

Original article

2016 ACR-EULAR adult dermatomyositis and polymyositis and juvenile dermatomyositis response criteria – methodological aspects

Lisa G. Rider^{1,*}, Nicolino Ruperto^{2,*}, Angela Pistorio³, Brian Erman⁴, Nastaran Bayat¹, Peter A. Lachenbruch¹, Howard Rockette⁵, Brian M. Feldman⁶, Adam M. Huber⁷, Paul Hansen⁸, Chester V. Oddis⁹, Ingrid E. Lundberg¹⁰, Anthony A. Amato¹¹, Hector Chinoy^{12,13}, Robert G. Cooper¹⁴, Lorinda Chung¹⁵, Katalin Danko¹⁶, David Fiorentino¹⁷, Ignacio García-De la Torre¹⁸, Ann M. Reed¹⁹, Yeong Wook Song²⁰, Rolando Cimaz²¹, Rubén J. Cuttica²², Clarissa A. Pilkington²³, Alberto Martini^{2,24}, Janjaap van der Net²⁵, Susan Maillard²³, Frederick W. Miller¹, Jiri Vencovsky^{26,†} and Rohit Aggarwal^{9,†}; the International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation

Abstract

Objective. The objective was to describe the methodology used to develop new response criteria for adult DM/PM and JDM.

Methods. Patient profiles from prospective natural history data and clinical trials were rated by myositis specialists to develop consensus gold-standard ratings of minimal, moderate and major improvement. Experts completed a survey regarding clinically meaningful improvement in the core set measures (CSM) and a conjoint-analysis survey (using 1000Minds software) to derive relative weights of CSM and candidate definitions. Six types of candidate definitions for response criteria were derived using survey results, logistic regression, conjoint analysis, application of conjoint-analysis weights to CSM and published definitions. Sensitivity, specificity and area under the curve were defined for candidate criteria using consensus patient profile data, and selected definitions were validated using clinical trial data.

Results. Myositis specialists defined the degree of clinically meaningful improvement in CSM for minimal, moderate and major improvement. The conjoint-analysis survey established the relative weights of CSM,

¹Environmental Autoimmunity Group, NIEHS, National Institutes of Health, Bethesda, MD, USA, ²Istituto Giannina Gaslini, Pediatria II - Reumatologia, PRINTO, ³Istituto Giannina Gaslini, Servizio di Epidemiologia e Biostatistica, Genoa, Italy, ⁴Social and Scientific Systems, Inc., Durham, NC, ⁵Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA, ⁶Division of Rheumatology, Hospital for Sick Children, Toronto, Ontario, ⁷Rheumatology Department, Izaak Walton Killam Health Centre, Halifax, Nova Scotia, Canada, ⁸Department of Economics, University of Otago, Dunedin, New Zealand, ⁹Department of Medicine, Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, PA, USA, ¹⁰Rheumatology Unit, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden, ¹¹Department of Neurology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, ¹²National Institute of Health Research Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals NHS Foundation Trust, ¹³Manchester Academic Health Science Centre, The University of Manchester, Manchester, ¹⁴Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, UK, ¹⁵Division of Rheumatology, Stanford University, Redwood City, CA, USA, ¹⁶3rd Department of Internal Medicine, Division of Immunology, University of Debrecen, Debrecen, Hungary, ¹⁷Department of Dermatology, Stanford University, Redwood City, CA,

USA, ¹⁸Hospital General de Occidente de la Secretaría de Salud, Guadalajara, Jalisco, Mexico, ¹⁹Department of Pediatrics, Duke University, Durham, NC, USA, ²⁰Department of Molecular Medicine and Biopharmaceutical Sciences, Medical Research Center, Seoul National University, Seoul, Korea, ²¹Pediatric Rheumatology, Azienda Ospedaliero Universitaria Meyer, University of Florence, Florence, Italy, ²²Department of Pediatric Rheumatology, Hospital de Niños Pedro de Elizalde, University of Buenos Aires, Buenos Aires, Argentina, ²³Department of Paediatric Rheumatology, Great Ormond Street Hospital for Children NHS Trust, London, UK, ²⁴Università degli Studi di Genova, Dipartimento di Pediatria, Genoa, Italy, ²⁵Wilhelmina Children's Hospital, University Medical Center, Utrecht, The Netherlands and ²⁶Institute of Rheumatology and Department of Rheumatology, Charles University, Prague, Czech Republic

Submitted 27 September 2016; revised version accepted 17 May 2017

* Lisa G. Rider and Nicolino Ruperto contributed equally to this study

† Jiri Vencovsky and Rohit Aggarwal contributed equally to this study

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

with muscle strength and Physician Global Activity as most important. Many candidate definitions showed excellent sensitivity, specificity and area under the curve in the consensus profiles. Trial validation showed that a number of candidate criteria differentiated between treatment groups. Top candidate criteria definitions were presented at the consensus conference.

Conclusion. Consensus methodology, with definitions tested on patient profiles and validated using clinical trials, led to 18 definitions for adult PM/DM and 14 for JDM as excellent candidates for consideration in the final consensus on new response criteria for myositis.

Key words: dermatomyositis, polymyositis, juvenile dermatomyositis, response criteria, 1000Minds software, conjoint analysis, hybrid measure, outcome assessment

Rheumatology key messages

- New criteria for minimal, moderate and major response in adult and juvenile myositis were developed.
- The criteria can be used to conduct more efficient drug trials for myositis.
- The criteria are data driven and expert consensus driven, with excellent sensitivity, specificity and face validity for myositis.

