

Published in final edited form as:

Arthritis Care Res (Hoboken). 2010 November ; 62(11): 1533–1541. doi:10.1002/acr.20280.

THE PEDIATRIC RHEUMATOLOGY INTERNATIONAL TRIALS ORGANIZATION PROVISIONAL CRITERIA FOR THE EVALUATION OF RESPONSE TO THERAPY IN JUVENILE DERMATOMYOSITIS

Nicolino Ruperto¹, Angela Pistorio², Angelo Ravelli^{1,3}, Lisa G. Rider⁴, Clarissa Pilkington⁵, Sheila Oliveira⁶, Nico Wulffraat⁷, Graciela Espada⁸, Stella Garay⁹, Ruben Cuttica¹⁰, Michael Hofer¹¹, Pierre Quartier¹², Jose Melo-Gomes¹³, Ann M. Reed¹⁴, Malgorzata Wierzbowska¹⁵, Brian M. Feldman¹⁶, Miroslav Harjacek¹⁷, Hans-Iko Huppertz¹⁸, Susan Nielsen¹⁹, Berit Flato²⁰, Pekka Lahdenne²¹, Harmut Michels²², Kevin J. Murray²³, Lynn Punaro²⁴, Robert Rennebohm²⁵, Ricardo Russo²⁶, Zsolt Balogh²⁷, Madeleine Rooney²⁸, Lauren M. Pachman²⁹, Carol Wallace³⁰, Philip Hashkes³¹, Daniel J. Lovell³², Edward H. Giannini³², and Alberto Martini³ for the Pædiatric Rheumatology International Trials Organisation (PRINTO) and the Pediatric Rheumatology Collaborative Study Group (PRCSG)

¹Nicolino Ruperto, MD, MPH, IRCCS G Gaslini, Pediatria II, Reumatologia, PRINTO, Genova, Italy ²Angela Pistorio, MD, PhD, IRCCS G Gaslini, Servizio di Epidemiologia e Biostatistica, Genova, Italy ³Angelo Ravelli, MD, Prof, Alberto Martini, MD, Prof, IRCCS G. Gaslini, Pediatria II, Reumatologia and Dipartimento di Pediatria, Università degli Studi, Genova, Italy ⁴Lisa G. Rider, MD, NIEHS, NIH, Bethesda, MD, USA ⁵Clarissa Pilkington, MD, Great Ormond Street Hospital for Sick Children, London, United Kingdom ⁶Sheila Oliveira, MD, Prof, Universidade Federal do Rio de Janeiro, Instituto de Puericultura e Pediatria Martagao Gesteira, Rio de Janeiro, Brazil ⁷Nico Wulffraat, MD, Wilhelmina Kinderziekenhuis, Utrecht, the Netherlands ⁸Graciela Espada, MD, Hospital de Ninos Ricardo Gutierrez, Buenos Aires, Argentina ⁹Stella Garay, MD, Hospital Sor Maria Ludovica, La Plata, Argentina ¹⁰Ruben Cuttica, MD, Hospital General de Ninos Pedro de Elizalde, Buenos Aires, Argentina ¹¹Michael Hofer, MD, Centre Multisite Romand de Rhumatologie Pédiatrique, Lausanne, Switzerland ¹²Pierre Quartier, MD, Université Paris-Descartes, Hôpital Necker-Enfants Malades, Paris, France ¹³Jose Melo-Gomes, MD, Instituto

Please address correspondence and reprint requests to: Nicolino Ruperto, M.D., M.P.H., Pædiatric Rheumatology International Trials Organisation (PRINTO), IRCCS G. Gaslini, Università di Genova, Pediatria II - Reumatologia, EULAR Centre of Excellence in Rheumatology 2008-2013, Largo Gaslini, 5, 16147 Genova ITALY, Tel: +39-010-38-28-54 or +39-010-39-34-25, Fax: +39-010-39-33-24 or +39-010-39-36-19, nicolaruperto@ospedale-gaslini.ge.it <http://www.printo.it> or www.pediatric-rheumatology.printo.it.

Organizers:

Italy: Alberto Martini, MD, Prof, Nicolino Ruperto, MD, MPH, Angelo Ravelli, MD, Angela Pistorio, MD, PhD; **USA:** Edward H Giannini, MSc, DrPH, Daniel J Lovell, MD, MPH; **Sweden:** Boel Andersson-Gäre, MD, PhD.

Attendees:

Argentina: Carmen De Cunto, MD, Ruben Cuttica, MD; **Belgium:** Rik Joos, MD; **Brazil:** Claudia Magalhaes Saad, MD, Sheila Oliveira, MD; **Bulgaria:** Dimitrina Mihaylova, MD; **Canada:** Brian M. Feldman, MD, MSc; **Croatia:** Miroslav Harjacek, MD; **Czech Republic:** Pavla Dolezalova, MD; **Denmark:** Susan Nielsen, MD; **Finland:** Pekka Lahdenne, MD; **France:** Anne Marie Prieur, MD; **Germany:** Hans Iko Huppertz, MD; **Greece:** Florence Kanakoudi Tsakalidou, MD; **Israel:** Philip Hashkes, Yosef Uziel, MD; **Latvia:** Ingrida Rumba, MD; **Mexico:** Ruben Burgos Vargas, MD; **Netherlands:** Nico Wulffraat, MD; **Norway:** Berit Flato, MD; **Poland:** Malgorzata Wierzbowska, MD; **Portugal:** Jose Antonio Melo-Gomes, MD; **Serbia and Montenegro:** Gordana Susic, MD; **Slovakia:** Richard Vesely, MD; **Slovenia:** Tadej Avcin, MD; **Switzerland:** Michael Hofer, MD; **Turkey:** Huri Ozdogan, MD; **United Kingdom:** Clarissa Pilkington, MD, Madeleine Rooney, MD; **USA:** Daniel J. Lovell, MD, MPH, Lauren M. Pachman, MD, Lisa G. Rider, MD, Ann M. Reed, MD, Robert Rennebohm, MD, Carol Wallace, MD.

