMAS = 3 points; CHAQ = 1.5 points, Extramuscular Global Activity or Disease Activity Score = 1.5 points, Enzyme or Physical Summary Score of the Child Health Questionnaire-Parent Form 50 = 1 point.

Table 3

Final top response criteria for minimal, moderate, and major improvement in JDM and combined adult DM/PM and JDM clinical trials and studies

Conjoint analysis–based continuous response criteria using absolute percent 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
Parent 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
MMT or CMAS	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
CHAQ	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) or CHQ-PhS	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 or Disease Activity Score	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†**
JDM thresholds	Minimal	≥30
	Moderate	≥45

Conjoint analysis–based continuous response criteria using absolute percent change in core set measures		
Core Set Measure*	**Level of Improvement**	**Level Score**
Adult DM/PM thresholds	Major	≥70
	Minimal	≥20
	Moderate	≥40
	Major	≥60

Abbreviations: JDM, juvenile dermatomyositis; DM, dermatomyositis; PM, polymyositis; MMT, manual muscle testing; CMAS, Childhood Myositis Assessment Scale; CHAQ, Childhood Health Assessment Questionnaire; Enzyme, most abnormal serum muscle enzyme level among creatine kinase, aldolase, alanine aminotransferase, aspartate aminotransferase, and lactate dehydrogenase; CHQ-PhS, Physical Summary Score of the Child Health Questionnaire-Parent Form 50.

*Note that either all the IMACS or all the PRINTO core set measures may be used.

†Note that this response criteria is also proposed for use in combined adult DM/PM and JDM clinical trials (22).

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 enzyme at baseline is used. The enzyme range was calculated based on 90% range of enzymes from natural history data (5;36), which for creatine kinase is 20 times the upper limit of normal (ULN), for aldolase is 6 times the ULN, and for lactate dehydrogenase, aspartate aminotransferase, and alanine transaminase is 5 times the ULN. An Improvement Score is assigned for each core set measure based on the absolute percent change. These are totaled among the six IMACS or PRINTO core set measures. The thresholds for minimal, moderate, and major improvement are provided. The Total Improvement Scores may also be compared among treatment arms in a trial. A Total Improvement Score between 0 and 100 also corresponds to the degree of improvement, with higher scores corresponding to a higher magnitude of improvement.