Introduction

Myositis is a heterogeneous, systemic autoimmune disease requiring multiple outcome measures and composite response criteria for meaningful clinical trials and therapeutic development. Myositis core set activity measures (CSM) for patients with adult DM/PM or JDM were established and validated by the International Myositis Assessment and Clinical Studies Group (IMACS) and the Paediatric Rheumatology International Trials Organisation (PRINTO) [1, 2]. The IMACS CSM include Physician and Patient/Parent Global Activity, muscle strength measured by manual muscle testing (MMT), physical function measured by the (Childhood) HAQ (CHAQ or HAQ), Extramuscular Global Activity measured using Myositis Disease Activity Assessment Tool and the most abnormal serum muscle enzyme (supplementary Table S1, available at *Rheumatology* Online) [1, 3]. The PRINTO CSM are similar but include the Childhood Myositis Assessment Scale instead of MMT, the global DAS instead of Extramuscular Global Activity and health-related quality of life (Physical Summary Score of the Child Health Questionnaire, but not muscle enzymes) [4, 5]. A clinically meaningful degree of change in each measure was established previously using a data-driven consensus process [6].

Preliminary definitions of improvement were developed through consensus ratings of patient profiles from natural history studies and therapeutic trials [7, 8]. Those definitions required different combinations of improvement in CSM above certain thresholds. However, they needed validation because they were developed with few patients, inadequate trial data and some retrospective data, and did not include validated criteria for moderate or major response [9, 10]. Newer methodology using continuous or hybrid measures of response, such as conjoint analysis, which was used to develop classification criteria for other rheumatic diseases [11–15], may increase the sensitivity and specificity of

response criteria. Trial validation is also now required for final response criteria according to ACR and EULAR guidelines [16].

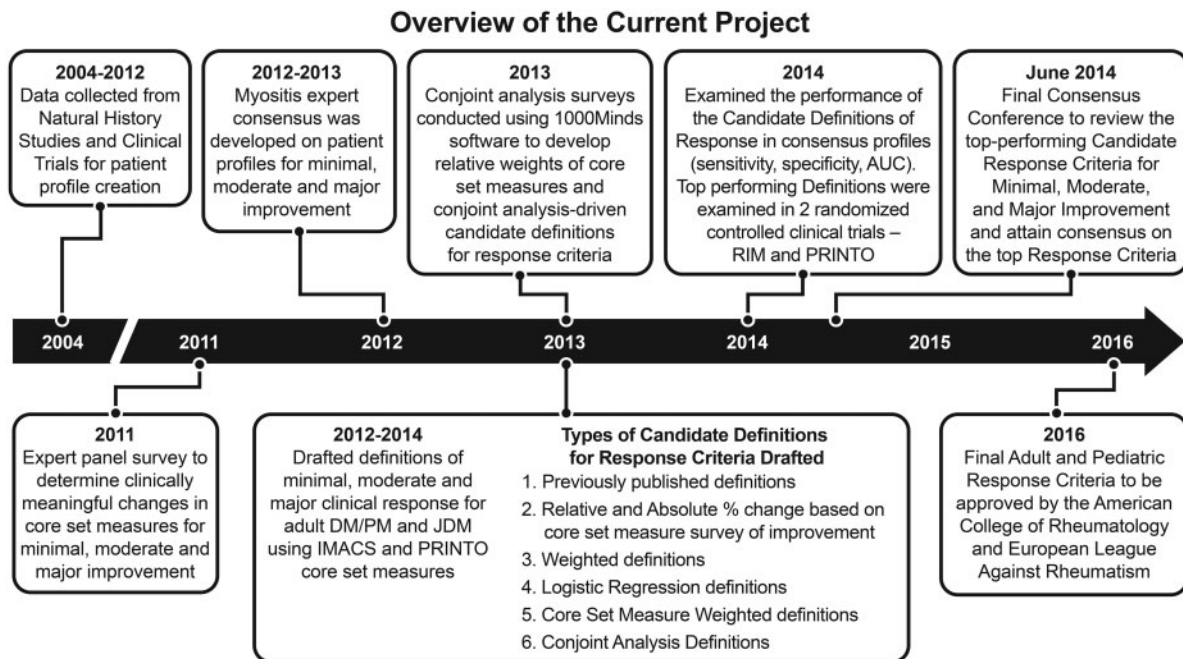
Our objective was to use data-driven and expert group decision-making (consensus) processes to develop new response criteria for DM/PM and JDM based on the six CSM of IMACS or PRINTO with data from new cohorts and clinical trials.

Methods

Overview

Adult and juvenile myositis experts with experience in using the CSM participated in expert working groups (supplementary Table S2, available at *Rheumatology* Online). They rated clinically meaningful improvement in the CSM for minimal, moderate and major improvement. Next, patient profiles were developed from natural history data and clinical trials [4, 5, 8] (supplementary Fig. S1, available at *Rheumatology* Online). Experts then rated the degree of improvement in patient profiles using the Delphi technique [17, 18]; $\geq 70\%$ consensus was reached on minimal, moderate and major improvement. Conjoint-analysis surveys (using 1000Minds software) [11] were completed by a subset of experts, and relative weights of CSM and conjoint analysis-based candidate definitions for response criteria were derived. Sensitivity, specificity and area under the curve (AUC) were derived for each candidate definition by using experts' consensus ratings of improvement from patient profiles as a gold standard. The best-performing candidate definitions were validated using data from randomized clinical trials, and top candidate definitions were taken to the consensus conference to develop new myositis response criteria for PM/DM and JDM [10, 19]. The overview and time line of the methodology are given in Fig. 1.

Fig. 1 Overview of the development of the new response criteria



Clinically meaningful improvement in myositis CSM

Adult and juvenile myositis specialists completed an online Delphi expert survey regarding the degree of change in each CSM deemed clinically significant, including the relative percentage change (defined as final minus baseline value, divided by baseline, multiplied by 100) and the absolute change deemed necessary for minimal, moderate and major clinical improvement in a therapeutic trial. The survey also assessed the minimal number of, and which, CSM must improve and the maximal number of, and which, CSM can worsen for a patient to be considered improved [20].

