ARTHRITIS & RHEUMATISM
Vol. 40, No. 7, July 1997, pp 1202-1209

PRELIMINARY DEFINITION OF IMPROVEMENT IN JUVENILE ARTHRITIS

EDWARD H. GIANNINI, NICOLINO RUPERTO, ANGELO RAVELLI, DANIEL J. LOVELL, DAVID T. FELSON, and ALBERTO MARTINI

Objective. **To identify a core set of outcome variables for the assessment of children with juvenile arthritis (JA), to use the core set to develop a definition of improvement to determine whether individual patients demonstrate clinically important improvement, and to promote this definition as a single efficacy measure in JA clinical trials by the kappa statistic.**

Methods. *A* **core set of outcome variables wasestablished using a combination of statistical and consensus formation techniques. Variables in the core set consisted of 1) physician global assessment of disease activity; 2) parent/patient assessment of overall well-being; 3) functional ability; 4) number of joints with active arthritis; 5) number of joints with limited range of motion; and 6) erythrocyte sedimentation rate. To establish a definition of improvement using this core set, 21 pediatric rheumatologists from 14 countries met, and, using consensus formation techniques, scored each of 72 patient profiles as improved or not improved. Using the physicians' consensus as the gold standard, the chi-square, sensitivity, and specificity were calculated for each of 240 possible definitions of improvement. Definitions with sensitivity or specificity of <80% were eliminated. The ability of the remaining definitions to discriminate between the effects of active agent and those of placebo, using actual trial data, was then observed. Each definition was also ranked for face validity, and the sum of the ranks was then multiplied**

Results. **The definition of improvement with the highest final score was as follows: at least 30% improvement from baseline in 3 of any 6 variables in the core set, with no more than 1 of the remaining variables worsening by >30%. The second highest scoring definition was closely related to the first; the third highest was similar to the Paulus criteria used in adult rheumatoid arthritis trials, except with different variables. Thisindicates convergent validity of the process used.**

Conclusion. **We propose a definition of improvement for *JA*. Use of a uniform definition will help standardize the conduct and reporting of clinical trials, and should help practitioners decide if a child with JA has responded adequately to therapy. We are in the process of prospectively validating this definition and several others that scored highly.**

The assessment of clinical response in juvenile arthritis (JA) clinical trials is not standardized. Multiple measures of outcome are in use, and different trials may use different end points. Some of these end points have low validity characteristics and are insensitive to change (1), some are redundant (2), and some are nonreliable (poor reproducibility) (3). Additionally, there is little consensus about the amount of change in end points which signifies clinically important improvement or worsening. This lack of standardization may lead to inefficient trials that require larger-than-necessary sample sizes, an increased risk of statistical error, possible reporting bias, multiple or ambiguous interpretations of results, and an inability to compare multiple therapies using meta-analysis techniques (4-7).

Presented at the Sixtieth National Scientific Meeting of the American College of Rheumatology, Orlando, FL, October 1996.

Supported by a Clinical Science Grant from the Arthritis Foundation, by the Children's Hospital Research Foundation, and by Clinica Pediatrica, Istituto di Ricovero a Cura a Carattere Scientifico Policlinico S. Matteo Pavia, Italy. Support for the meeting in Pavia, Italy, was provided by Centacor (US), CibaGeneva Pharmaceutical (US), GenDerm Corporation (US), Lepetit (Italy), Nordmark (Italy), Pfizer, Inc. (US), Sandoz (Italy), Universita degli Studi di Pavia (Italy), and Wyeth-Ayerst Labs (US).

Edward H. Giannini MSc DrPH, Daniel J. Lovell, MD, MPH: Children's Hospital Medical Center, Cincinnati, Ohio; Nicolino Ruperto, MD, Angelo Ravelli, MD, Alberto Martini, MD: University di Pavia, Clinica Pediatrica, Istituto di Ricovero a Cura a Carattere Scientifico Policlinico S. Matteo, Pavia, Italy; David T. Felson, MD, MPH: Boston University Arthritis Center, Boston, Massachusetts.

Address reprint requests to Edward H. Giannini, MSc, DrPH, Rheumatology, Children's Hospital Medical Center, Pavilion Building 2-129, 3333 Burnet Avenue Cincinnati, OH 45229-3039.

