

SPECIAL ARTICLE

# 2016 American College of Rheumatology/European League Against Rheumatism Criteria for Minimal, Moderate, and Major Clinical Response in Juvenile Dermatomyositis

An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative

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

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

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

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

Supported in part by the American College of Rheumatology, the European League Against Rheumatism, the NIH (Intramural Research Programs of the National Institute of Environmental Health Sciences [NIEHS], National Center for Advancing Translational Sciences, and National Institute of Arthritis and Musculoskeletal and Skin Diseases), Istituto G. Gaslini and the Paediatric Rheumatology International Trials Organisation (PRINTO), Cure JM Foundation, Myositis UK, and the Myositis Association. Dr. Vencovsky's work was supported by the Ministry of Health, Czech Republic (Institute of Rheumatology project for conceptual development of a research organization, 00023728).

<sup>1</sup>Lisa G. Rider, MD, Nastaran Bayat, MD, Peter A. Lachenbruch, PhD, Frederick W. Miller, MD, PhD: NIEHS, NIH, Bethesda, Maryland; <sup>2</sup>Rohit Aggarwal, MD, MSc, Howard Rockette, PhD: University of Pittsburgh, Pittsburgh, Pennsylvania;

<sup>3</sup>Angela Pistorio, MD, PhD: Istituto Giannina Gaslini, Servizio di Epidemiologia e Biostatistica, Genoa, Italy; <sup>4</sup>Brian Erman, MS: Social and Scientific Systems, Inc., Durham, North Carolina; <sup>5</sup>Brian M. Feldman, MD, MSc, FRCPC: The Hospital for Sick Children, Toronto, Ontario, Canada; <sup>6</sup>Adam M. Huber, MD: IWK Health Centre, Halifax, Nova Scotia, Canada; <sup>7</sup>Rolando Cimaz, MD: University of Firenze, Florence, Italy; <sup>8</sup>Rubén J. Cuttica, MD: Hospital de Niños Pedro de Elizalde, University of Buenos Aires, Buenos Aires, Argentina; <sup>9</sup>Sheila Knupp de Oliveira, MD: Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; <sup>10</sup>Carol B. Lindsley, MD: University of Kansas City Medical Center, Kansas City, Kansas; <sup>11</sup>Clarissa A. Pilkington, BSc, MBBS, MRCP: Great Ormond Street Hospital for Children NHS Trust, London, UK; <sup>12</sup>Marilynn Punaro, MD: University of Texas Southwestern Medical Center, Dallas; <sup>13</sup>Angelo Ravelli, MD: Istituto Giannina Gaslini, Pediatria II - Reumatologia, and Università degli Studi di Genova, Genoa, Italy; <sup>14</sup>Ann M. Reed, MD: Duke University, Durham, North Carolina; <sup>15</sup>Kelly Rouster-Stevens,

**Objective.** To develop response criteria for juvenile dermatomyositis (DM).

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

**Results.** Consensus was reached for a conjoint analysis-based continuous model with a total improvement score of 0–100, using absolute percent change in core set measures of minimal ( $\geq 30$ ), moderate ( $\geq 45$ ), and major ( $\geq 70$ ) improvement. The same criteria were chosen for adult DM/polymyositis, with differing thresholds for improvement. The sensitivity and specificity were 89% and 91–98% for minimal improvement, 92–94% and 94–99% for moderate improvement, and 91–98% and 85–86% for major improvement, respectively, in juvenile DM patient cohorts using the IMACS and PRINTO core set measures. 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 RIM trial for significantly differentiating the physician's rating for improvement ( $P < 0.006$ ).

**Conclusion.** The response criteria for juvenile DM consisted of a conjoint analysis-based model using a continuous improvement score based on absolute percent change in core set measures, with thresholds for minimal, moderate, and major improvement.

MD: Emory University School of Medicine, Atlanta, Georgia; <sup>16</sup>Annet van Royen-Kerkhof, MD, PhD: University Medical Centre Utrecht, Wilhelmina Children's Hospital, Utrecht, The Netherlands; <sup>17</sup>Frank Dressler, MD: Hannover Medical School, Hannover, Germany; <sup>18</sup>Claudia Saad Magalhaes, MD: Universidade Estadual Paulista Júlio de Mesquita Filho, Botucatu, São Paulo, Brazil; <sup>19</sup>Tamás Constantin, MD, PhD: Semmelweis University, Budapest, Hungary; <sup>20</sup>Joyce E. Davidson, FRCP, FRCPCH: Royal Hospital for Sick Children, Glasgow, UK, and Royal Hospital for Sick Children, Edinburgh, UK; <sup>21</sup>Bo Magnusson, MD: Karolinska University Hospital, Stockholm, Sweden; <sup>22</sup>Ricardo Russo, MD: Hospital de Pediatría Garrahan, Buenos Aires, Argentina; <sup>23</sup>Luca Villa, MA, Mariangela Rinaldi, MEng, Nicolino Ruperto, MD, MPH: Istituto Giannina Gaslini, Pediatria II - Reumatologia, PRINTO, Genoa, Italy; <sup>24</sup>Jiri Vencovsky, MD, PhD: Charles University, Prague, Czech Republic. See Appendix A for members of the International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation who contributed to developing the response criteria.