Adult and paediatric patient profiles

Six natural history studies and five trials were used to create 270 adult DM/PM patient profiles consisting of IMACS CSM (supplementary Table S3, available at *Rheumatology* Online) [21–25] and representing a spectrum of from no improvement to major clinical improvement [7, 8]. The Bohan and Peter [26, 27] criteria were used for classification. The PRINTO prospective cohort and IMACS natural history study were used to create 299 JDM patient profiles [9, 10]. For each JDM patient, two profiles were created, one using IMACS and one using PRINTO CSM. These IMACS and PRINTO profiles were administered separately to the paediatric experts in a randomized order. For each profile, the CSM at baseline and at 4–12 months' follow-up, the absolute change and relative percentage change in each measure were presented. Raters scored each profile as unchanged or worse, minimally improved, moderately improved or majorly improved, and rated the degree of improvement on a

scale of 0–10 (supplementary Fig. S1, available at *Rheumatology* Online). Profiles that did not reach consensus were re-rated, and those without consensus after two rounds were discarded. Informed patient consents and institutional review board approvals were obtained for the original studies and trial by their respective investigators (supplementary Table S3, available at *Rheumatology* Online). In this study, only the de-identified data were used for development and validation of the response criteria.

Derivation of candidate definitions

Candidate definitions for response criteria were derived and tested for sensitivity, specificity and AUC, using the experts' consensus ratings of patient profiles as the gold standard (supplementary Table S4, available at *Rheumatology* Online). Definitions were examined using relative and absolute percentage change (final minus baseline divided by the range, multiplied by 100) in CSM. Supplementary Table S4, available at *Rheumatology* Online, shows the types of candidate definitions. Three types of candidate definitions were categorical, which included previously published, newly drafted definitions using Delphi expert survey and weighted definitions. The previously published and newly drafted categorical candidate definitions required a certain number of CSM to improve by varying thresholds while allowing up to two CSM to worsen by certain thresholds. The weighted definitions were derived by applying relative weights to the categorical definitions.

Three other types of candidate definitions were continuous definitions, where total improvement scores are generated with thresholds for minimal, moderate and major

improvement. These included logistic regression, CSM-weighted and conjoint analysis-based definitions (supplementary Table S4, available at *Rheumatology* Online). Continuous candidate definitions were hybrid definitions, where the same definition can be used either as a continuous outcome measure using the total improvement score or as a categorical outcome measure using the thresholds for minimal, moderate and major improvement. The logistic regression-derived definitions were developed using randomly selected patient profile data (two-thirds of the data set) and standard logistic regression modelling. Expert consensus rating of minimal, moderate and major improvement was the dependent variable (gold standard), and relative or absolute percentage changes in each CSM were independent variables. The cut-off for an optimal sensitivity-specificity trade-off for improved vs not improved was selected based on the Youden index [28]. Each model underwent internal validation using the remaining third of the data set and external validation using clinical trial data.

Weighted CSM candidate definitions were derived by multiplying the relative or absolute percentage improvement or worsening in each CSM by its relative weight (developed in the conjoint-analysis surveys) and then summing the results of all CSM as total improvement score.

Conjoint-analysis surveys and candidate definitions

Conjoint-analysis surveys were administered to subgroups of myositis experts using 1000Minds online software [11]. Experts were presented with pairs of hypothetical patients; each patient had different levels of improvement in the same two CSM, assuming other CSM remained the same. Experts chose which patient had greater improvement, that both had equal improvement, or that the scenario was not possible (supplementary Fig. S2, available at *Rheumatology* Online). Based on the rater's response, all other hypothetical patients that could be pairwise ranked were eliminated via the property of transitivity, thereby significantly reducing the number of scenarios presented. The Potentially All Pairwise Rankings of All Possible Alternatives (PAPRIKA) method determined the mean weights and relative importance of the CSM, using mathematical methods based on linear programming [11].

Adult PM/DM conjoint-analysis surveys were administered using pairs of patients representing different degrees of improvement in IMACS CSM, whereas two separate surveys representing the same patients were administered for JDM using IMACS and PRINTO CSM. For each CSM, the total potential range of the relative percentage improvement was divided into five levels, ensuring that the patient profile data were distributed among them. Six CSM with five levels resulted in an average of 85 pairwise-ranking questions.

Results from the conjoint-analysis surveys resulted in development of six conjoint-analysis candidate definitions that combine continuous improvement scores and categorical threshold criteria into a single definition (therefore

called a hybrid definition) [29]. Relative weights of CSMs and their levels of improvement were used to develop a scoring system such that when all six CSMs are considered together, the maximal score possible for representing a patient's improvement is 100 and the minimal score is 0. The performance of each candidate criterion was evaluated by using randomly selected two-thirds of the patient profile data and by using consensus ratings for minimal, moderate or major improvement as the gold standard. Appropriate improvement thresholds in adult and paediatric patients were selected for each conjoint-analysis definition based on optimal sensitivity and specificity (using the Youden index [28]). We simplified the points given to each level of improvement in CSM by rounding to the nearest multiple of 2.5, and then we re-evaluated the sensitivity, specificity and AUC. These thresholds for minimal, moderate and major improvement were then validated on the remaining third of the patient profile data and further validated using clinical trial data. The thresholds differed for adult PM/DM and JDM for each candidate definition.

External validation

The candidate definitions that performed well using consensus profile data were externally validated using data from two controlled trials. The Rituximab in Myositis (RIM) trial [23], which was used for adult DM/PM patients (n=142), was a randomized double-blind placebo-phased trial where one group received active treatment early (week 0) and the other group received treatment late (week 8); primary outcome was time to improvement. As the trial did not discriminate between treatment arms, we were unable to determine whether the candidate definitions differentiated between early vs late treatments. The treating physician's rating of improvement (0–7 scale) at 24 weeks in the RIM trial was used for validation, and each point change in physician rating was considered clinically significant. The Mann-Whitney *U*-test was used to evaluate whether each candidate definition could differentiate between the treating physician's median rating of improvement, if the improvement criteria were met compared with not met.