Submitted for publication August 26, 1996; accepted in revised form February 3, 1997.

A similar situation in adult rheumatoid arthritis (RA) led to the development of a core set of outcome variables and a preliminary definition of improvement (8,9). This definition of improvement in adult RA is not appropriate for use in children with JA because of several factors: JA is considered a different disease entity, some core variables are less often abnormal or have lower scores in children than in adults, and their measurement is compromised due to age-related cognitive problems (e.g.. self-reported pain).

The purpose of this project was to develop and promulgate a core set of end points that can be used in future clinical trials in children with JA, to describe the amount of change in each variable that is considered clinically important, and to use the entire core set to develop a definition of improvement to aid in the classification of individual patients as either improved or not improved. The long-term goals are to increase the efficacy of JA clinical trials, and facilitate future meta-analyses of therapies for JA. We anticipated that the definition of improvement might also be useful to physicians when assessing patient improvement in routine practice. We used an approach similar to that developed by the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) project that led to the development of the American College of Rheumatology (ACR) core set and definition of improvement used in adult RA (10). The OMERACT core set has now been endorsed by the World Health Organization and International League Against Rheumatism. (11).

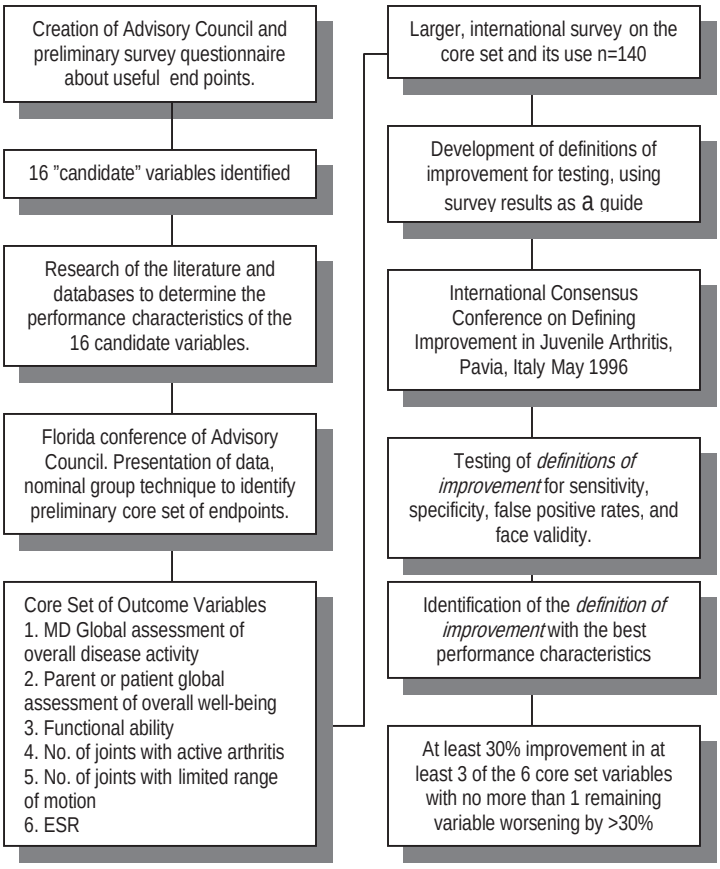

Figure 1. Process of choosing a core set of outcome variables, and, using the core set, a definition of improvement, for use in juvenile arthritis. ESR = erythrocyte sedimentation rate.

PATIENTS AND METHODS

A multistep process was used in developing the JA core set of variables and definition of improvement. This process is described below and is summarized in Figure 1.

Selection of the preliminary core set of response variables. In 1993, a 16-member Advisory Council was formed consisting of 1) members of the Rheumatology Section of the American Academy of Pediatrics, the Pediatric Section of the ACR, and the Arthritis Foundation, 2) OMERACT participants, and 3) private and academic practitioners. Prior to convening as meeting of this committee, a brief questionnaire was mailed to each member asking about response variables used when assessing clinical response in patients with A. The questionnaire listed 25 variables that had been used in the reporting of JAC clinical trials and asked the physicians to rank-order their top 6 choices. An "other" category was provided to add variables not included in the list. Variables were ranked in order of priority votes received. A total of 16 variables received votes, and these became known as the candidate variables for inclusion in the core set.