External Observers:

Brazil: Marcia Bandeira, MD; **Greece:** Jenny Pratsidou, MD; **Argentina:** Stella Maris Garay, MD.

Portugues de Reumatologia, Lisbon, Portugal ¹⁴Ann M. Reed, MD, Mayo Clinic School of Medicine and Mayo Foundation, Rochester, MN, USA ¹⁵Malgorzata Wierzbowska, MD, Institute of Rheumatology, Warsaw, Poland ¹⁶Brian M. Feldman, MD, MSc, FRCPC, University of Toronto, The Hospital for Sick Children, Toronto, Canada ¹⁷Miroslav Harjacek, MD, Children's Hospital Zagreb, Zagreb, Croatia ¹⁸Hans-Iko Huppertz, MD, Klinikum Bremen-Mitte, Bremen, Germany ¹⁹Susan Nielsen, MD, Juliane Marie Centret, Rigshospitalet, Copenhagen, Denmark ²⁰Berit Flato, MD, Rikshospitalet University Hospital, Oslo, Norway ²¹Pekka Lahdenne, MD, Hospital for Children and Adolescents, Helsinki, Finland ²²Harmut Michels, MD, Rheumakinderklinik Garmisch-Partenkirchen, Germany ²³Kevin J. Murray, MD, Princess Margaret Hospital for Children, Perth, WA, Australia ²⁴Lynn Punaro, MD, Texas Scottish Rite Hospital, Dallas, TX, USA ²⁵Robert Rennebohm, MD, Alberta Children's Hospital, Calgary, Canada ²⁶Ricardo Russo, MD, Hospital de Pediatria Juan P. Garrahan, Buenos Aires, Argentina ²⁷Zsolt Balogh, MD, National Institute of Rheumatology and Physiotherapy, Budapest, Argentina ²⁸Madeleine Rooney, MD, Musgrave Park Hospital, Belfast, Northern Ireland ²⁹Lauren M. Pachman, MD, Children's Memorial Hospital, Chicago, IL, USA ³⁰Carol Wallace, MD, University of Washington/Children's Hospital, Seattle, WA, USA ³¹Philip Hashkes, MD, Shaare Zedek Medical Center, Jerusalem, Israel ³²Daniel J. Lovell, MD, MPH, Edward H. Giannini, MSc, DrPH, Children's Hospital Medical Center, Cincinnati, OH, USA

Abstract

Objective—To develop a provisional definition for the evaluation of response to therapy in juvenile dermatomyositis (JDM) based on the PRINTO JDM core set of variables.

Methods—Thirty-seven experienced pediatric rheumatologists from 27 countries, achieved consensus on 128 difficult patient profiles as clinically improved or not improved using a stepwise approach (patients rating, statistical analysis, definition selection). Using the physicians' consensus ratings as the “gold-standard measure”, chi-square, sensitivity, specificity, false positive and negative rate, area under the ROC, and kappa agreement for candidate definitions of improvement were calculated. Definitions with kappa >0.8 were multiplied with the face validity score to select the top definitions.

Results—The top definition of improvement was: at least 20% improvement from baseline in 3/6 core set variables with no more than 1 of the remaining worsening by more than 30%, which cannot be muscle strength. The second highest scoring definition was at least 20% improvement from baseline in 3/6 core set variables with no more than 2 of the remaining worsening by more than 25%, which cannot be muscle strength which is definition P1 selected by the IMACS group. The third is similar to the second with the maximum amount of worsening set to 30%. This indicates convergent validity of the process.

Conclusion—we proposes a provisional data driven definition of improvement that reflects well the consensus rating of experienced clinicians, which incorporates clinically meaningful change in core set variables in a composite endpoint for the evaluation of global response to therapy in JDM.

Keywords

juvenile dermatomyositis; core set; response to therapy; disease activity; consensus

The standardization of the criteria to evaluate improvement in rheumatic diseases has been a goal of numerous research groups. This work led to establishment of definition of response in rheumatoid arthritis (1), juvenile arthritis (2–4), systemic lupus erythematosus (SLE) both in adults (5–7) and children (8–10).

The International Myositis Outcome Assessment and Clinical Studies (IMACS) group proposed a core set of outcome variables for inclusion in clinical trials in adult and juvenile inflammatory myopathies and defined the degree of change in each core set variables that is clinically meaningful, as well as guidelines for performing clinical trials (11–14). However, until now these proposals have not yet been formally validated in the context of external prospective pediatric studies or clinical trials. Although children/adolescents and adults with DM share many signs and symptoms of disease, they differ in the clinical features and outcome (15–17), and treatment approaches should consider the peculiarities of juvenile patients as well as their longer life expectancy. Therefore, all outcome measures developed for adults need to be subjected to a critical evidence-based evaluation of their measurement properties in children and adolescents.