For JDM, data from 48 patients in the RIM trial were used with IMACS CSM, as described for adults. We also used data from 139 patients in a PRINTO trial of prednisone alone vs prednisone combined with either MTX or CSA, which used both IMACS and PRINTO CSM [19]. For the PRINTO trial, the ability of a candidate definition to differentiate between treatment arms at 24 weeks was evaluated using a χ^2 test to compare the frequency of patients meeting the candidate response criteria between the two treatment arms. The candidate definitions for JDM had to meet performance standards for both IMACS and PRINTO measures.

After external validation, we selected top candidate definitions for response criteria to be evaluated by expert consensus at the conference held in Paris on 9–10 June 2014. Adult and paediatric expert groups ranked them

using a nominal group technique to achieve consensus on final adult and paediatric response criteria [30, 31].

Results

Clinically meaningful improvement in myositis CSM

Adult and paediatric myositis experts required a median of 20% relative percentage change in each IMACS CSM to classify adult DM/PM and JDM patients as having minimal improvement, respectively (Table 1), except for muscle enzymes, which required greater improvement. The absolute percentage change values were similar, but sometimes of lower magnitude, to classify patients as improved. Greater degrees of change were required for moderate and major improvement (Table 1).

For adult DM/PM and JDM, muscle strength was ranked as the most important CSM, followed by Physician Global Activity, and must improve for a patient to be considered clinically improved. Experts suggested that two to four measures must improve for a patient to be considered improved. A small degree of deterioration was allowable in up to two measures.

Inter-rater reliability

There was substantial agreement among all raters ($\kappa = 0.64$). There was high agreement between groups of raters (e.g. adult vs paediatric, rheumatologists vs other specialists, North America vs other geographical region, more vs less experienced) for minimal improvement ($\kappa = 0.74$ – 1.0 , 90–100% agreement) and major improvement ($\kappa = 0.83$ – 0.93 , 93–97% agreement). The intra-class correlation coefficient was high for all raters (0.82) and among different groups of raters (0.80–0.83). Among paediatric raters, agreement for both minimal and major improvement between IMACS and PRINTO measures was excellent (81–90%, $\kappa = 0.67$ for both).

Patient profiles

Baseline and follow-up CSM values, as well as relative and absolute percentage changes in the consensus profiles rated as minimally, moderately and majorly improved, are presented in supplementary Tables S5 and S6, available at *Rheumatology* Online. Consensus was achieved for 157 profiles (48%) as minimally improved, 72 profiles (22%) as moderately improved and 12 (4%) as majorly improved for adult DM/PM. Likewise, 231 (86%), 155 (58%) and 63 (24%) profiles were rated as minimally, moderately and majorly improved, respectively, for both IMACS and PRINTO JDM cases.

Conjoint-analysis survey

Table 2 shows the relative weights and rank order of importance for each CSM derived from the conjoint-analysis survey. In the survey of adult myositis experts, the weights were as follows: baseline weight 1.0 for Patient Global Activity; MMT, 2.0; Physician Global Activity, HAQ and Extramuscular Activity, 1.5; and muscle enzyme, 0.5. For the paediatric IMACS survey, baseline weight 1.0 for Parent Global Activity; MMT, 3.0; Physician Global

Activity, 2.0; and CHAQ and Extramuscular Activity, 1.5; and muscle enzyme, 0.5. For the paediatric PRINTO survey, baseline weight 1.0 for Parent Global Activity; Childhood Myositis Assessment Scale, 3.5; Physician Global Activity and DAS, 2.0; and CHAQ and Physical Summary Score of the Child Health Questionnaire, 1.0. Based on myositis experts' preferences for various CSM, all CSM were ranked in each conjoint-analysis survey (Table 2), with the rank order paralleling the relative weights in the CSM.

Candidate definitions for response criteria and their performance in patient profiles and trial analysis

A total of 287 adult, 284 paediatric IMACS and 312 paediatric PRINTO candidate definitions were derived and tested for sensitivity, specificity and AUC, using the consensus patient profile ratings as the gold standard (supplementary Table S4, available at *Rheumatology* Online): 102 adult, 101 paediatric IMACS and 99 PRINTO candidate definitions with sensitivity and specificity $\geq 80\%$, AUC ≥ 0.9 for minimal and AUC ≥ 0.8 for moderate and major improvement in the patient profile analysis were evaluated in trial analysis. For adult DM/PM, 36 candidate definitions with significant differentiating potential in the RIM trial and with AUC ≥ 0.7 were advanced to final selection. For JDM, 29 IMACS and 30 PRINTO candidate definitions with significant differentiating potential in the RIM ($P < 0.05$) and PRINTO trials ($P \leq 0.057$) as well as AUC ≥ 0.8 for minimal and AUC ≥ 0.7 for moderate/major improvement advanced to final selection. Up to four top-performing candidate definitions from each category were selected for presentation at the consensus conference. Thus, 18 adult and 14 paediatric candidate definitions for the response criteria were evaluated at the consensus conference using nominal group technique. Results of the consensus conference and final adult and paediatric myositis response criteria are described in the final response criteria manuscripts [30, 31].

Discussion

Through the efforts of many experts, by combining data sets from observational studies and clinical trials and by using several different statistical approaches, we developed new candidate definitions for response criteria for minimal, moderate and major improvement in adult DM/PM and JDM. Experts' consensus ratings of a large number of patient profiles were used as a gold standard to evaluate the candidate definitions. There was high inter-rater reliability among the experts.