The questionnaire also asked if it would be acceptable to combine a cores set of variables into a single definition of improvement in order to dichotomously divide patients into those who improve by a clinically important amount and those who do not. A total of 91% of the respondents indicated that a valid definition for designating improvement would be acceptable to them.

The performance characteristics (validity, reliability sensitivity to change, redundancy) of all variables that received votes in the survey questionnaire were investigated using the literature and the core data bank of the Pediatric Rheumatology Collaborative Study Group (PRCSG) described previously (12). (Briefly, this databank contains all data from all trials of second0line agents studied by the PRCSG [n=551])

In 1994, a conference of the Advisory Council was convened in Florida, and the performance characteristics of each candidate variable were reviewed for attendees by the conference organizer (EHG). Next, using consensus formation techniques (nominal group technique[13]), attendees developed a preliminary core set of response variables. The cores set included the following 6 end points: 1) physicians' global assessment of overall disease activity (measured on a 10-cm visual analog scale [VAS]); 2) parent (or, if appropriate in age, patient) global assessment of overall well-being (measured on a 10-cm VAS); 3) functional ability; 4) number of joints with active arthritis (as defined by the ACR criteria: presence of swelling [not due to currently inactive synovitis or to bony

Table 1. Correlation of different juvenile arthritis core set variables cross-sectionally*

	MD global	Overall well-being	Functional ability	No. of active joints	No. of joints with LROM
Overall well-being	0.49				
Functional abilityt	0.38	0.57			
No. of active joints	0.56	0.30	0.36		
No. of joints with LROM	0.39	0.24	0.51	0.82	
ESR	0.53	0.43	0.33	0.22	0.29

* Correlations (r values) ?0.7 are considered evidence of redundancy. LROM = limited range of motion; ESR = erythrocyte sedimentation rate.
† Derived through regression analysis

.

enlargement] or, if no swelling is present, limitation of motion accompanied by heat, pain, or tenderness [14,15]); 5) number of joints with limited range of motion; and 6) erythrocyte sedimentation rate (ESR).

Ascertainment of international consensus on the core set of variables. Following the conference in Florida, several issues and questions remained to be addressed: 1) A broader consensus about the core set and its use to define improvement had to be obtained from other North American and international practitioners. 2) If the core set was to be used to define improvement, decisions would have to be made regarding how many of the variables would have to improve, and by how much, before practitioners could classify the patient as improved. 3) Since not all variables can be expected to improve, decisions would have to be made regarding how many variables could worsen, and by how much, while still allowing the patient to be classified as improved.

To address these issues, a questionnaire was designed and was mailed to a larger, more international sample of practitioners to obtain their reaction to the core set and its proposed use. The questionnaire was designed to provide preliminary data for organization of the International Consensus Conference, described below. The following mailing lists were used; to sample practitioners: 1) all members of the PRCSG and Pediatric Rheumatology Database Research Group (a consortium of pediatric rheumatology centers that collects data on the diagnoses of patients visiting their clinics); 2) all attendees of the Second European International Congress on Pediatric Rheumatology; and 3) other practitioners of pediatric rheumatology known to us who were not otherwise on the above mailing lists.

A total of 198 questionnaires were mailed, and 140 (71 %) were returned (88 from Europe, 52 from North America). Results validated the conclusions reached during the Florida conference: When variables were ranked according to their priority score, the same 6 chosen in Florida were scored highest in the larger survey. Based on the results, it was also determined that the median improvement that should be observed in order to classify a patient as improved within a given variable was 30% for all variables except for functional ability, which was 35%. It was further determined that the number of variables that should improve by the specified amount in order to classify a patient as clinically importantly improved was 3, and the number that could be ignored if they worsened was 2 (>30%). One hundred twenty four (89%) of the 140 respondents said that they would be willing to use the core set in combination to define improvement in individual patients. Differences between results from the US and the international respondents were inconsequential.