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

Address correspondence to Nicolino Ruperto, MD, MPH, Istituto Giannina Gaslini, Pediatria II, PRINTO, Via G. Gaslini 5, 16147 Genoa, Italy. E-mail: nicolaruperto@gaslini.org.

Juvenile dermatomyositis (DM) is a systemic autoimmune disease characterized by chronic skeletal muscle inflammation and weakness. Core set measures to assess juvenile DM 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 value and extramuscular global activity, whereas PRINTO includes instead a health-related quality of life measure, the Child Health Questionnaire (7) and a global activity score, the Disease Activity Score (8). IMACS measures muscle strength using manual muscle testing, and PRINTO measures muscle strength using 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 the sample sizes needed, 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 3 of 6 core set measures with no more than 1 or 2 worsening (which cannot be muscle strength) had been established as preliminary response criteria, and additional combinations of improvement in the core set measures serve as secondary response criteria (9,10). PRINTO adapted its top criteria for minimal clinical improvement to moderate and major improvement by using cutoffs of 50% and 70%, similar to the improvement criteria for juvenile idiopathic arthritis (JIA) (11–13).

Although the preliminary response criteria for juvenile DM advanced the assessment of patients and their responses to treatment, those criteria were limited by differences in the core set measures 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 (14). The preliminary response criteria also considered each core set measure equally rather than differentially weighting them. However, most myositis experts agree that some core set measures are more important, such as physician global activity and muscle strength (3,15). For PRINTO studies, physician global evaluation of disease activity, muscle strength, and parent global evaluation of the child's overall well-being were weighted as the most important core set measures in a logistic regression analysis (3,10). 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 the standardized criteria developed for rheumatoid arthritis (RA) and JIA.

For these reasons, IMACS and PRINTO engaged in a joint effort to develop fully validated response criteria for juvenile DM, including criteria for minimal, moderate, and major clinical response. This report focuses on the consensus conference in which the top candidate definitions of response leading to the final juvenile DM response criteria were considered.

## Methods

In previous reports (16,17), we described the methodology used a) to create patient profiles using natural history data and obtain expert consensus on minimal, moderate, and major improvement (16), b) to determine differential weights of the core set measures using conjoint analysis, and c) to draft 6 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 (19–22). In the conjoint analysis surveys administered using 1000Minds online software (23), experts were presented with pairs of hypothetical patient scenarios; each patient had different levels of improvement in the same 2 core set measures, assuming other core set measures remained the same. Experts rated which of the 2 scenarios had greater improvement. Based on the rater's response, the relative weights of core set measures and their levels of improvement were established and used to develop a scoring system by mathematical methods based on linear programming (24) such that when all 6 core set measures are considered together, the maximum score (total improvement score) possible for representing a patient's improvement is 100, and the minimum score is 0.

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 (18). 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 a 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 both  $\geq 80\%$ , AUC  $\geq 0.9$  for minimal, and AUC  $\geq 0.8$  for moderate and major improvement, using IMACS or PRINTO core set measures [1]) were externally

validated. The PRINTO trial randomized patients with new-onset juvenile DM to receive prednisone alone ( $n = 47$ ) or prednisone combined with methotrexate or cyclosporine ( $n = 46$  patients per treatment arm) (13). Chi-square analysis was used to compare the percentages of patients meeting the candidate definitions for response at the primary end point (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 core set measures were available in this trial.

A second trial validation data set included 48 juvenile DM patients enrolled in the Rituximab in Myositis (RIM) trial for treatment-refractory patients. It had a randomized placebo-phase design in which patients received either rituximab or placebo at weeks 0 and 1, and at weeks 8 and 9 their treatment assignment was reversed in a blinded manner (25). 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 that in the PRINTO trial. For the RIM trial, only the IMACS core set measures were available.

We then selected the top candidate definitions, up to 4 top-performing definitions from each of the 6 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 (LGR and NR, 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 juvenile DM. 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 core set measures, including sensitivity, specificity, AUC, as well as kappa values and odds ratios. AUC was defined as the average of the sensitivity and specificity values for all categorical candidate definitions, as well as for thresholds for 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 (ROC) curve as a plot of sensitivity versus ( $1 - \text{specificity}$ ) for total improvement scores as well as for thresholds (26–28). Results of the external validation for each candidate definition from the PRINTO and RIM clinical trial data sets 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 in juvenile DM. The participants were informed of the secondary goal of reaching consensus on response criteria for both juvenile DM and adult DM/polymyositis (PM). Participants were first asked to rank their top 5 choices, considering the data presented, based on face validity, feasibility, and