Clinical trial data from two studies were used to validate the top-performing candidate definitions. As there was no difference between the treatment arms in the primary end point in the RIM trial, we used the treating physician's rating of improvement to evaluate the candidate definitions' performance as the criterion for trial validation. For the PRINTO trial of new-onset JDM patients [19], we examined the discriminative power of each candidate definition by examining the magnitude of difference

TABLE 1 Results of the core set measure survey^a

Core set measure	Minimal improvement		Moderate improvement		Major improvement	
	Relative percentage change ^b	Absolute percentage change ^c	Relative percentage change ^b	Absolute percentage change ^c	Relative percentage change ^b	Absolute percentage change ^c
Adult						
MD global	20 (15–30)	20 (10–20)	30 (25–40)	35 (20–40)	50 (40–70)	50 (30–70)
Patient/parent global activity	20 (15–30)	20 (15–30)	30 (25–40)	35 (25–40)	50 (40–60)	50 (35–60)
MMT	20 (10–20)	13 (10–20)	30 (20–40)	31 (10–44)	50 (30–70)	50 (23–63)
HAQ	20 (15–25)	33 (10–33)	30 (30–40)	33 (20–50)	50 (40–60)	67 (33–67)
ExtraMusc	20 (15–25)	20 (10–20)	30 (25–40)	40 (20–40)	50 (30–60)	50 (30–60)
Enzyme ^d	30 (20–40)	20 (10–20)	40 (30–50)	30 (20–50)	55 (50–80)	50 (30–100)
Paediatric						
MD Global	20 (20–30)	20 (20–20)	40 (30–50)	40 (24–40)	60 (50–70)	50 (40–70)
Patient/parent global activity	20 (20–30)	20 (20–30)	40 (30–50)	30 (30–50)	60 (50–70)	60 (40–70)
MMT	20 (15–30)	19 (12–25)	40 (25–50)	36 (23–45)	50 (50–70)	50 (33–66)
CHAQ	20 (20–30)	17 (13–33)	40 (30–50)	33 (20–33)	60 (50–70)	67 (40–67)
ExtraMusc	20 (20–30)	20 (20–30)	30 (30–50)	30 (30–50)	50 (50–70)	55 (50–70)
Enzyme ^d	30 (23–50)	25 (15–30)	50 (30–60)	50 (30–50)	70 (50–80)	60 (50–75)
CMAS	20 (15–30)	13 (10–19)	40 (25–50)	20 (19–38)	50 (50–70)	49 (29–58)
DAS	20 (20–30)	20 (20–25)	30 (30–50)	40 (25–50)	50 (50–70)	60 (40–75)
CHQ-PF50	20 (20–30)	15 (13–25)	40 (30–50)	25 (25–38)	55 (50–70)	40 (38–53)

^aResults are presented as the median (interquartile range). They show the degree of change deemed clinically relevant by an expert panel of myositis clinicians for minimal, moderate and major improvement for adult (DM/PM) and paediatric (JDM) groups.

^bRelative percentage change = [(final – baseline value)/baseline value] × 100. ^cAbsolute percentage change = [(final – baseline value)/potential range for test] × 100. ^dEnzyme (expressed as percentage of upper limits of normal) details for patients include

aldolase, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase and creatine kinase. CHAQ: Childhood Health Assessment Questionnaire; CHQ-PF50: physical summary score of the Child Health Questionnaire–Parent Form 50; CMAS: Childhood Myositis Assessment Scale; ExtraMusc: extramuscular global activity score; MD Global: physician global activity score; MMT: Manual Muscle Test.

between prednisone-alone vs the combined-treatment arm (prednisone plus either MTX or CSA).

We used a rigorous, systematic approach to determine candidate definitions for consideration in a final consensus meeting to select a new response criterion for adult DM/PM and JDM, based on data-driven development of candidate definitions and validation with real patient and trial data in the selection process. One strength of this approach was that we allowed many different methods to compete in the data-driven process and in final expert consensus. We used the results from the first expert survey and previously published criteria to draft traditional categorical candidate response criteria that combined different CSM for minimal, moderate and major improvement. According to guidance from representatives of the US Food and Drug Administration, we also tested candidate response criteria that required patient-reported outcome measures to improve or be weighted more heavily; however, those definitions did not perform as well as the general definitions. The conjoint-analysis surveys had been used to weight clinical and laboratory features differentially to develop scoring systems, as has been done for new classification criteria for RA, SSc and gout [12, 32–34]. We administered conjoint-analysis surveys in which experts chose the more improved hypothetical patient in a pair, defined by the

degree of change in two CSM at a time and involving a trade-off between them. These expressed choices served efficiently to reveal each CSM's relative importance and were used to differentially weight CSM in the candidate response criteria. This process also generated continuous definitions where the continuous improvement score corresponded to the magnitude of improvement, with categorical thresholds for achieving minimal, moderate and major improvement [29]. Therefore, they are truly hybrid definitions, in which the same definitions can be used as a continuous outcome measure by calculating the degree of improvement, as well as a categorical outcome measure by using thresholds of improvement. Similar hybrid candidate definitions were also developed through traditional logistic regression approaches, as well as by applying a relative weight to the degree of change in each CSM. The improvement score as a continuous outcome measure could also be used to differentiate between treatment and placebo by comparing mean improvement scores between treatment arms, and such scores could potentially provide more discriminative power, resulting in smaller sample size requirements in future trials [35]. Lastly, we selected the top-performing candidate definitions from each category for presentation at the consensus conference, as we did not know which response criterion structure would have better face validity and acceptability to

TABLE 2 Rounded weights^a and relative rank of importance of each core set measure

Core set measure	Adult IMACS		Paediatric IMACS		Paediatric PRINTO	
	Weight	Relative rank	Weight	Relative rank	Weight	Relative rank
MD global activity	1.5	2	2.0	2	2.0	2
Patient/parent global activity	1.0	5	1.0	6	1.0	4
MMT8	2.0	1	3.0	1	N/A	N/A
CMAS	N/A	N/A	N/A	N/A	3.5	1
HAQ/CHAQ	1.5	4	1.5	4	1.0	3
Extramuscular activity	1.5	3	1.5	3	N/A	N/A
DAS	N/A	N/A	N/A	N/A	2.0	2
Enzyme	0.5	6	1.0	5	N/A	N/A
CHQ-PhS	N/A	N/A	N/A	N/A	1.0	5