Assessment of multicollinearity. Having verified that the end points in the core set were acceptable to a large proportion of practitioners, the amount of multicollinearity existing between the variables had to be observed in larger and more diverse data sets. For this exercise, we used the PRCSG core databank (n = 551), an inception cohort of 227 children with JRA from the Cincinnati Special Treatment Center and from Pavia, Italy, who had taken part in a study of predictors of outcome (16,17), and a cohort of 55 patients from Pavia, Italy who had been followed up by one of us (AR) and assessed for each variable in the core set (18); r values greater than 0.7 were taken as evidence of collinearity. Only the number of joints with limited range of motion and the number of active joints showed evidence of collinearity (r = 0.82) (Table 1). The data banks used for this exercise had repeated measures for these 2 variables, and the r value for the change in the number of joints with limited range of motion versus the change in the number of joints with active arthritis was 0.65 (2). Thus, we concluded from this exercise that the core set variables were correlated but not excessively redundant.

Development and selection of a definition of improvement. Because the literature contained no definitions of improvement using combinations of the variables in the core set, we developed, for testing, a set of 240 definitions that seemed reasonable in consideration of the data from the international survey questionnaire discussed -above. A second conference, entitled International Consensus Conference on Defining Improvement in Juvenile Arthritis, was held in Pavia, Italy in May 1996. The meeting was attended by 21 pediatric rheumatologists from 14 different countries (see Acknowledgments for list of the 18 attendees who are not authors of this report), and was facilitated by 2 of us (EHG, NR) with expertise in nominal group process. The overall goal of the meeting was to decide upon a preliminary definition of improvement based on the core set of end points, using a combination of statistical and consensus formation techniques. In order to achieve this, there were 5 objectives, which are described in consecutive order below.

1. Rate each of 72 paper patient profiles as clinically importantly improved or not improved, using nominal group technique. Existing clinical trial data were used for patient profiles presented to conference attendees for evaluation of response. The Italian Pediatric Rheumatology Study Group had conducted an open-label, uncontrolled trial of methotrexate (MTX), 10 mg/m2/week (n = 94). This trial was selected as

a source of patients because it was the only one known to us that had measured each of the core set variables. The profiles elected were those near the threshold level of improvement, as determined by the international survey described above (e.g., patients who showed 100% improvement in all outcome variables were not good candidates for inclusion because everyone would agree that the patient had improved, and all the definitions of improvement would categorize the patient as improved.) The Juvenile Arthritis Functional Assessment Report (19) had, been used to assess functional ability. For each core set variable, absolute values at baseline and 6 months were shown, as well as the absolute difference and percent change from baseline. Participants were randomized into equally sized nominal groups and asked to silently rate each of 72 patient profiles as clinically importantly improved or not improved. The moderator then asked each member how he or she had voted on each patient. If an 80% consensus about whether the patient was improved or not improved was not achieved, the case was discussed in round-robin fashion and a second vote taken. If 80% consensus was still not attained, the patient profile was declared uninterpretable and not used further in the nominal group. A plenary session was then held in an attempt to resolve those cases scored discordantly by the groups.

2. Using the physicians' consensus judgment as the gold standard, calculate the percent false positive and false-negative rates, chi-square, sensitivity, and specificity for each definition of improvement. We evaluated the ability of the 240 candidate definitions of improvement to classify individual patients as improved or not improved and then assessed the agreement between the "decision" of the criteria and the consensus of the physicians. We used only patient profiles for which physician consensus was achieved. For each definition, we calculated the chi-square (1 degree of freedom) and the corresponding P value, sensitivity (ability of the definition to identify a patient as improved who had been classified as improved by the physicians), specificity (ability of the definition to identify a patient as not improved who had been classified as not improved by the physicians), rate of false-positivity ([number falsely identified as improved by criteria/all patients identified as improved] x 100), and rate of false-negativity ([number falsely identified as not improved by the criteria/all. Patients identified as not improved] x 100). Those definitions of improvement showing either a sensitivity or specificity of <80% were eliminated from further consideration. On the next day, the results of the statistical exercises were presented to the group of physicians.

3. Observe the ability of the remaining definitions of improvement to discriminate between active agent and placebo using existing trial data. This phase of the exercise was very limited in scope because placebo-controlled clinical trial data-bases in which all of the core set variables had been measured were not in existence. We used data from the 10 mg/m /week MTX-versus-placebo trial published previously (20), for 2 reasons: the active treatment arm (MTX) produced large effect sizes in the original trial analysis, and many (but not all) of the core set variables had been measured. Those core set variables that had not been measured were derived through regression analysis using variables that had been assessed and that correlate with core set variables, as discussed in greater detail below.