**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. Three of any 6 improved by $\geq 20\%$ , no more than 1 worse by $>30\%$ (which cannot be CMAS) (10) Moderate. Three of any 6 improved by $\geq 50\%$ , no more than 1 worse by $>30\%$ (which cannot be CMAS) (11) Major. Three of any 6 improved by $\geq 70\%$ , no more than 1 worse by $>30\%$ (which cannot be CMAS) (11)
Newly drafted (categorical definition)	Drafted relative or absolute percent change in candidate definitions of response, based on recent CSM survey	Minimal. MD global, muscle strength (MMT or CMAS), and 1 other CSM improved by $\geq 20\%$ Moderate. MD global, muscle strength (MMT or CMAS), and 1 other CSM improved by $\geq 30\%$ Major. MD global, muscle strength (MMT or CMAS), and 1 other CSM improved by $\geq 50\%$
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 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, C-HAQ = 1.5 points, extramusc/DAS = 1.5 points, enzyme/CHQ-PhS = 1 point Minimal. Improvement points given when CSM $\geq 20\%$ ; worsening points given when CSM worse by $>30\%$ Moderate. Improvement points given when CSM $\geq 50\%$ ; worsening points given when CSM worse by $>30\%$ Major. Improvement points given when CSM $\geq 75\%$ ; 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 % change	Improvement score = (MD global % change) + 0.5 $\times$ (parent global activity % change) + 0.5 $\times$ (extramusc activity or DAS % change) Minimal. Improvement score $\geq 15$ Moderate. Improvement score $\geq 30$ Major. Improvement score $\geq 60$
Core set measure-weighted (continuous definition)†	Multiply the % change in each CSM by the weights derived from conjoint analysis, then sum (% change in each CSM $\times$ 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 $\times$ (MD global % change) + (parent global % change) + 3 $\times$ (MMT or CMAS % change) + 1.5 $\times$ (C-HAQ % change) + 1.5 $\times$ (extramusc or DAS % change) + (enzyme or CHQ-PhS % change) Minimal. Improvement score $\geq 100$ Moderate. Improvement score $\geq 250$ Major. Improvement score $\geq 400$
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 the improvement score is above the cutoff for minimal improvement; similarly for moderate and major improvement	Cut points for the model for juvenile DM are: Minimal. Improvement score $\geq 30$ Moderate. Improvement score $\geq 45$ Major. Improvement score $\geq 70$ ‡

\* CMAS = Childhood Myositis Assessment Scale; CSM = core set measure; MD global = physician global activity score; MMT = manual muscle testing; parent global = parent global activity score; C-HAQ, 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.

‡ The full absolute percent change model is shown in Table 3 and in Supplementary Table 2 (available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract>).

generalizability, and to determine which response criteria were most clinically meaningful. The voting process was conducted in a systematic manner with a predetermined format using nominal group technique (29,30) facilitated by an internet-based system developed by staff at the PRINTO coordinating center (31,32). 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 his or her top- and bottom-ranked choices to the group. The rounds of voting continued in the same manner until consensus was reached ( $\geq 80\%$  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 core set measures and response criteria for juvenile DM that would interchange both the IMACS and PRINTO measures. Participants also voted on retesting the performance of the top candidate response criteria in future trials.

**Combined pediatric and adult working group.** After consensus was attained for juvenile DM 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 juvenile DM and adult DM/PM. Common response criteria that would include both juvenile DM and adult DM/PM were considered for use in clinical trials, which might facilitate drug approvals for myositis treatment. Experienced moderators (LGR, RA, FWM, and NR) 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 his or her top-choice candidate definition, using nominal group technique in a round-robin manner. At each round, participants were asked to select only 1 candidate top response criteria set; discussion was stopped once consensus of  $\geq 80\%$  was reached. For determining the thresholds of improvement for the selected definition, the required consensus was  $\geq 70\%$ , which was done by post-conference voting.

## Results

The performance characteristics of 101 of 312 candidate definitions were excellent (sensitivity and specificity of  $\geq 80\%$ , AUC  $\geq 0.90$  for minimal improvement), and 30 candidate definitions also performed well in 2 clinical trials, in which they differentiated between treatment arms ( $P < 0.05$  for minimal improvement) and differentiated the treating physician's improvement score at week 24 ( $P < 0.001$ ) (15).

### 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 Tables 1 and 2, available on the *Arthritis & Rheumatology* web site at [\[onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract\]\(http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract\)\). These candidate criteria included 9 categorical definitions in which different criteria were set for minimal, moderate, and major improvement and 5 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 9 categorical definitions, 2 were previously published IMACS and PRINTO response criteria \(9–11\), 4 were newly drafted definitions based on a survey of experts, and 3 were weighted definitions. Among the continuous definitions, 2 were developed by logistic regression, and 3 were developed from the conjoint analysis survey. Among the 14 candidate criteria considered, 11 were based on relative percent change, and 3 were based on absolute percent change in the core set measures.](http://</a></p></div><div data-bbox=)

The performance characteristics of these 14 candidate definitions are shown in Table 2 and Supplementary Table 1 (available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract>). In the patient profiles, with expert consensus as a gold standard, all definitions presented at the conference had sensitivity and specificity of  $\geq 87\%$  (AUC  $\geq 0.90$ ) for minimal improvement (Table 2 and Supplementary Table 1). For moderate improvement, specificity decreased but was  $\geq 80\%$  (AUC  $\geq 0.88$ ), and for major improvement specificity was generally  $\geq 75\%$  (AUC  $\geq 0.84$ ). For continuous definitions, the AUCs (from ROC curves) for the total improvement score were generally better than the AUCs (average of sensitivity and specificity) for the thresholds of minimal, moderate, and major improvement. Performance was similar between the IMACS and PRINTO core set measures for each definition.