^aWeights were generated based on the results of the 1000Minds survey. Weights were rounded to the nearest multiple of 0.5 from their original weightings and were derived as the average weights of all adult or paediatric experts participating in the conjoint analysis surveys, with higher weight indicating experts' stronger importance associated with a particular core set measure in defining the degree of improvement in patients. Relative rank is based on mean relative importance scores, which consist of the average rank order of core set measures among experts participating in the conjoint analysis surveys, with lower scores indicating a higher degree of importance of that measure in determining whether a patient is improved or not. CHAQ: Childhood Health Assessment Questionnaire; CHQ-PhS: physical summary score of the Child Health Questionnaire; CMAS: Childhood Myositis Assessment Scale; IMACS: International Myositis Assessment and Clinical Studies Group; MMT8: Manual Muscle Test on 8 muscles; N/A: not applicable; PRINTO: Paediatric Rheumatology International Trials Organisation.

expert physicians. The results of the consensus conference are published elsewhere [30, 31].

We encountered several issues while developing these new response criteria. First, the CSM have varying relative importance, with muscle strength followed by Physician Global Activity as the most important measures. However, muscle strength changed much less than other CSM, particularly in adult patients, but was nevertheless considered by the experts to be important. Consequently, trials that rely on improvement in muscle strength as a sole primary end point [36, 37] would probably not succeed in demonstrating improvement, because strength changes very little in most adult DM/PM patients. Second, in developing new response criteria, we separated the IMACS and PRINTO CSM for JDM and tested these candidate definitions separately. However, because several CSM overlap, we paired these candidate definitions and found that they showed similar performance characteristics, thus allowing the interchange of IMACS or PRINTO CSM in a final response criterion. While developing preliminary response criteria, we sought candidate definitions that performed well for both adult DM/PM and JDM. However, in this new process, we separated adult DM/PM and JDM candidate definitions for response criteria and conducted the analyses in parallel, maintaining separate data sets and expert groups. In this way, the best-performing response criteria could be established for adult DM/PM as well as JDM, and either distinct or the same criteria could be established as the response criterion for combined adult DM/PM and JDM trials.

Potential limitations of this work include over-fitting the models and the increased likelihood of overestimating the performance owing to the large number of candidate

definitions considered. We decreased the likelihood of selecting a low-performing model by means of both internal and external validation. Another limitation is that the decisions made at various stages of development of the criteria were influenced by several considerations, including available resources and data sets, statistical properties of different approaches used, methods used for reaching consensus and experts' preferences regarding similarly performing definitions. Moreover, most of the CSM are subjective, and although there is need to develop better objective CSM in myositis, they represented the most validated and accepted CSM currently available to develop data- and consensus-driven composite response criteria. Lastly, improvement thresholds of minimal, moderate and major response were based on expert consensus as a gold standard and not on quality of life or other outcomes, such as survival.

In summary, through several online consensus-rating exercises and statistical approaches using large, pooled natural history and trial data sets, we developed new candidate definitions for response criteria for minimal, moderate and major improvement for adult DM/PM and JDM. These candidate definitions all perform acceptably, and their structure is similar—but not identical—to myositis expert opinion recommendations on the requirements for clinical improvement. A consensus selection of the top response criteria for myositis based on performance characteristics, face validity and acceptability was made at the consensus conference [30, 31]. Uniform and consensus-driven, validated response criteria in myositis, along with new and improved classification criteria that are being developed [38], may lead to better clinical trials in myositis.

Acknowledgements

We thank Saad Feroz for assistance in drafting candidate definitions for response criteria and developing patient profiles. We thank the following people for invaluable input and feedback on project development and support: B.M.F., American College of Rheumatology; Daniel Aletaha, European League Against Rheumatism; Suzette Peng and Sarah Yim, the US Food and Drug Administration; 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 thank Drs Michael Ward and Steven Pavletic for critical reading of the manuscript. P.H. owns the 1000Minds software referred to in this article, which he co-invented with Franz Ombler. J.V.'s work in myositis is supported by the project (Ministry of Health, Czech Republic) for conceptual development of research organization 00023728 (Institute of Rheumatology). H.C. is supported by an Arthritis Research UK (18474) and the National Institute for Health Research Biomedical Research Unit Funding Scheme.

International Myositis Assessment and Clinical Studies Group, Paediatric Rheumatology International Trials Organization contributor list: Steering Committee: L.G.R. (Co-Principal Investigator), N.R. (Co-Principal Investigator), R.A. (Methodology Lead), F.W.M., J.V. Statistical Team: R.A., B.E., N.B., A.P., A.M.H., P.H., B.M.F., H.R., P.A.L., N.R., L.G.R. Adult Core Set Survey Group: A.A.A., H.C., Lisa Christopher-Stine, L.C., R.G.C., Lisa Criscione-Schreiber, Leslie Crofford, Mary E. Cronin, D.F., I.Gdl.T., K.D., Patrick Gordon, Gerald Hengstman, James D. Katz, Andrew Mammen, Galina Marder, Neil McHugh, C.V.O., Elena Schiopu, Albert Selva-O'Callaghan, Y.W.S., J.V., Gil Wolfe, Robert Wortmann. Paediatric Core Set Survey Group: Maria Apaz, Suzanne Bowyer, R.C., Tamás Constantin, Megan Curran, Joyce Davidson, B.M.F., Thomas Griffin, A.M.H., Olcay Jones, Susan Kim, Bianca Lang, Carol Lindsley, Daniel Lovell, Claudia Saad Magalhaes, Lauren M. Pachman, C.A.P., Andrea Ponyi, Marilyn Punaro, Pierre Quartier, Athimalaipet V. Ramanan, Angelo Ravelli, A.M.R., 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: A.A.A., H.C., L.C., R.G.C., K.D., D.F., I.Gdl.T., M.G., I.E.L., F.W.M., C.V.O., P.P., A.M.R., L.G.R., N.R. and members of PRINTO, Y.W.S., J.V. Adult Patient Profile Working Group: R.A., A.A.A., Dana Ascherman, Richard Barohn, Olivier Benveniste, Jan De Bleecker, Jeffrey Callen, Christina Charles-Schoeman, H.C., Lisa Christopher-Stine, L.C., R.G.C., Leslie Crofford, Mary E. Cronin, K.D., Sonye Danoff, Maryam Dastmalchi, I.Gdl.T., Mazen Dimachkie, Steve DiMartino, Lyubomir Dourmishev, Floranne Ernste, D.F., Takahisa Gono, Patrick Gordon, M.G., David Isenberg, Yasuhiro Katsumata, James D. Katz, John Kissel, Richard L. Leff, Todd Levine, I.E.L., Andrew Mammen, Herman Mann, Galina Marder, Isabelle

