ACR and EULAR Improvement Criteria Have Comparable Validity in Rheumatoid Arthritis Trials

ANKE M. van GESTEL, JENNIFER J. ANDERSON, PIET L.C.M. van RIEL, MAARTEN BOERS, CEES J. HAAGSMA, BILL RICH, GEORGE WELLS, MARY L.M. LANGE, and DAVID T. FELSON

ABSTRACT. We compared the validity of the American College of Rheumatology (ACR) and the European League of Associations for Rheumatology (EULAR) definitions of response in rheumatoid arthritis (RA) clinical trials. US: ACR and EULAR improvement criteria were calculated in 7 large randomized RA clinical trials. The discriminant validity of the response criteria between treatment groups was studied using the Mantel-Haenszel chi-squared value. To compare both sets of criteria the chi-squared ratio was determined for each trial. Europe: In 2 large randomized RA clinical trials, ACR and EULAR criteria were calculated, once with extensive and once with 28 joint counts. The classification of patients with these 4 criteria were compared with each other using cross tables. We further studied the difference in response between treatment groups per trial, the association of response with patient and investigator assessment of improvement, and the association of response with radiological progression. US: The chi-squared ratio for most trials was close to 1. There was no clear pattern suggesting that the discriminant validity of the ACR criteria was stronger than the discriminant validity of the EULAR definition of response or vice versa. Europe: Conflicting results between ACR and EULAR were present in only 3% of patients in both trials. The discriminant validity of all 4 criteria (ACR and EULAR with reduced and extensive joint counts) was comparable. All criteria were related with the overall assessment of improvement by both investigator and patient. The association with radiographic progression was comparable for EULAR and ACR improvement criteria. There is a high level of agreement between ACR and EULAR improvement classification, and their validity is equivalent. The discriminating potential of the criteria between treatment groups is comparable, as is the association with patient's and investigator's overall assessment and with radiographic progression. (J Rheumatol 1999;26:705-11)

Key Indexing Terms:
RHEUMATOID ARTHRITIS TRIAL EVALUATION IMPROVEMENT CRITERIA

There is still no cure for rheumatoid arthritis (RA), but in recent years the number of available disease modifying agents has grown considerably. To evaluate which agents perform best, the reader should be able to compare the results of respective trials. However, in addition to different study designs, most trials present different endpoint measures. Therefore a minimum set of valid disease activity variables, the core set, was proposed a few years ago[1,2]. Using standard measurement techniques and measurement protocols, one could then compare the group results of these 7 or 8 separate variables among trials. Besides group results, it is important to know how many patients actually improved, i.e., is a good group result based on a large number of patients improving moderately, or on a small number of patients with a considerable improvement. The choice for such an individual response measure should also be uniform among trials to improve comparability.

Using the core set of disease activity variables, two different definitions of response or improvement in RA clinical trials have been promulgated. One has been recommended by the American College of Rheumatology (ACR) Committee as a definition of improvement in RA clinical trials[3]. For each individual patient in a trial, it tests whether the patient has experienced clinical improvement by evaluating their percentage of improvement in core set variables during the trial. Generally speaking, if the patient experiences at least 20% improvement in multiple variables simultaneously, they are defined as having satisfied the definition of clinical improvement, a definition that corresponds to clini-

From the Department of Rheumatology, University Hospital Nijmegen, Nijmegen, The Netherlands; CHQOER Unit, VAMC, Bedford, MA, USA; Department of Clinical Epidemiology, Free University, Amsterdam, The Netherlands; Amgen rhIL-1ra Clinical Research Product Team, Boulder, CO; Immunex Corporation, Seattle, WA; Boston University Arthritis Center, Boston, MA, USA.

The US analyses were supported by NIH grant AR20613.

A.M. van Gestel, MSc, Department of Rheumatology, University Hospital Nijmegen; J.J. Anderson, PhD, CHQOER Unit, VAMC; P.L.C.M. van Riel, MD, PhD, Professor of Rheumatology, Department of Rheumatology, University Hospital Nijmegen; M. Boers, MSc, MD, PhD, Professor of Clinical Epidemiology, Department of Clinical Epidemiology, Free University; C.J. Haagsma, MD, PhD, Department of Rheumatology, University Hospital Nijmegen; B. Rich, MSc, Amgen rhIL-1ra Clinical Research Product Team; G. Wells, MSc, PhD, Professor, Department of Clinical Epidemiology, Ottawa Civic Hospital, Ottawa; M.L.M. Lange, MS, Manager Statistics Biometrics, Immunex Corporation; D.T. Felson, MD, MPH, Boston University Arthritis Center.