To help standardize the conduct and reporting of juvenile dermatomyositis (JDM) clinical trials and enhance identification of new therapeutic agents, the Pediatric Rheumatology International Trials Organization (PRINTO) (18), in collaboration with the Pediatric Rheumatology Collaborative Study Group (PRCSG) and with the support of the European Union and the U.S. National Institutes of Health, undertook in year 2000 a multinational effort to develop, and promulgate a core set of outcome variables and a definition of clinical improvement to evaluate response to therapy in patients with JDM and in juvenile SLE. The first two phases of the project, previously published (8;19), led to the development of a prospectively evidence-based validated core set of six variables for the evaluation of response to therapy that is now known as the provisional PRINTO/American College of Rheumatology/European League Against Rheumatism Disease Activity Core Set for the evaluation of response to therapy in JDM (PRINTO/ACR/EULAR JDM core set) (Table 1).

In this paper we report the results of the third phase of the project, which was aimed at developing a provisional validated definition of improvement to aid in the classification of individual patients in future therapeutic trials and in current clinical practice as either *improved* or *not improved*.

PATIENTS AND METHODS

The overall methodology of this phase of the project was based on a methodological framework used successfully in previous work in rheumatoid arthritis (1) juvenile arthritis (2–4), juvenile SLE (8–10), and inflammatory myopathies (13).

Table 1 gives the six core variables validated previously and the respective tools for their assessment. The PRINTO JDM core set includes the following six variables: 1) physician's global assessment of the patient's overall disease activity measured with a 10-cm visual analogue scale (VAS) (0=no activity; 10=maximum activity) (20); 2) muscle strength as assessed by the Childhood Myositis Assessment Scale (CMAS) (0 = worst; 52 = best) (21–23); 3) global disease activity assessment through the Disease Activity Score (DAS) (24) or alternatively the Myositis Disease Activity Assessment (MDAA, this instrument (25) combines two partially overlapping tools named the Myositis Disease Activity Assessment Visual Analogue Scale [MYOACT] and the Myositis Intention to Treat Activity Index A–E version [MITAX) (25)); 4) parent's global assessment of the overall child's well-being on a 10-cm VAS (0 = very well; 10 = very poor) (20;26;27); 5) functional ability, as measured by the Childhood Health Assessment Questionnaire (C-HAQ) (26;27) (0 = best; 3 = worst); 6) health-related quality of life (HRQOL) assessment using the physical summary score (PhS) of the Child Health Questionnaire (CHQ) parent version (27;28). The methods for calculating the scores of the PRINTO JDM core set variables are reported in Ruperto et al (19).

The variables underwent extensive evidence-based evaluation, the process of which has been described previously (19). In particular, all variables were found to be feasible, and have good construct validity, discriminant ability, and internal consistency. Furthermore, they were not redundant, proved responsive to clinically important change in disease activity, and were strongly associated with treatment outcome and thus were included in the final core set.

Following this selection of variables for the evaluation of response to therapy, a second consensus conference was held attended by 37 experienced pediatric rheumatologists from 27 different countries to ensure wide international acceptance of the results, and was facilitated by 4 of the authors (NR, EHG, BAG, AP) with expertise in nominal group process (29;30). The overall goal of the meeting was to reach consensus on a provisional validated definition of improvement, incorporating the PRINTO core set of variables, using a combination of statistical criteria and consensus formation techniques. In order to achieve this objective, four steps (process and analysis) were pursued as briefly described in order below and whose full details can be found elsewhere (2;19).

Step 1: Rate each of 128 paper patient profiles as “*clinically importantly improved*” or “*not improved*”, using nominal group technique. Data from the 294 JDM patients analysed for the PRINTO/ACR/EULAR JDM core set (19) were used to select a subgroup of 128 difficult/atypical patient profiles presented to conference attendees for evaluation of therapeutic response. The profiles selected (see examples in Table 2) were those that were judged by the conference organizers to be near a putative threshold level of improvement. For example, patients who showed 100% improvement in all outcome variables were not good candidates for inclusion because all would agree that the patient had improved, and all the definitions of improvement would categorize the patient as *improved*. Each profile contained only information related to the six validated JDM core set variables with absolute values at baseline and at 6 months, as well as absolute and percent change from baseline, (Table 1 and Table 2). Participants were randomized into three “nominal groups” of equal size, and asked to rate independently all 128 difficult patient profiles as either *clinically importantly improved* or *not improved*. If an 80% consensus was not achieved, the case was discussed in a round-robin fashion at each table and if necessary also in a plenary session. We expected to reach consensus for at least 80% of the patients discussed.

Step 2 (statistical analysis): Using the physicians’ consensus judgment as the “gold standard”, we performed several statistical evaluations (see below) to identify the definition of improvement with the best performance characteristics. We were unable to find in the literature any definitions of improvement that used combinations of the core set variables. Therefore, we tested 999 different definitions of improvement that were deemed clinically reasonable by the the Steering Committee of the project (NR, AP, AR, DHL, EHG, AM). Some of the definitions of improvement tested were provided by the IMACS group (13).

Each definition of improvement was classified as either “generic” or “specific” (9). An example of “generic definition” is as follows: at least 20% improvement from baseline in any 2 of the 6 core set variables with no more than 1 of the remaining worsening by more than 30%. An example of a “specific definition” is as follows: physician’s global assessment of the patient’s overall disease activity and muscle strength improved by at least 30%, two of any remaining three improved by at least 20%, and none worsening by more than 30%.