Almost all candidate criteria were validated using the PRINTO trial at 6 months, when 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 juvenile DM patients in the RIM trial (25). All definitions could differentiate the median treating physician's improvement score at week 24 ( $P \leq 0.006$ ).

**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 3, 6 candidate definitions were chosen, each receiving a similar number of votes. These 6 included the 3 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 criteria (8,9). In the fourth round of voting and discussion, participants reached consensus

**Table 2.** Detailed performance characteristics of patient profiles for the top 5 candidate definitions presented at the consensus conference\*

Candidate definition type based on final consensus rank order, improvement category, core set measure	RIM trial††											
	PRINTO trials§			Total improvement score AUC‡		Specificity, %		Threshold AUC†		Response criteria, not improved#		Rank
	Tx (%)	Ctrl (%)	P	improvement score	AUC‡	%	AUC†	improved#	Response criteria, not improved#	P		
Conjoint analysis, absolute % change (model 3)**												1
Minimal ( $\geq 30$ )												
IMACS	89	91	0.90	0.98	75	53	0.009	2.0	3.0	<0.001		
PRINTO	89	98	0.93	0.99	73	55	0.038					
Moderate ( $\geq 45$ )												
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 ( $\geq 70$ )												
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					
Conjoint analysis, relative % change (model 1)††												2
Minimal ( $\geq 33$ )												
IMACS	99	87	0.93	0.98	75	55	0.018	2.0	4.0			
PRINTO	96	98	0.97	1.00	74	55	0.027					
Moderate ( $\geq 60$ )												
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 ( $\geq 80$ )												
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					
Conjoint analysis, relative % change (model 2)††												3
Minimal ( $\geq 33$ )												
IMACS	95	94	0.94	0.98	75	53	0.009	2.0	4.0	<0.001		
PRINTO	94	98	0.96	0.99	74	55	0.027					
Moderate ( $\geq 55$ )												
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 ( $\geq 77$ )												
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					
Weighted definition, relative % change‡‡												4
Minimal (improvement points given when CSM $\geq 20$ , worsening points given when CSM worse by $>30$ )												
IMACS	95	100	0.97	NA	70	51	0.032	2.0	3.0	<0.001		
PRINTO	92	98	0.95	NA	73	53	0.021					
Moderate (improvement points given when CSM $\geq 50\%$ , worsening points given when CSM worse by $>30\%$ )												
IMACS	95	91	0.93	NA	68	51	0.045	2.0	3.0	<0.001		
PRINTO	95	92	0.94	NA	71	51	0.023					
Major (improvement points given when CSM $\geq 75\%$ , worsening points given when CSM worse by $>30\%$ )												
IMACS	100	81	0.91	NA	64	47	0.050	1.5	3.0	<0.001		
PRINTO	98	85	0.91	NA	62	49	0.142					

Table 2. (Cont d)

Candidate definition type based on final consensus rank order, improvement category, core set measure	Sensitivity, %	Specificity, %	Threshold AUC†	Total improvement score AUC‡	PRINTO trials§		RIM trial¶		Rank	
					Tx (%)	Ctrl (%)	Response criteria, improved#	Response criteria, not improved#		
										P
Previously published definition (10,11), relative % change									5	
Minimal (3 of any 6 improved by ≥20%, no more than 1 worse by >30%) (which cannot be MMT/CMAS) (10)	93	100	0.97	NA	70	51	0.032	2.0	3.0	<0.001
PRINTO	88	100	0.94	NA	71	51	0.023			
Moderate (3 of any 6 improved by ≥50%, no more than 1 worse by >30%) (which cannot be MMT/CMAS) (11)	90	95	0.93	NA	66	51	0.081	2.0	3.0	<0.001
PRINTO	90	96	0.93	NA	68	51	0.045			
Major (3 of any 6 improved by ≥70%, no more than 1 worse by >30%) (which cannot be MMT/CMAS) (11)	99	83	0.91	NA	63	49	0.111	2.0	3.0	<0.001
PRINTO	99	89	0.94	NA	60	49	0.223			

\* The performance characteristics of patient profiles for definitions ranked 6-14 are shown in Supplementary Table 1 (available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract>). Note that either International Myositis Assessment and Clinical Studies (IMACS) or Paediatric Rheumatology International Trials Organisation (PRINTO) core set measures (CSMs) may be used in these candidate definitions of response; the candidate definitions were developed in parallel with IMACS or PRINTO CSMs. Tx = treatment arm of prednisone in combination with methotrexate or cyclosporine; Ctrl = control; NA = not applicable.