4. *Using nominal group technique, decide upon which of the remaining definitions of improvement is easiest to use and most credible (highest face validity).* The attendees were again split into 2 groups, and, using nominal group technique, were asked to decide upon which of the definitions of improvement that performed best were easiest to use and most credible (face validity), ranking the 5 best from 5 highest face validity) to 1 (lowest).

5. *Multiply the face validity score by the kappa values to obtain the "best" definition.* We used the kappa statistic as an additional measure of agreement between the physicians' evaluation and the definitions; K values ≥ 0.7 were considered to be evidence of agreement. Finally, we combined the face validity rankings by the 2 nominal groups and multiplied this sum by its kappa statistic to obtain the "best" definition (9).

RESULTS

Results of scoring the patient profiles. The 21 physicians scored 31 of the 72 patient profiles as clinically importantly improved, 27 as not improved, and 14 as uninterpretable. In no case did one nominal group rate a patient as improved and the other group rate the same patient as not improved.

Identification of 9 definitions of improvement as the best performers. Nine of the 240 definitions of improvement showed a sensitivity and specificity ≥ 80%. These 9 definitions, their corresponding chi-square values, *P* values, percent false-positive and false-negative rates, and kappa statistics are shown.in Table 2.

Face validity of the 9 definitions of improvement, and final resolution. After presentation of the above data, the attendees, using nominal group technique, ranked the 9 definitions for face validity on a 1-5 scale, with 5 being the highest. Another 5 definitions were added by the participants to the list to be ranked for face validity, but all received a low ranking in the final vote (data not shown). The sum of the combined ranks from the 2 groups is presented in Table 2 (range 9-70). Finally, the sum of the ranking was multiplied by its kappa statistics to obtain the final. score (range 7.4460.80), and the definitions of improvement with the highest final score were identified. The definition of improvement that scored highest was as follows: at least 30% improvement in at least 3 core set variables, with no more than 1 of the remaining variables deteriorating by more than 30%.

As can be seen in Table 2, the definitions that scored second and third highest were very similar to the gist. The third definition is similar to the Paulus criteria used in adult RA (21). The similarity of the top-ranking definitions indicates convergent validity of the process.

Table 2. Final results for the 9 best definitions of improvement*

Definition of improvement	χ^2†	Sensitivity, %	Specificity, %	False-positive, %	False-negative, %	*K* statistic	Sum of face validity scores	Final score
3 of any 6 improved by ≥30%; no more than 1 worse by >30%	43.8	100	85	11	0	0.87	70	60.80
3 of any 6 improved by ≥30%o; no more than 2 worse by >30%	40.7	100	81	14	0	0.84	54	45.23
4 of any 6 improved by ≥20%; no more than 1 worse by >30%	43.0	94	93	6	7	0.86	41	35.32
MD global improved by ≥20%; 3 of any remaining 5 improved by ≥20%; no more than 1 worse by >30%	43.0	94	93	6	7	0.86	33	28.43
4 of any 6 improved by ≥20%; no more than 2 worse by >30%	39.7	94	89	9	8	0.83	25	20.67
3 of any 6 improved by ≥30%; none worse by >30%	27.9	81	89	11	20	0.69	22	15.27
MD global improved by ≥20%; 2 of any remaining 5 improved by ≥30%; no more than 1 worse by >30%	36.9	97	81	14	4	0.80	19	15.16
2 of any 6 improved by ≥40%; no more than 1 worse by >30%	30.3	90	81	15	12	0.72	11	7.96
MD global improved by ≥20%; 3 of any remaining 5 improved by ≥20%; no more than 2 worse by >30%	39.7	94	89	9	8	0.83	9	7.44

* See text for definitions of false-positive and false-negative rates and for other details.
† P values < 0.001.

Discriminant validity of the 9 definitions of improvement. As discussed above, our ability to assess the discriminant ability of these 9 definitions of improvement was compromised due to the lack of placebo-controlled clinical trial data sets that use the core set of variables.