We evaluated the ability of the 999 candidate definitions of improvement to classify individual patients as *improved* or *not improved*, and then assessed the agreement between the definitions and consensus of the physicians. We used only patient profiles for which physician consensus was achieved. For each definition, we calculated the chi-square test (1

df) and the corresponding p value, sensitivity, specificity, percent of false-positives, percent of false-negatives, and area under the receiver operating characteristic curve (ROC) (31). The kappa statistic (32) was used to measure the strength of concordance between the definitions and consensus of the physicians. The kappa statistic was converted to a Likert-like scale using the conversion proposed by Landis & Koch (33): 0.01–0.2 = slight; 0.21–0.4 = fair; 0.41–0.6 = moderate; 0.61–0.8 = substantial; 0.81–1 = almost perfect agreement. While the statistical properties of all 999 definition were presented to the consensus attendees only definitions with a kappa > 0.7 (substantial agreement), sensitivity and specificity > 80%, and percent false positive and false negative < 20%, were retained in the further analysis. Results of the statistical analyses were then presented to the conference attendees.

Step 3: We then used nominal group technique to decide which of the definitions of improvement with the highest statistical performance is easiest to use and most credible (highest face validity). The attendees were again randomly split into three groups and, using nominal group technique, were asked to decide which definitions of improvement (selected among the 999 definition tested) that performed best (in the analysis described above) were easiest to use and most credible (content validity), ranking the 5 best from 1 (lowest) to 5 (highest content validity).

Step 4: We multiplied the content validity score by the kappa values to obtain the “best” definitions. For each definition, the three content validity rankings obtained by the 3 nominal groups were summed up and the resulting sum was multiplied by the corresponding value of the kappa statistic, to obtain the “final score” that incorporated both statistical evaluations and experts’ judgment.

Association between changes in each of the 5 core variables and the overall outcome

The association between the change in each core set variables and the evaluation of response to therapy was analyzed by multiple logistic regression, which used as explanatory variables the baseline-to-6-month change in each core set variable and as the dependent outcome the physician’s consensus evaluation of patient’s improvement. Odds ratios (OR) with 95% confidence intervals (95% CI) were reported. Continuous variables were dichotomized according to the best cut-offs provided by the ROC analysis (31). The purposes of this post-consensus analysis was to evaluate which were the core set variables that influenced most the consensus decision and to establish the best cut-offs for absolute change for the variables included in the model. The best cut-offs for each core set variable should help physicians decide if a patient is improved based on the absolute change of that particular measure.

Data were entered into an Access XP database and analyzed with Excel XP (Microsoft), XLSTAT 6.1.9 Addinsoft, Statistica 6.0 (StatSoft, Inc), and Stata 7.0 (Stata Corporation).

RESULTS

Table 3 shows the comparison of demographic features and baseline and 6-month values of the core set variables between the subgroup of 128 difficult patients used to create the patient’ profiles used in this exercise, and the remaining 166 patient-cohort; the entire cohort of 294 patients was analysed for the PRINTO/ACR/EULAR JDM core set(19). In general, the features were comparable between cohorts, although the former had longer disease duration. Similarly, the two cohorts were comparable at baseline for five of the core set variables; the exception being the parent’s global assessment of the overall child’s well-being. The differences observed at 6 months between the 128 patient-cohort and the remaining sample was expected because this 128 subgroup was composed of the difficult/atypical patients selected for the consensus exercise that overall responded less to the 6-

month treatment given by the treating physicians (see Methods section). The remaining 166 patient-cohort consisted of patients who achieved the most pronounced levels of improvement, after the 6-month of treatment, and who were not useful for the purposes of the consensus exercise.

Results of scoring the patient profiles

Consensus $\geq 80\%$ was achieved for 121 (95%) of the 128 difficult patients, with 98/121 (81%) patients being judged as *clinically importantly improved*, and 23/121 (19%) patients as *not improved*. All three nominal groups reached the same consensus opinion as to patient status on all profiles.

Identification of the top definitions of improvement as the best performers

Thirteen of the 999 definitions of improvement reached a kappa ≥ 0.8 (almost perfect agreement); their corresponding chi-square values, p values, sensitivity, specificity, percent false positive and false negative rates, AUC, and kappa statistics are reported in Table 4.

Face validity of the top definitions of improvement and final resolution

After presentation of the above data, attendees used nominal group technique to rate content validity (Step 3) using a 1–5 scale, with five being the highest. The sums of the combined ranks from the three nominal groups are presented in Table 4 (min-max 1–131). Next, the sum of the ranking was multiplied by its respective kappa statistic to obtain the final score (min-max 1–113), thereby allowing identification of the definitions of improvement with the highest final score. The definition of improvement that scored highest was the following: At least 20% improvement from baseline in 3 of any 6 variables with no more than one of the remaining worsening by more than 30%, which cannot be muscle strength (as measured by the CMAS).

As can be seen in Table 4, the definitions that scored second (IMACS P1) and third highest are similar to the first all requiring an improvement $\geq 20\%$ in at least 3 core set variable, but required a different number (2 instead of 1) or a different degree of worsening (25% instead of 30%) in the remaining variables (13). The similarity of the top ranking definitions indicates convergent validity of the measures. Since the statistical performance of the best definitions had all kappa > 0.8 , the selection of the final definition of improvement was driven mainly by the ranking (content validity) of the top 5 definitions.

Association between changes in each of the 6 core variables and the overall outcome

The association between the change in each core set measure and response to therapy was analyzed in a multivariate analysis, as described in the Methods section. In the final model (Table 5), the physician's global assessment of the patient's overall disease activity appeared to be the strongest predictor of response to therapy (OR, 11), followed by the CMAS (OR, 10.2) and the parent's global assessment of the overall child's well-being (OR, 5.5). The remaining three core set variables, the DAS, the C-HAQ and the CHQ PhS did not reach statistical significance. In the footnote of Table 5 are also reported the best cut-offs for absolute change for the variables included in the model.