Marie, Neil McHugh, Joseph Merola, F.W.M., C.V.O., Marzena Olesinska, Nancy Olsen, Nicolo Pipitone, Sindhu Ramchandren, Seward Rutkove, Lesley Ann Saketkoo, Adam Schifffenbauer, Albert Selva-O'Callaghan, Samuel Katsuyuki Shinjo, Rachel Shupak, Y.W.S., Katarzyna Swierkocka, J.V., Marianne de Visser, Julia Wanschitz, Victoria P. Werth, Irene Whitt, Robert Wortmann, Steven Ytterberg. Paediatric Patient Profile Working Group: Maria Apaz, Tadej Avcin, Mara Becker, Michael W. Beresford, R.C., Tamás Constantin, Megan Curran, R.J.C., Joyce Davidson, Frank Dressler, Jeffrey Dvergsten, B.M.F., Virginia Paes Leme Ferriani, Berit Flato, Valeria Gerloni, Thomas Griffin, Michael Henrickson, Claas Hinze, Mark Hoeltzel, A.M.H., Maria Ibarra, Norman Ilowite, Lisa Imundo, Olcay Jones, Susan Kim, Daniel Kingsbury, Bianca Lang, Carol Lindsley, Daniel Lovell, A.M., Claudia Saad Magalhaes, Bo Magnusson, Sheilagh Maguiness, S.M., Pernille Mathiesen, Liza McCann, Susan Nielsen, Sheila Knupp Feitosa de Oliveira, Lauren M. Pachman, Murray Passo, C.A.P., Marilyn Punaro, Pierre Quartier, Eglá Rabinovich, Athimalaipet V. Ramanan, Angelo Ravelli, A.M.R., Robert Rennebohm, L.G.R., 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, J.vd.N., Dawn Wahezi, Carol Wallace, Francesco Zulian. Conjoint Analysis—Adult Group: R.A., A.A.A., H.C., Lisa Christopher-Stine, L.C., R.G.C., Mary E. Cronin, K.D., Mazen Dimachkie, Steve Di Martino, D.F., I.Gdl.T., Patrick Gordon, I.E.L., Herman Mann, F.W.M., C.V.O., Albert Selva-O'Callaghan, J.V., Victoria P. Werth, Robert Wortmann, Steven Ytterberg. Conjoint Analysis—Paediatric Group: R.C., Tamás Constantin, R.J.C., Joyce Davidson, Frank Dressler, B.M.F., Thomas Griffin, Michael Henrickson, A.M.H., Lisa Imundo, Bianca Lang, Carol Lindsley, Claudia Saad Magalhaes, Bo Magnusson, S.M., Sheila Knupp Feitosa de Oliveira, Lauren M. Pachman, Murray Passo, C.A.P., Marilyn Punaro, Angelo Ravelli, A.M.R., L.G.R., Kelly Rouster-Stevens, Annet Van Royen-Kerkhof, Ricardo Russo, Bracha Shaham, Robert Sundel, J.vd.N.

Funding: This work was supported in part by the ACR, the EULAR, the intramural research programme of the National Institutes of Health, 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.

Disclosure statement: R.A. has received consultancies from BMS, Novartis, Octapharma and Mallinckrodt and research grants from Pfizer, BMS and Mallinckrodt. A.A. has received consultancies and/or is on the medical

advisory board for Idera, Novartis, Akashi and Acceleron. H.C. has received research support from Novartis. I.L. has received grants from BMS and Astra-Zeneca, and has served on the advisory board for BMS. N.R. has received speaker's bureaus and consulting fees from AbbVie, Amgen, Biogenidec, Alter, AstraZeneca, Baxalta Biosimilars, Biogenidec, Boehringer, BMS, Celgene, CrescendoBio, EMD Serono, Hoffman-La Roche, Italfarmaco, Janssen, MedImmune, Medac, Novartis, Novo Nordisk, Pfizer, Sanofi Aventis, Servier, Takeda, UCB Biosciences GmbH and their institute, the G. Gaslini Hospital, has received contributions from industries for the coordination activity of the PRINTO network from BMS, GlaxoSmithKline, Hoffman-La Roche, Novartis, Pfizer, Sanofi Aventis, Schwarz Biosciences, Abbott, Francesco Angelini S.P.A., Sobi, Merck Serono; this money has been reinvested for the research activities of the hospital in fully independent manners besides any commitment with third parties. L.R. has received research grants from American College of Rheumatology, National Institute of Environmental Health Sciences, NCATS, NIAMS of NIH, Cure JM and The Myositis Association. P.H. owns the 1000Minds software referred to in this article, which he co-invented with Franz Omblér. N.B. received grants/research support from the Cure JM Foundation. H.E.R. has received grant/research support from the University of Pittsburgh. All other authors have declared no conflicts of interest.

Supplementary data

Supplementary data are available at *Rheumatology* Online.