To make some attempt at determination of discriminant validity, we used the 10 mg/m^2/week MTX clinical trial data set of the PRCSG (20). The results are presented in Table 3. While the percent of MTX-treated patients who improved was high (range 56.7-80.0%) for

Table 3. Discriminant validity of the 9 definitions of improvement that yielded sensitivity and specificity ≥80%*

Definition of improvement	% of MTX-treated patients improved (n = 38)	% of placebo-treated patients improved (n = 39)
3 of any 6 improved by ≥30%; no more than 1 worse by >30%	63.3	40.0
3 of any 6 improved by ≥30%; no more than 2 worse by >30%	80.0	43.3
4 of any 6 improved by ≥20%; no more than 1 worse by >30%	63.3	43.3
MD global improved by ≥20%; 3 of any remaining 5 improved by ?20%; no more than 1 worse by >30%	63.3	43.3
4 of any 6 improved by ≥20%; no more than 2 worse by >30%	76.7	46.7
3 of any 6 improved by ≥30%; none worse by >30%	56.7	30.0
MD global improved by ≥20%; 2 of any remaining 5 improved by ≥30%; no more than 1 worse by >30%	63.3	40.0
2 of any 6 improved by ≥40%; no more than 1 worse by >30%	63.3	40.0
MD global improved by ≥20%; 3 of any remaining 5 improved by ≥20%; no more than 2 worse by >30%	66.7	43.3

* Data from a clinical trial of methotrexate (MTX; 10 mg/m^2/week) versus placebo in juvenile arthritis (20) were used for this analysis

all definitions, the percent of placebo-treated patients same who improved according to these definitions was also high (30.0-46.7%). These data must be interpreted with considerable caution. Since this data set does not contain all of the core set variables, some had to be derived and converted from other scales of measure. The physician global evaluation and parent/patient global evaluation had to be derived from a scale that contained only the categories much better, better, same, worse, and much worse. Functional ability had to be derived through regression analysis using the number of joints with limited range of motion, with which it has shown good correlation in other data sets.

DISCUSSION

Using a consensus formation and statistical approach, our results suggest that improvement in patients with JA can, be defined as follows: 3 of any 6 core set variables improved by at least 30%, with no more than 1 of the remaining variables worsened by more than 30%. The variables included in the core set are 1) physician global assessment of disease activity; 2) parent/patient global assessment of overall well-being (each scored on a 10-cm VAS); 3) functional ability; 4) number of joints with active arthritis; 5) number of joints with limited range of motion; and 6) ESR. This set of end points has an intuitive appeal to the clinician in that it combines aspects of the articular examination with true outcome (functional ability and parent/patient assessment of overall well-being).

We propose these end points as a core set only; investigators can measure as many other variables as they deem appropriate. Indeed, we believe other variables should still be measured and reported. Furthermore, the core set does not have to serve as the basis for the primary outcome, but the core set variables should always be measured. Patients should be evaluated as improved or not improved by comparing the values of the core end points at the end of the trial, or at withdrawal from trial (intent-to-treat approach), with baseline values.

The definition of improvement developed here shows high sensitivity and specificity, and low false-positive and false-negative rates. Moreover, the top 3 definitions with the highest final score showed convergent validity: all 3 are very similar to each other, the third highest is similar to the Paulus criteria (with slightly different variables) that have been used by adult rheumatologists for many years, and the international questionnaire survey done in 1995 yielded nearly the definition of improvement in terms of the number of variables that must improve, the percent change that must be attained in order to call the variables improved, and the number of variables that could worsen and the patient still be classified as improved. Because of the relatedness and similarity in performance of the top 3 definitions identified here, we intend to observe the relative performance of each in prospective validation studies.

Several issues remain unresolved. Because of a lack of adequate data sets, we were not able to provide firm conclusions about the discriminant ability of the definitions under placebo-controlled trial. Prospective validation of the definition will be a necessary next step.

We did not specify which of the functional ability tools currently available or under development should be used. This issue may be of little importance provided the instrument used has been validated in the pediatric population, and all investigators use the same instrument throughout a trial. When the instruments now under development complete validity testing, it may become possible to recommend one that combines measurements of functional ability and health-related quality of life.

The issue of redundancy between the number of joints with active arthritis and the number of joints with limited range of motion will have to be reviewed carefully as the definitions are tested further. One suggestion is to use the number of joints with swelling, rather than active joints, as in the core set for adult RA. However, this is not likely to solve the problem. The number of joints with swelling also correlates with the number with limited range of motion, and the further problem of joints that are swollen due to bony enlargement without currently active synovitis arises.