DISCUSSION

Using a combination of data-driven and consensus-formation processes, pediatric rheumatologists with specific expertise in the assessment of JDM developed a provisional validated definition of improvement that PRINTO proposes for use in future JDM clinical trials. Based on the best performing definition, improvement in individual patients with JDM can be defined as follows: any three among the six core set variables improved by at

least 20% versus baseline, with no more than one of the remaining variables worsening by more than 30%, which cannot be muscle strength.

The provisional definition selected by the consensus panel performed well in the available data set, with high sensitivity and specificity, and low false-positive and false-negative rates. The consensus process indicated that this definition had the best content validity as well. The main strength of the definition lies in the consensus of a large number of experienced pediatric rheumatologists from many countries, that provided wide international acceptance of the project, and in its strong statistical properties. Furthermore its core set variables (19) were selected with by an evidence-based process and validated through a large scale data collection in patients who had been assessed in a prospective fashion.

During the discussion phase in the content validity session participants made it clear that muscle strength is one of the essential components for the evaluation of response to therapy in JDM. For this reason, all definitions that required muscle strength to not worsen were highly ranked.

Of note, the second highest scoring definition was at least 20% improvement from baseline in 3 of any 6 core set variables with no more than 2 of the remaining worsening by more than 25% which cannot be muscle strength, is definition P1 selected by the IMACS group (13). This demonstrates convergent validity of the approaches used by the two groups which confirm the validity of the 2 parallel works and the respective findings but in different cohorts. The main difference between the PRINTO and the IMACS group definition of improvement is that we focused on response criteria for use only in JDM and not also in adult patients with DM and PM. Other differences, fully discussed elsewhere (17;19), are related to the core set of variables with serum muscle enzymes included in IMACS core set and excluded in PRINTO core set for their poor statistical performance, and second the inclusion of HRQOL assessment as a distinct core set variables specific for children by the PRINTO group, whereas the IMACS investigators did not incorporate it in the core set, though they recommended to include this measure in therapeutic trials of patients with IIM. Future studies in external cohort will allow the comparison and final validation of the 2 proposed core set and definitions.

The provisional validated definition of improvement was based on a composite combination of outcome measures that were set up to detect a broad range of clinical change. The PRINTO JDM core set includes both objective and subjective measures from both, the physician and patient/parents' perspective. The evaluation of response to therapy from different perspectives has the advantage of covering all changes induced by the agent under study and of providing information related to the entire spectrum of disease manifestations and consequences. It is also expected to provide better discriminant validity than previous clinical trials which used only muscle strength as the primary outcome (12).

For the practical application of the provisional PRINTO definition of improvement we reported in Table 1 the domains and suggested variables included in the final core set for the evaluation of response to therapy in JDM (adapted from ref. (19)). The suggested variables to measure each domain are the ones used for the validation of the core set and of the definition of improvement but researchers can use other variables that might be more appropriate based on their study design or new validation data that may appear in the literature. In addition in Table 2 are reported 2 examples with data from real patients used at the consensus conference that will help readers, by using the related formulas, to apply the PRINTO definition of improvement for JDM. In the footnote of Table 5 are also reported the best cut-offs for absolute change for the variables included in the model that might help physician in daily practice to decide if a variable has improved significantly.

A possible limitation of our study is the lack of analysis in the context of a real clinical trial and the fact that the cohort used for the definition/consensus generation is the same as per the provisional validation. Another potential limitation is the small sample of not improved patients since prevalence of the outcome could have the false positive/negative rate. The main strength resides in the large prospective collected data, which rarely is attempted in rheumatic diseases (1;2;13) and that enables a comprehensive evidence-based provisional validation of the JDM core set (19) and related definition of improvement.

In summary, PRINTO developed and validated a data driven provisional definition of improvement that will help standardize the conduct of JDM clinical trials and assist clinicians in daily practice when attempting to classify patients as either responders or non-responders. The definition of improvement derived here should undergo final validation in future controlled studies in different external cohorts of patients. This will allow examination of its discriminant validity in detecting a therapeutic response greater than placebo or an active comparator, and to establish whether refinements in currently available instruments are required.

Acknowledgments

We are indebted to Drs. Anna Tortorelli, Monica Tuffillo, and Elisabetta Maggi for their help in data handling, organization skills and overall management of the project. We are also thankful to Dr Luca Villa and Mr Michele Pesce for their help in data base development.

The authors wish to acknowledge the attendees of the Camogli, Italy “International Consensus Conference on defining improvement in JSLE and JDM” for their work during the meeting.

Supported by a grant from the European Union (contract no. QLGI-CT-2000-00514), by IRCCS G. Gaslini, Genoa, Italy, and by the National Institute of Health (Grant RO3 AI 44046). Lisa G. Rider was supported by the intramural research program of the NIH, National Institute of Environmental Health Sciences