† Calculated as the area under the curve (AUC) from the receiver operating characteristic (ROC) curve for the total improvement score and the threshold for minimal, moderate, and major improvement.

‡ Calculated as the AUC from the ROC curve, using the total improvement score and the threshold cutoffs for minimal, moderate, and major improvement, which applies only to continuous definitions.

§ PRINTO juvenile dermatomyositis (DM) trial of prednisone alone versus prednisone with methotrexate or cyclosporine (n = 139) (13).

¶ Rituximab in Myositis (RIM) trial juvenile DM arm (n = 48). Comparison of the treating physician's rating of improvement if the improvement criteria are met versus not met at week 24 (25). A 1-point difference in physician's rating of improvement from no improvement to minimal improvement was considered not only statistically significant but also clinically significant.

# Median score for physician's rating of improvement.

\*\* The conjoint analysis-based continuous candidate response criteria using absolute percent change in core set measures (absolute percent change model) are shown in Table 3. These criteria are also the top response criteria for adult DM/polymyositis (PM), but with different thresholds for the total improvement score for minimal, moderate, and major improvement (18).

†† The conjoint analysis-based continuous candidate definitions using relative percent change in core set measures are shown in Supplementary Table 3 (available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract>). These criteria are also the second- and third-choice criteria for adult DM/PM, but with different thresholds in the total improvement score for minimal, moderate, and major improvement (18).

‡‡ Improvement = at least 3.5 improvement points, and no more than 1.5 worsening points, where physician global activity = 2 points, parent global activity = 1 point, manual muscle testing (MMT) or Childhood Myositis Assessment Scale (CMAS) = 3 points, Childhood Health Assessment Questionnaire = 1.5 points, extramuscular global activity or Disease Activity Score = 1.5 points, and enzyme or physical summary score of the Child Health Questionnaire-Parent Form 50 = 1 point.

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

Core set measure, level of improvement based on absolute percent change	Improvement score
Physician global activity	
Worsening to 5% improvement	0
>5% to 15% improvement	7.5
>15% to 25% improvement	15
>25% to 40% improvement	17.5
>40% improvement	20
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
Manual muscle testing 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
Childhood Health Assessment Questionnaire	
Worsening to 5% improvement	0
>5% to 15% improvement	5
>15% to 25% improvement	7.5
>25% to 40% improvement	7.5
>40% improvement	10
Enzyme (most abnormal) or CHO-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

The **total improvement score** is the sum of all 6 improvement scores associated with the change in each core set measure. A total improvement score of  $\geq 30$  represents **minimal improvement**, a score of  $\geq 45$  represents **moderate improvement**, and a score of  $\geq 70$  represents **major improvement**.

\* Either all of the International Myositis Assessment and Clinical Studies Group (IMACS) or all of the Paediatric Rheumatology International Trials Organisation (PRINTO) core set measures may be used. Note that these response criteria are also proposed for use in combined adult DM/polymyositis (DM/PM) and juvenile DM trials (18). For comparison, the thresholds of improvement in the total improvement score for adult DM/PM are  $\geq 20$  for minimal improvement,  $\geq 40$  for moderate improvement, and  $\geq 60$  for major improvement.

**How to calculate the improvement score:** The absolute percent change ( $[(\text{final value} - \text{baseline value})/\text{range}] \times 100$ ) is calculated for each core set measure. For muscle enzymes, the most abnormal serum muscle enzyme level at baseline (creatinine kinase, aldolase, alanine transaminase, aspartate aminotransferase, lactate dehydrogenase) is used. The enzyme range was calculated based on a 90% range of enzymes from natural history data (5,38), 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.<sup>†</sup> The ULN is determined according to the individual laboratories in the participating centers. The ranges for the other core set activity measures are based on the instrument scale used (13,15,25). An improvement score is assigned for each core set measure based on the absolute percent change. These are totaled among the 6 IMACS or PRINTO core set measures. The thresholds for minimal, moderate, and major improvement are provided. The total improvement score itself may also be compared among treatment arms in a trial. A total improvement score between 0 and 100 corresponds to the degree of improvement, with higher scores corresponding to a greater degree of improvement. CMAS = Childhood Myositis Assessment Scale; CHO-PhS=Physical Summary Score of the Child Health Questionnaire-Parent Form 50.

on final top response criteria, a conjoint analysis-based continuous model using absolute percent change in the IMACS or PRINTO core set measures (Table 3).

Table 2 and Supplementary Table 1 (available on the *Arthritis & Rheumatology* web site at <http://online-library.wiley.com/doi/10.1002/art.40060/abstract>) show the

<sup>†</sup>Correction added after online publication 12 February 2018: The information on calculation of enzyme ranges in this sentence has been corrected.