References

- Rider LG, Werth VP, Huber AM *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* 2011;63 (Suppl 11):S118-57.
- Ruperto N, Martini A. Networking in paediatrics: the example of the Paediatric Rheumatology International Trials Organisation (PRINTO). *Arch Dis Child* 2011;96:596-601.
- Miller FW, Rider LG, Chung YL *et al.* Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. *Rheumatology* 2001;40:1262-73.
- Ruperto N, Ravelli A, Pistorio A *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.
- Ruperto N, Ravelli A, Murray KJ *et al.* Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. *Rheumatology* 2003;42:1452-9.
- Rider LG, Giannini EH, Harris-Love M *et al.* Defining clinical improvement in adult and juvenile myositis. *J Rheumatol* 2003;30:603-17.
- Rider LG, Giannini EH, Brunner HI *et al.* International consensus on preliminary definitions of improvement in adult and juvenile myositis. *Arthritis Rheum* 2004;50:2281-90.
- Ruperto N, Pistorio A, Ravelli A *et al.* The Pediatric Rheumatology International Trials Organisation provisional criteria for the evaluation of response to therapy in juvenile dermatomyositis. *Arthritis Care Res* 2010;62:1533-41.
- Lazarevic D, Pistorio A, Palmisani E *et al.* The PRINTO criteria for clinically inactive disease in juvenile dermatomyositis. *Ann Rheum Dis* 2013;72:686-93.
- Hasija R, Pistorio A, Ravelli A *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;63:3142-52.
- Hansen P, Omblér 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.
- Neogi T, Jansen TL, Dalbeth N *et al.* 2015 Gout Classification Criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheumatol* 2015;67:2557-68.
- Utz KS, Hoog J, Wentrup A *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.
- Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, eds. *Using discrete choice experiments to value health and health care*. Dordrecht: Springer, 2008.
- 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.
- Singh JA, Solomon DH, Dougados M *et al.* Development of classification and response criteria for rheumatic diseases. *Arthritis Rheum* 2006;55:348-52.
- Ruperto N, Meiorin S, Iusan SM *et al.* Consensus procedures and their role in pediatric rheumatology. *Curr Rheumatol Rep* 2008;10:142-6.
- 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.
- Ruperto N, Pistorio A, Oliveira S *et al.* Prednisone versus prednisone plus ciclosporin versus prednisone plus

- methotrexate in new-onset juvenile dermatomyositis: a randomised trial. *Lancet* 2016;387:671–8.
- 20 Rider LG, Lee J, Jansen A *et al.* Defining clinically relevant changes in core set activity measures for adult and juvenile idiopathic inflammatory myopathies (IIM). *Arthritis Rheum* 2011;63 (Suppl):S89.
- 21 Lundberg IE, Svensson J. Registries in idiopathic inflammatory myopathies. *Curr Opin Rheumatol* 2013;25:729–34.
- 22 Zong M, Dorph C, Dastmalchi M *et al.* Anakinra treatment in patients with refractory inflammatory myopathies and possible predictive response biomarkers: a mechanistic study with 12 months follow-up. *Ann Rheum Dis* 2014;73:913–20.
- 23 Oddis CV, Reed AM, Aggarwal R *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.
- 24 Chung L, Genovese MC, Fiorentino DF. A pilot trial of rituximab in the treatment of patients with dermatomyositis. *Arch Dermatol* 2007;143:763–7.
- 25 Dastmalchi M, Grundtman C, Alexanderson H *et al.* A high incidence of disease flares in an open pilot study of infliximab in patients with refractory inflammatory myopathies. *Ann Rheum Dis* 2008;67:1670–7.
- 26 Bohan A, Peter JB. Polymyositis and dermatomyositis (first of two parts). *New Engl J Med* 1975;292:344–7.
- 27 Bohan A, Peter JB. Polymyositis and dermatomyositis (second of two parts). *New Engl J Med* 1975;292:403–7.
- 28 Youden W. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- 29 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.
- 30 Rider LG, Aggarwal R, Pistorio A *et al.* International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism Criteria for Minimal, Moderate, and Major Clinical Response in Juvenile Dermatomyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. *Ann Rheum Dis* 2017;76:782–91.
- 31 Aggarwal R, Rider LG, Ruperto N *et al.* International Myositis Assessment and Clinical Studies Group and the Paediatric Rheumatology International Trials Organisation. 2016 American College of Rheumatology/European League Against Rheumatism criteria for minimal, moderate, and major clinical response in adult dermatomyositis and polymyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. *Ann Rheum Dis* 2017;76:792–801.
- 32 Neogi T, Aletaha D, Silman AJ *et al.* The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. *Arthritis Rheum* 2010;62:2582–91.
- 33 Johnson SR, Naden RP, Fransen J *et al.* Multicriteria decision analysis methods with 1000Minds for developing systemic sclerosis classification criteria. *J Clin Epidemiol* 2014;67:706–14.
- 34 Van Den Hoogen F, Khanna D, Fransen J *et al.* 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League against Rheumatism collaborative initiative. *Arthritis Rheum* 2013;65:2737–47.
- 35 Streiner DL. Breaking up is hard to do: the heartbreak of dichotomizing continuous data. *Can J Psychiatry* 2002;47:262–6.
- 36 Villalba L, Hicks JE, Adams EM *et al.* Treatment of refractory myositis: a randomized crossover study of two new cytotoxic regimens. *Arthritis Rheum* 1998;41:392–9.
- 37 Cherin P, Herson S, Wechsler B *et al.* Efficacy of intravenous gammaglobulin therapy in chronic refractory polymyositis and dermatomyositis: an open study with 20 adult patients. *Am J Med* 1991;91:162–8.
- 38 Lundberg IE, Miller FW, Tjärnlund A, Bottai M. Diagnosis and classification of idiopathic inflammatory myopathies. *J Intern Med* 2016;280:39–51.