Reference List

1. 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; 38:727–735. [PubMed: 7779114]
2. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum.* 1997; 40(7):1202–1209. [PubMed: 9214419]
3. Ruperto N, Ravelli A, Falcini F, Lepore L, De Sanctis R, Zulian F, et al. Performance of the preliminary definition of improvement in juvenile chronic arthritis patients treated with methotrexate. *Ann Rheum Dis.* 1998; 57(1):38–41. [PubMed: 9536821]
4. Alborno MA. ACR formally adopts improvement criteria for juvenile arthritis (ACR Pediatric 30). *ACR News.* 2002; 21(7):3.
5. Renal Disease Subcommittee of the American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria. The American College of Rheumatology response criteria for proliferative and membranous renal disease in systemic lupus erythematosus clinical trials. *Arthritis Rheum.* 2006; 54(2):421–432. [PubMed: 16453282]
6. Strand V, Gladman D, Isenberg D, Petri M, Smolen J, Tugwell P. Outcome measures to be used in clinical trials in systemic lupus erythematosus. *J Rheumatol.* 1999; 26(2):490–497. [PubMed: 9972993]
7. Smolen JS, Strand V, Cardiel M, Edworthy S, Furst D, Gladman D, et al. Randomized clinical trials and longitudinal observational studies in systemic lupus erythematosus: Consensus on a preliminary core set of outcome domains. *J Rheumatol.* 1999; 26(2):504–507. [PubMed: 9972996]
8. Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B, Feldman BM, et al. Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. *Rheumatology (Oxford).* 2003; 42(12):1452–1459. [PubMed: 12832713]

9. Ruperto N, Ravelli A, Cuttica R, Espada G, Ozen S, Porras O, et al. The Pediatric Rheumatology International Trials Organization criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: Prospective validation of the disease activity core set. *Arthritis Rheum.* 2005; 52(9):2854–2864. [PubMed: 16142708]
10. Ruperto N, Ravelli A, Oliveira S, Alessio M, Mihaylova D, Pasic S, et al. The Pediatric Rheumatology International Trials Organization/American College of Rheumatology provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus. Prospective validation of the definition of improvement. *Arthritis Rheum.* 2006; 55(3):355–363. [PubMed: 16739203]
11. Rider LG, Giannini EH, Harris-Love M, Joe G, Isenberg D, Pilkington C, et al. Defining Clinical Improvement in Adult and Juvenile Myositis. *J Rheumatol.* 2003; 30(3):603–617. [PubMed: 12610824]
12. 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.* 2001; 40(11):1262–1273. [PubMed: 11709610]
13. Rider LG, Giannini EH, Brunner HI, Ruperto N, James-Newton L, Reed AM, et al. International consensus on preliminary definitions of improvement in adult and juvenile myositis. *Arthritis Rheum.* 2004; 50(7):2281–2290. [PubMed: 15248228]
14. Oddis CV, Rider LG, Reed AM, Ruperto N, Brunner HI, Koneru B, et al. International consensus guidelines for trials of therapies in the idiopathic inflammatory myopathies. *Arthritis Rheum.* 2005; 52(9):2607–2615. [PubMed: 16142757]
15. Feldman BM, Rider LG, Reed AM, Pachman LM. Juvenile dermatomyositis and other idiopathic inflammatory myopathies of childhood. *Lancet.* 2008; 371(9631):2201–2212. [PubMed: 18586175]
16. Ramanan AV, Feldman BM. Clinical features and outcomes of juvenile dermatomyositis and other childhood onset myositis syndromes. *Rheum Dis Clin North Am.* 2002; 28(4):833–857. [PubMed: 12506775]
17. Rider LG. Outcome assessment in the adult and juvenile idiopathic inflammatory myopathies. *Rheum Dis Clin N Am.* 2002; 28:935–977.
18. Ruperto N, Martini A. International research networks in pediatric rheumatology: the PRINTO perspective. *Curr Opin Rheumatol.* 2004; 16(5):566–570. [PubMed: 15314496]
19. Ruperto N, Ravelli A, Pistorio A, Ferriani V, Calvo I, Ganser G, et al. The provisional Pediatric Rheumatology International Trial Organization/American College of Rheumatology/European League Against Rheumatism disease activity core set for the evaluation of response to therapy in juvenile dermatomyositis: a prospective validation study. *Arthritis Rheum.* 2008; 59(1):4–13. [PubMed: 18163404]
20. Rider LG, Feldman BM, Perez MD, Rennebohm RM, Lindsley CB, Zemel LS, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies. I. Physician, parent, and patients global assessments. *Arthritis Rheum.* 1997; 40(11):1976–1983. [PubMed: 9365086]
21. Lovell DJ, Lindsley CB, Rennebohm RM, Ballinger SH, Bowyer SL, Giannini EH, et al. Development of validated disease activity and damage indices for the juvenile idiopathic inflammatory myopathies - II. The Childhood Myositis Assessment Scale (CMAS): a quantitative tool for the evaluation of muscle function. *Arthritis Rheum.* 1999; 42(10):2213–2219. [PubMed: 10524696]
22. Rennebohm RM, Jones K, Huber AM, Ballinger SH, Bowyer SL, Feldman BM, et al. Normal scores for nine maneuvers of the childhood myositis assessment scale. *Arthritis Rheum Arthritis Care Res.* 2004; 51(3):365–370.
23. Huber AM, Feldman BM, Rennebohm RM, Hicks JE, Lindsley CB, Perez MD, et al. Validation and clinical significance of the childhood myositis assessment scale for assessment of muscle function in the juvenile idiopathic inflammatory myopathies. *Arthritis Rheum.* 2004; 50(5):1595–1603. [PubMed: 15146430]
24. 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 Arthritis Care Res.* 2003; 49(1):7–15.