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 core set measures, the sensitivity and specificity in the patient profiles was generally >90% and the 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 core set measures, which experts thought were consistent with their assessment of the relative importance of each of the core set measures. The top response criteria were based on absolute percent change in core set measures, which was also favored by the participants because, given the various visual analog scale (VAS) measurements used in the core set measures, the absolute percent changes were more congruent than relative percent changes with actual changes that the myositis experts see in clinical practice.

**Final response criteria chosen by the combined pediatric and adult working group.** For this round of votes, the top 2 pediatric (Table 2) and adult definitions (18) were considered. Two rounds of voting resulted in final consensus response criteria, with 91% of participants voting for the conjoint analysis-based continuous response criteria based on absolute percent change in the core set measures (Table 3). It was agreed that the top response criteria would be used in future clinical trials that combined juvenile DM and adult DM/PM. Because the final response criteria were similar, participants favored using response criteria that would be common to juvenile DM 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 using the Delphi method, 74% of the participants agreed to use the following pediatric threshold values for minimal, moderate, and major response in juvenile DM: total improvement score  $\geq 30$  (on a scale of 0–100) for minimal,  $\geq 45$  for moderate, and  $\geq 70$  for major improvement. In contrast, the final

thresholds for minimal, moderate, and major response in adult DM/PM were  $\geq 20$ ,  $\geq 40$ , and  $\geq 60$ , respectively. The pediatric working group also reached consensus that, given the overall similarity between the IMACS and PRINTO response criteria, joint IMACS/PRINTO response criteria for juvenile DM are being proposed. The current development of the response criteria in parallel between the IMACS and PRINTO core set measures necessitates that either all of the IMACS or all of the PRINTO core set measures be used. The pediatric experts, however, committed to measure both IMACS and PRINTO core set measures in future therapeutic trials, with 92% agreement, and to continue to test the interchangeability of the IMACS and PRINTO core set measures. The group also unanimously agreed to retest the validity of the top 5 candidate definitions for response criteria and to utilize the other 4 definitions as secondary end points in future clinical trials. The top 3 of these criteria, the conjoint analysis definitions, are the same for both juvenile DM and adult DM/PM, with different thresholds of improvement (Table 3 and Supplementary Table 3, available on the *Arthritis & Rheumatology* web site at <http://onlinelibrary.wiley.com/doi/10.1002/art.40060/abstract>).

## Discussion

Conjoint analysis-based continuous response criteria, based on absolute percent change in the core set measures, were developed as the consensus- and data-driven response criteria for minimal, moderate, and major improvement in juvenile DM. For the response criteria, either IMACS or PRINTO core set measures could be used. In addition, it was agreed that the same response criteria, using the IMACS core set measures but with different thresholds for improvement, would be the consensus response criteria for adult DM/PM trials and combined juvenile DM and adult DM/PM trials in the future (18).

The comprehensive process used to develop final response criteria for minimal, moderate, and major improvement in juvenile DM included the use of large, prospective, natural history data sets for juvenile DM and data from 2 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 (13,15,25). The involvement of many clinical experts who had experience using the core set measures in juvenile DM patients was also critical. They provided input at several points throughout the process, including determining thresholds for improvement in core set measures by which definitions of response were drafted, achieving gold standard ratings of improvement by evaluating and developing

consensus patient profiles, completing the conjoint analysis surveys to develop differential weights for the core set measures, and participating in the final consensus conference to achieve consensus for common response criteria with the greatest clinical face validity. The current response criteria (Table 3) also resolve the differences between PRINTO and IMACS core set measures by testing candidate definitions of response criteria in parallel using both sets of measures and showing 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 core set measures, have many advantages. For each measure, improvement points are calculated based on the level of change in that measure, and each core set measure is differentially weighted, such that changes in muscle strength and physician global activity are weighted more heavily than changes in the most abnormal enzyme value 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 (33). 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 (33,34). 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 core set measures instead of relying on only a few measures to determine whether a patient has experienced improvement. However, although these response criteria were developed using all 6 core set measures, the response criteria could still be used if fewer core set measures 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 6 core set measures are used. As such, the response

criteria signify a major advance in assessing improvement in therapeutic trials and other clinical research studies by providing data-driven response criteria that 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 (35,36), 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 core set measures used a 10-cm VAS, and the experts thought 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 core set measure is 0 and give similar results for similar degrees of change at either end of the VAS.

The participants also favored using the same response criteria for juvenile DM 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 (25), 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 end points for future clinical trials. Among these were not only other conjoint analysis-based continuous models but also the published PRINTO preliminary response criteria (10,11). Future work should also evaluate whether a baseline composite score threshold derived from the PRINTO or IMACS core set measures 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 juvenile DM that had 3 treatment arms to use for external validation (13), in which we evaluated the ability of the candidate definitions to differentiate between treatment arms. Although thresholds for major improvement were developed and validated in fewer patients, we believe that 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 core set measures; however, this generally does not affect the outcome, because when patients are rated as improved, no more than 1 or 2 measures worsen in our clinical data sets. Also, although we tested the interchange of IMACS and PRINTO core set measures, we tested these variations as

2 parallel core set measures but did not examine intermixing the PRINTO and IMACS core set measures. Further work to examine the interchangeability of the IMACS and PRINTO core set measures will be needed.