The lack of valid, widely available laboratory markers of inflammation in children with JA leaves the core set with only the ESR as a biochemical marker of response. Some children enrolled in trials of second-line agents have a normal ESR throughout the study, thus compromising the utility of the definition of improvement. As the science advances, investigators should feel free to replace the ESR with another, more specific laboratory marker. This can be done in the prospective validation phase.

We did not attempt to create different definitions of improvement for the various phenotypes of JA. Rather, we attempted to make the definition robust enough to cover all types of JA, focusing on the central features of arthritis, function, and overall well-being. The primary outcome in a study of systemically ill

children may be fever reduction, but the core set will still be useful to assess.

Other issues that remain include whether it would be more desirable to define levels of improvement (none, mild, moderate, marked) by using an approach similar to that used by the European League Against Rheumatism (22), the possibility of using reduced joint counts, the acceptance of the definition of improvement and Drug Administration, and the impact of the chang-ing classification criteria for the idiopathic arthritides of childhood (23).

In summary, we propose a definition of JA improvement that has >80% sensitivity and specificity and high face validity. Use of a uniform definition will help standardize the conduct and reporting of clinical trials, and may help practitioners decide if a child with JA has responded adequately to therapy.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the following indi- viduals, who were attendees of the Pavia, Italy, International Consensus Conference on Defining Improvement in Juvenile Arthritis and are members of the Pediatric Rheumatology Inter- national Trials Organization, for their diligent work during the meeting: Boel Andersson Gare, MD (Jonkoping, Sweden), Zsolt Balogh, MD (Budapest, Hungary), James T. Cassidy, MD (Co- lumbia, MO), Jaime de Inocencio, MD (Madrid, Spain), Ciaran M. Duffy, MD (Montreal, Quebec, Canada), Flavio Fantini, MD (Milan, Italy), Wietse Kuis, MD (Utrecht, The Netherlands), Joseph E. Levinson, MD (Cincinnati, OH), Jose A. Melo-Gomes, MD (Lisbon, Portugal) Hartmut Michels, MD (Garmisch-Partenkirchen**,** Germany), Ross E. Petty, MD, PhD (Vancouver, British Columbia, Canada), Anne-Marie Prieur, MD (Paris, France), Lisa G. Rider, MD (Bethesda, MD), Anneli Savolainen, MD (Heinola, Finland), Alexander Shaikov, MD (Moscow, Rus-sia), Earl D Silverman, MD (Toronto, Ontario, Canada), Filip van den Bosch, MD (Gent, Belgium), Patricia Woo, MD (Lon-don UK). The authors also wish to thank the members of the OMERACT Committee for their advice concerning the method- ology used in this project. A special thanks to Enrico Solcia, MD, Scientific Director of the Istituto di Ricovero a Cura a Carattere Scientifico Policlinico S. Matteo, Pavia, Italy for extending his support for the Pavia meeting.