25. Isenberg DA, Allen E, Farewell V, Ehrenstein MR, Hanna MG, Lundberg IE, et al. International consensus outcome measures for patients with idiopathic inflammatory myopathies. Development and initial validation of myositis activity and damage indices in patients with adult onset disease. *Rheumatology*. 2004; 43(1):49–54. [PubMed: 12867580]
26. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis Rheum*. 1994; 37:1761–1769. [PubMed: 7986222]
27. Ruperto N, Ravelli A, Pistorio A, Malattia C, Cavuto S, Gado-West L, et al. Cross-cultural adaptation and psychometric evaluation of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ) in 32 countries. Review of the general methodology. *Clin Exp Rheumatol*. 2001; 19(4):S1–S9. [PubMed: 11510308]
28. Landgraf, JM.; Abetz, L.; Ware, JE. *The CHQ User's Manual*. First Edition. Boston, MA, USA: The Health Institute, New England Medical Center; 1996.
29. Delbecq, AL.; Van de Ven, AH.; Gustafson, DH. *Group Techniques for Program Planning. A guide to nominal group and Delphi processes*. 1 ed.. Glenview, Ill, Scott: Foresman and Company; 1975.
30. Ruperto N, Meiorin S, Iusan SM, Ravelli A, Pistorio A. for the Paediatric Rheumatology International Trials Organisation (PRINTO). Consensus procedures and their role in pediatric rheumatology. *Curr Rheumatol Rep*. 2008; 10(2):142–146. [PubMed: 18460270]
31. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978; 8(4):283–298. [PubMed: 112681]
32. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press; 1977.
33. Landis JR, Koch GC. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33(1):159–174. [PubMed: 843571]

Table 1

Final domains and suggested variables included in the final PRINTO/ACR EULAR core set for the evaluation of response to therapy in JDM (adapted from ref. (19)).

Final Domains	Final core set Suggested variable(s)
Physician's global assessment of the patient's overall disease activity	10 cm VAS
Muscle strength	CMAS (or MMT)
Global JDM disease activity tool	DAS (or MYOACT or MITAX)
Parent's global assessment of the overall child's well-being	10 cm VAS
Functional ability assessment	C-HAQ
Health-related quality of life assessment	CHQ PhS

JDM = juvenile dermatomyositis, VAS = visual analogue scale; CMAS = Childhood Myositis Assessment Scale; MMT = Manual Muscle Testing; DAS = Disease Activity Score; C-HAQ = Childhood Health Assessment Questionnaire; CHQ PhS = Child Health Questionnaire physical summary score

Table 2

Example of 2 patients evaluated at the consensus. Readers, by using the related formulas, can calculate improvement/worsening of each variable and apply the JDM definition of improvement: at least 20% improvement from baseline in 3 of any 6 core set variables with no more than 1 of the remaining worsening by more than 30% which cannot be muscle strength

Variable (range) ↑worse, ↓worse* Formulas	Month 0 a	Month 6 b	Absolute difference c=b-a	% difference d=(c/a)*100	Outcome
Patient 1 (example of a patient who improved)					
Physician's global assessment of the patient's overall disease activity (0-10 cm VAS)↑	6.8	0.3	-6.5	-96%	Improved
Parent's global assessment of the overall child's well-being (0-10 cm VAS)↑	5.2	0	-5.2	-100%	Improved
CMAS (0-52 score) ↓	16	42	26	163%	Improved
DAS (0-20 score) ↑	12	4	-8	-67%	Improved
C-HAQ (0-3 score) ↑	2.3	0.5	-1.8	-78%	Improved
CHQ Physical summary score (PhS) (40-60 score) ↓	29.1	53.4	24.3	84%	Improved
Patient 2 (example of a patient who did not improve)					
Physician's global assessment of the patient's overall disease activity (0-10 cm VAS)↑	5.6	9.8	4.2	75%	Not improved
Parent's global assessment of the overall child's well-being (0-10 cm VAS)↑	1.5	5.6	4.1	273%	Not improved
CMAS (0-52 score) ↓	28	16	-12	-43%	Not improved
DAS (0-20 score) ↑	8	12	4	50%	Not improved
C-HAQ (0-3 score) ↑	1	1.5	0.5	50%	Not improved
CHQ Physical summary score (PhS) (40-60 score) ↓	23.9	18.6	-5.2	-22%	Not improved

* The ↑ indicates that higher tool score of that variable denotes worse activity (eg physician's global assessment of the patient's overall disease activity) while the ↓ indicates that lower tool score denotes worse values (eg CHQ physical summary score). VAS = visual analogue scale; CMAS = Childhood Myositis Assessment Scale; DAS = Disease Activity Score; C-HAQ = Childhood Health Assessment Questionnaire; CHQ PhS = Child Health Questionnaire physical summary score

Comparison between the difficult patients evaluated at the consensus conference (N=128) and the remaining patients of the sample collected (N=166); the total sample of 294 patients was used for the analysis of the final PRINTO/ACR/EULAR JDM core set of variables for the evaluation of response to therapy (19).

Table 3

Variables Mean±SD ↑ higher worse; ↓ lower worse*	Month 0		Month 6		p-value
	Validation patients N=166	Consensus Patients N=128	Validation patients N=166	Consensus Patients N=128	
Age at onset (yrs)	7.6±4.1	7.5±3.5			0.80*
Age at first observation at the center (yrs)	8.2±4.1	8.3±3.3			0.79*
Age at study visit (yrs)	8.7±4.1	9.6±3.8	9.3±4.1	10.1±3.8	0.09*
Disease duration (yrs)	1.1±1.8	2.1±2.4	1.7±1.9	2.6±2.4	<0.0001#
Females no (%)	96/166 (57.8%)	81/128 (63.3%)			0.34§
PRINTO/ACR/EULAR JDM core set:					
Physician's global assessment of the patient's overall disease activity (0–10 cm)↑	5.5±2.5	5.2±2.3	1.4±1.8	2.4±2.5	<0.0001#
Parent's global assessment of the overall child's well-being (0–10 cm)↑	5.6±2.8	4.8±2.9	1±1.5	2.4±2.3	<0.0001#
CMAS (0–52 score) ↓	24.1±14.7	26.4±14.0	43.6±10.4	38.9±11.9	0.16#
DAS (0–20 score) ↑	12.2±3.7	11.7±3.7	4.3±3.2	6.9±4.1	0.24#
C-HAQ (0–3) ↑	1.7±1.0	1.6±0.9	0.4±0.6	0.8±0.8	0.20#
CHQ Physical summary score (PHS) (40–60 score)↓	32.6±11.9	33.7±11.7	48.9±8.8	44.8±9.9	0.47#

* The ↑ indicates that higher tool score of that variable denotes worse activity (e.g. physician's global assessment of the patient's overall disease activity) while the ↓ indicates that lower tool score denotes worse values (e.g. Physical summary score).