The data sets used to develop the new response criteria primarily contained information about patients with a recent diagnosis or those experiencing a disease flare, and further work is needed to determine how the response criteria perform in patients with longstanding disease or those with significant disease-related damage. Finally, although 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 manner in which the core set measures are evaluated (e.g., similar to the Juvenile Arthritis Disease Activity Score for JIA) (37). The time required to apply these criteria is estimated to be 25–35 minutes to complete the core set measures 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 for either RA or JIA, despite being regularly reported in clinical trials.

In summary, conjoint analysis-based continuous response criteria that establish different thresholds for minimal, moderate, and major improvement and utilize the absolute percent change in core set measures were chosen as the consensus response criteria for juvenile DM and were validated 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 worldwide. They should be sensitive in detecting differences in improvement and in quantitating the degree of improvement, as seen in the 2 clinical trials. Thus, clinical trials that test new therapies for juvenile DM should be easier to design, conduct, and compare.

#### ACKNOWLEDGMENTS

We thank the following individuals for providing invaluable input and feedback on project development and support: 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), Bob Goldberg and Theresa Curry (The Myositis Association), Rhonda McKeever and Patti Lawler (Cure JM Foundation), and Irene Oakley (Myositis UK). We also thank

Drs. Michael Ward and Steven Pavletic for their critical review of the manuscript. Paul Hansen, who with Franz Ombler owns and co-invented the 1000Minds software referred to in the article, provided intellectual and logistic support for this project.

#### AUTHOR CONTRIBUTIONS

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

**Study conception and design.** Rider, Aggarwal, Miller, Vencovsky, Ruperto.

**Acquisition of data.** Rider, Aggarwal, Pistorio, Bayat, Erman, Feldman, Huber, Cimaz, Cuttica, de Oliveira, Lindsley, Pilkington, Punaro, Ravelli, Reed, Rouser-Stevens, van Royen-Kerkhof, Dressler, Magalhaes, Constantin, Davidson, Magnusson, Russo, Villa, Rinaldi, Rockette, Lachenbruch, Vencovsky, Ruperto.

**Analysis and interpretation of data.** Rider, Aggarwal, Pistorio, Bayat, Erman, Feldman, Huber, Rockette, Lachenbruch, Miller, Vencovsky, Ruperto.

#### REFERENCES

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:S118–57.
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;40:1262–73.
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;30:603–17.
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;42:1452–9.
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;59:4–13.
6. Ruperto N, Martini A. Networking in paediatrics: the example of the Paediatric Rheumatology International Trials Organisation (PRINTO). *Arch Dis Child* 2011;96:596–601.
7. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis Rheum* 1994;37:1761–9.
8. Bode RK, Klein-Gitelman MS, Miller ML, Lechman TS, Pachman LM. Disease Activity Score for children with juvenile