REFERENCES

1. Giannini EH, Brewer EJ: Poor correlation between the erythro-cyte sedimentation rate and clinical activity in juvenile rheumatoid arthritis. Clin Rheumatol 6:197-201, 1987
2. Ruperto N, Giannini EH: Redundancy of conventional articular response variables used in juvenile chronic arthritis clinical trials. Ann Rheum Dis 55:73-75, 1996
3. Giannini EH, Stillman CM, Brewer EJ: Measuring grip strength inchildren with the Martin Vigorimeter and adapted sphygmoma-nometer cuff. Occup Ther J Res 4:234-236, 1984
4. Anderson JJ, Felson DT, Meenan RE, Williams HJ: Which traditional measures should be used in rheumatoid arthritis clini-cal trials? Arthritis Rheum 32:1093-1099, 1989
5. Felson DT, Anderson JJ, Meenan RF: Time for changes in the design, analysis, and reporting of rheumatoid arthritis trials. Arthritis Rheum 33:140-149, 1990
6. Felson DT: Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials. J Rheumatol 20:531-534, 199
7. Giannini EH, Lovell DJ, Hepburn B: FDA draft guidelines for the 3 clinical evaluation of antiinflammatory and antirheumatic drugs in children: executive summary. Arthritis Rheum 38:715-718, 1995
8. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, Furst D, Goldsmith C, Kieszak S, Lightfoot R, Paulus H, Tugwell P, Weinblatt M Widmark R, Williams HJ, Wolfe F: The American College of Rheumatology preliminary core set of dis-ease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 36:729-740, 1993
9. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, Katz LM, Lightfoot R Jr, Paulus H, Strand V, Tugwell P, Weinblatt M, Williams HJ, Wolfe F, Meszak S: American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 38:727-735, 1995
10. Fried BJ, Boers M, Baker PRA: A method for achieving consensus on rheumatoid arthritis outcome measures: the OMERACT conference process. J Rheumatol 20:548-551, 1993
11. Boers M, Tugwell P, Felson DT, van Riel PLCM, Kirwan JR, Edmonds JP, Smolen JS, Khaltaev N, Muirden KD: World Health Organization and International League of Associations for Rheu-matology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol 21 (suppl 41):86-89, 1994
12. Giannini EH, Cassidy JT, Brewer EJ, Shaikov A, Maximov A, Kuzmina N: Comparative efficacy and safety of advanced drug therapy in children with juvenile rheumatoid arthritis. Semin Arthritis Rheum 23:34-46 1993
13. Delbecq AL, van de Ven AH, Gustafson DH: Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes. Glenview, IL, Scott, Foresman and Co., 1975
14. Brewer EJ Jr, Bass J Baum J, Cassidy JT Fink C, Jacobs J, Hanson V, Levinson JE, Schaller J, Stillman JS: Current proposed revision of JRA criteria. Arthritis Rheum 20:195-199, 1977
15. Cassidy JT, Levinson JE, Bass JC, Baum J, Brewer EJ Jr, Fink CW, Hanson V, Jacobs JC, Masi AT, Schaller JG, Fries JF, McShane D, Young D: A study of classification criteria for a diagnosis of juvenile rheumatoid arthritis. Arthritis Rheum 29: 274-281, 1986
16. Ruperto N, Levinson JE, Ravelli A, Shear ES, Tague BL, Murray K, Martini A, Giannini EH: Long-term health outcomes and quality of life in American and Italian inception cohorts of patients with juvenile rheumatoid arthritis. I. Outcome status. J Rheumatol24:945-951, 1997
17. Ruperto N, Ravelli A, Levinson JE, Shear ES, Murray K, Tague BL, Martini A, Glass DN, Giannini EH: Long-term health out comes and quality of life in American and Italian inception cohorts of patients with juvenile rheumatoid arthritis. II. Early predictors of outcome. J Rheumatol 24:952-958, 1997
18. Ravelli A, Viola S, Ruperto N, Corsi B, Ballardini G, Martini A: Correlation between conventional disease activity measures in juvenile chronic arthritis. Ann Rheum Dis (in press)
19. Howe S, Levinson J, Shear E, Hartner S, McGirr G, Schulte M, Lovell D: Development of a disability measurement tool for juvenile rheumatoid arthritis: the Juvenile Arthritis Functional Assessment Report for children and their parents. Arthritis Rheum 34:873-880, 1991
20. Giannini EH, Brewer EJ, Kuzmina N, Shaikov A, Maximov A, Vorontsov 1, Fink CW, Newman AJ, Cassidy JT, Zemel LS, for the

Pediatric Rheumatology Collaborative Study Group: Methotrexate in resistant juvenile rheumatoid arthritis: results of the U.S.A.-U.S.S.R. double-blind, placebo-controlled trial. N Engl J Med 326:1043-1049, 1992

21. Paulus HE, Egger MJ, Ward JR, Williams HJ, and the Cooperative Systematic Studies of Rheumatic Diseases Group: Analysis of improvement in individual rheumatoid arthritis patients treated with disease-modifying antirheumatic drugs, based on the findings in patients treated with placebo. Arthritis Rheum 33:477-484, 1990
22. Van Gestel AM, Prevoo MLL, van't Hof MA, van Rijswijk MH,van de Putte LBA, van Riel PLCM: Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism criteria.Arthritis Rheum 39:34-40, 1996
23. Fink CW: Proposal for the development of classification criteria for idiopathic arthritides of childhood. J Rheumatol 22:1566-1569, 1995