CMAS: Childhood Myositis Assessment Scale; DAS: Disease Activity Score; C-HAQ: Childhood Health Assessment Questionnaire; CHQ: Child Health Questionnaire

* Student's t-test for independent samples

Mann-Whitney U test for independent samples

§ Pearson's chi-square

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Final results for the best definitions of improvement (DI) all with Kappa >0.8. Definitions are ordered according to the final score.

Definitions	Chi square*	% Sensitivity	% Specificity	% False Neg	% False Pos	AUC	Kappa	Rank #	Final Score#
3 of any 6 improved by at least 20%, no more than 1 worsened by more than 30% which cannot be muscle strength	90.3	98	87	9	3	92	0.86	131	113
3 of any 6 improved by at least 20%, no more than 2 worsened by ≥ 25%, which cannot be muscle strength (IMACS definition P1) (13)	90.3	98	87	9	3	92	0.86	104	90
3 of any 6 improved by at least 20%, no more than 2 worsened by more than 30% which cannot be muscle strength	90.3	98	87	9	3	92	0.86	81	70
2 of any 6 improved by at least 40%, no more than 1 worsened by more than 30% which cannot be muscle strength	85.2	97	87	13	3	92	0.84	61	51
2 of any 6 improved by at least 30%, no more than 1 worsened by more than 30% which cannot be muscle strength	90.1	100	78	0	5	89	0.85	46	39
3 of any 6 improved by at least 20%, no more than 1 worsened by more than 30%	84.3	98	83	10	4	90	0.83	36	30
3 of any 6 improved by at least 20%, no more than 2 worsened by more than 30%	84.3	98	83	10	4	90	0.83	17	14
3 of any 6 improved by at least 20%, no more than 2 worsened by ≥ 25% (IMACS definition P2) (13)	84.3	98	83	10	4	90	0.83	13	11
2 of any 6 improved by at least 40%, no more than 1 worsened by more than 30%	79.2	97	83	14	4	90	0.81	13	11
2 of any 6 improved by at least 40%, no more than 2 worsened by more than 30%, which	79.2	97	83	14	4	90	0.81	13	11

Definitions	Chi square*	% Sensitivity	% Specificity	% False Neg	% False Pos	AUC	Kappa	Rank#	Final Score#
cannot be muscle strength									
2 of any 6 improved by at least 30%, no more than 2 worsened by more than 30%, which cannot be muscle strength	84.3	100	74	0	6	87	0.82	6	5
3 of any 6 improved by at least 20% (IMACS definition P3) (13)	84.3	98	83	10	4	90	0.83	3	3
2 of any 6 improved by at least 30%, no more than 1 worsened by more than 30%	84.3	100	74	0	6	87	0.82	1	1

* All chi-squares correspond to a p value < 0.0001

The ranks were obtained by asking the attendees of the consensus meeting to decide upon which of the definitions of improvement that performed best were easiest to use and most credible (content validity). Then for each definition, the content validity rankings obtained were summed up and the resulting sum was multiplied by the corresponding value of the kappa statistic, to obtain the "Final score", that incorporated both statistical criteria and experts' judgments."

Table 5

Logistic regression model to predict improvement according to the evaluation of the participants at the consensus conference. Prediction was based on absolute change of the variables included in the final core set. Variables have been dichotomized according to the best cut-offs obtained from the ROC analysis (see footnote). Area under ROC curve of the model = 0.9.

Sample=102 ↑ higher worse; ↓ lower worse	Odds ratio	95% CI	Likelihood ratio test p value
Physician's global assessment of the patient's overall disease activity (0–10 cm)↑	11.0	2.1–56.7	0.003
CMAS (0–52 score) ↓	10.2	1.6–65.4	0.009
Parent's global assessment of the overall child's well-being (0–10 cm)↑	5.5	1.1–26.7	0.029
DAS (0–20 score) ↑	1.2	0.2–6.5	0.81
C-HAQ (0–3) ↑	0.9	0.1–5.5	0.88
CHQ Physical summary score (PhS) (40–60 score)↓	1.2	0.2–7.3	0.85

Best cut-offs for the variables included in the model: physician's global assessment of the patient's overall disease activity (absolute change): ≤ -1.3 (sensitivity 84.7%; specificity 82.6%); CMAS (absolute change): >4 (sensitivity 85.7%; specificity 87.0%); parent's global assessment of the overall child's well-being (absolute change): ≤ -1.4 (sensitivity 72.4%; specificity 78.3%); DAS (absolute change): ≤ -4 (sensitivity 78.6%; specificity 73.9%); C-HAQ (absolute change): ≤ -0.375 (sensitivity, 77.6% specificity 73.9%); CHQ PhS (absolute change): > 10.75 (sensitivity, 49.4% specificity 73.7%).