- dermatomyositis: reliability and validity evidence. *Arthritis Rheum* 2003;49:7–15.
9. Rider LG, Giannini EH, Brunner HI, Ruperto N, James-Newton L, Reed AM, et al. International consensus on preliminary definitions of improvement in adult and juvenile myositis. *Arthritis Rheum* 2004;50:2281–90.
  10. Ruperto N, Pistorio A, Ravelli A, Rider LG, Pilkington C, Oliveira S, et al. The Paediatric Rheumatology International Trials Organization provisional criteria for the evaluation of response to therapy in juvenile dermatomyositis. *Arthritis Care Res (Hoboken)* 2010;62:1533–41.
  11. 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 multi-center PRINTO study. *Arthritis Rheum* 2011;63:3142–52.
  12. 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.
  13. Ruperto N, Pistorio A, Oliveira S, Zulian F, Cuttica R, Ravelli A, et al. Prednisone versus prednisone plus ciclosporin versus prednisone plus methotrexate in new-onset juvenile dermatomyositis: a randomised trial. *Lancet* 2016;387:671–8.
  14. Felson DT, Furst DE, Boers M. Rationale and strategies for reevaluating the ACR20. *J Rheumatol* 2007;34:1184–7.
  15. Rider LG, Ruperto N, Pistorio A, Erman B, Bayat N, Lachenbruch PA, et al. 2016 development of adult dermatomyositis and polymyositis and juvenile dermatomyositis response criteria: methodological aspects: an American College of Rheumatology/European League Against Rheumatism/International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. *Rheumatology (Oxford)* 2016. In press.
  16. 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 [abstract]. *Arthritis Rheumatol* 2014;66 Suppl:S578.
  17. Aggarwal R, Rider LG, Ruperto N, Bayat N, Erman B, Feldman BM, et al. A consensus hybrid definition using a conjoint analysis is the proposed response criteria for minimal and moderate improvement for adult polymyositis and dermatomyositis clinical trials [abstract]. *Arthritis Rheumatol* 2014;66 Suppl:S404.
  18. Aggarwal R, Rider LG, Ruperto N, Bayat N, Erman B, Feldman BM, et al. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: an International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation collaborative initiative. *Arthritis Rheumatol* 2017;69:898–910.
  19. Neogi T, Jansen TL, Dalbeth N, Fransen J, Schumacher HR, Berendsen D, et al. 2015 gout classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheumatol* 2015;67:2557–68.
  20. Utz KS, Hoog J, Wentrup A, Berg S, Lammer A, Jainsch B, et al. Patient preferences for disease-modifying drugs in multiple sclerosis therapy: a choice-based conjoint analysis. *Ther Adv Neurol Disord* 2014;7:263–75.
  21. 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.
  22. 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.
  23. De Lautour H, Taylor WJ, Adebajo A, Alten R, Burgos-Vargas R, Chapman P, et al. Development of preliminary remission criteria for gout using Delphi and 1000Minds consensus exercises. *Arthritis Care Res (Hoboken)* 2016;68:667–72.
  24. Hansen P, Omblor 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.
  25. Oddis CV, Reed AM, Aggarwal R, Rider LG, Ascherman DP, Levesque MC, et al. Rituximab in the treatment of refractory adult and juvenile dermatomyositis and adult polymyositis: a randomized, placebo-phase trial. *Arthritis Rheum* 2013;65:314–24.
  26. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283–98.
  27. Finney DJ. Statistical method in biological assay. 3rd ed. London: Charles Griffin; 1978.
  28. Liengme BV. A guide to Microsoft Excel 2002 for scientists and engineers. 3rd ed. Oxford: Butterworth-Heinemann; 2002.
  29. Ruperto N, Meiorin S, Iusan SM, Ravelli A, Pistorio A, Martini A. Consensus procedures and their role in pediatric rheumatology. *Curr Rheumatol Rep* 2008;10:142–6.
  30. 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.
  31. Ruperto N, Ozen S, Pistorio A, Dolezalova P, Brogan P, Cabral DA, et al. EULAR/PRINTO/PRES criteria for Henoch-Schonlein purpura, childhood polyarteritis nodosa, childhood Wegener granulomatosis and childhood Takayasu arteritis: Ankara 2008. Part I: overall methodology and clinical characterisation. *Ann Rheum Dis* 2010;69:790–7.
  32. Piram M, Kone-Paut I, Lachmann HJ, Frenkel J, Ozen S, Kummerle-Deschner J, et al. Validation of the auto-inflammatory diseases activity index (AIDAI) for hereditary recurrent fever syndromes. *Ann Rheum Dis* 2014;73:2168–73.
  33. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. *Arthritis Rheum* 2007;57:193–202.
  34. Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002;47:262–6.
  35. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727–35.
  36. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum* 1997;40:1202–9.
  37. Consolaro A, Ruperto N, Bazzo 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;61:658–66.
  38. Volochayev R, Csako G, Wesley R, Rider LG, Miller FW. Laboratory test abnormalities are common in polymyositis and dermatomyositis and differ among clinical and demographic groups. *Open Rheumatol J* 2012;6:54–63.

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

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

**Statistical team:** Rohit Aggarwal, Brian Erman, Nastaran Bayat, Angela Pistorio, Adam M. Huber, Brian M. Feldman, Paul Hansen, Howard Rockette, Peter A. Lachenbruch, Nicolino Ruperto, Lisa G. Rider.

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

**Clinical trial or natural history study data set contributors:** Frederick W. Miller, Chester V. Oddis, Ann Reed, Lisa G. Rider, Nicolino 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, Sheila Knupp Feitosa de Oliveira, Brian M. Feldman, Virginia Paes Leme Ferriani, Berit Flato, Valeria Gerloni, Thomas Griffin, Michael Henrickson, Claas Hinze, Mark Hoeltzel, Adam M. 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, Lauren M. Pachman, Murray Passo, Clarissa Pilkington, Marilyn Punaro, Pierre Quartier, Eglá Rabinovich, Athimalaipet V. Ramanan, Angelo Ravelli, Ann Reed, Robert Rennebohm, Lisa G. Rider, Rafael Rivas-Chacon, Angela Byun Robinson, Kelly Rouster-Stevens, 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, Annet van Royen-Kerkhof, Dawn Wahezi, Carol Wallace, Francesco Zulian.

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

**Participants in consensus conference, pediatric working group:** Rolando Cimaz, Rubén Cuttica, Sheila Knupp Feitosa de Oliveira, Brian M. Feldman, Adam M. Huber, Carol B. Lindsley, Clarissa Pilkington, Marilyn 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, Marianne de Visser, David Fiorentino, David Isenberg, James Katz, Andrew Mammen, Chester V. Oddis, Jiri Vencovsky, Steven R. Ytterberg.