Address reprint requests to Dr. A.M. van Gestel, Department of Rheumatology, University Hospital Nijmegen, PO Box 9101, 6500 HB Nijmegen, the Netherlands; E-mail: A.vanGestel@reuma.azn.nl

cians' impressions that the patient has experienced improvement. The ACR definition focuses on patient change, not on the absolute state of activity of their disease. The European League Against Rheumatism (EULAR) also developed a definition of response in RA using some core set variables[4]. This definition of response differs from the ACR response conceptually. First, rather than defining improvement or no improvement, the EULAR definition classifies persons into 3 groups, those experiencing no response, those experiencing moderate response, and those experiencing a good response. Also, the EULAR response criteria factors in both change and the absolute level of disease. If a patient improves greatly, yet does not reach a certain level of disease inactivity, she cannot be characterized as having experienced a good response. The EULAR response criteria are computed using an index of activity in RA, the Disease Activity Scale (DAS)[5]. The DAS combines information from the Ritchie Articular Index, the swollen joint count, the erythrocyte sedimentation rate, and patient global assessment of their disease activity.

The criteria sets have been validated separately both for full and limited joint counts, and appear to perform well. It is not clear which definition of improvement or response is better able to differentiate between treatments in RA clinical trials. Discriminant validity can be defined as the ability to distinguish between clinically relevant levels of treatment efficacy. In the setting of this study, this was equated with the statistical ability to distinguish between the responses of treatment groups in a trial. The goal of our analysis was to compare the validity of the ACR and EULAR definitions of response. For this purpose datasets from several mainly North American groups[6-12] were analyzed by a US research group (JA/DF), and two European datasets were analyzed by a European (EU) research group (AG/PR). Whereas the US group focused on discriminant validity, the EU group also studied construct and criterion validity.

MATERIALS AND METHODS

US group. To compare the discriminant validity to the ACR and EULAR definitions of response or improvement, we used data from a group of randomized double blind controlled trials in RA, most of them placebo controlled. Using intent-to-treat (last observation brought forward) data on patients at the end of the trial, compared to patient baseline, we characterized the number of responders in each treatment group in each trial according to ACR improvement criteria (Table 1). Using a similar approach after computing DAS scores, we characterized EULAR non, moderate, and good responder rates in each treatment group in each trial (Table 1). For the DAS computation, we used the following formula[5]:

$$DAS = 0.53938 \text{ sqrt (RAI)} + 0.06465 \text{ (swollen joint count)} + 0.3330 \ln (\text{ESR}) + 0.00722 \text{ (patient global)}$$

For each trial using both ACR and EULAR response definitions, we computed a Mantel-Haenszel chi-squared statistic to differentiate response on active treatment from response control treatment. This chi-squared value was our measure of discriminant validity.

The trials we analyzed are as follows:
(1) Cooperative systematic studies of rheumatic disease trial of methotrexate versus placebo[6]. (2) Cooperative systematic studies of rheumatic disease trial of penicillamine versus placebo from which we removed the low dose of penicillamine group[7]. (3) Cooperative systematic studies of rheumatic disease trial of injectable gold, auranofin versus placebo[8]. (4) The COBRA (Combination Therapy in Rheumatoid Arthritis) trial testing high dose corticosteroids, methotrexate, and sulfasalazine versus sulfasalazine and double placebo in early RA[9]. (5) Trial of methotrexate versus auranofin in early RA[10]. (6) Trial of the combination of methotrexate and cyclosporine versus methotrexate and placebo in partial methotrexate responders with RA[11]. (7) Trial of TNFR:Fc (16 mg/m^2 dose only) versus placebo in RA[12].

As in previous analyses[3] we combined data from 3 Cooperative

Table 1. RA improvement criteria.

ACR Improvement Criteria	
≥ 20% improvement in	Tender joint count, and Swollen joint count, and At least 3 of the following: ESR of CRP Investigator assessment of global disease activity Patient assessment of global disease activity Patient assessment of pain Physical disability

EULAR (EULAR28) Response Criteria

Reached Value		Change in DAS or DAS28 from Baseline		
DAS28	DAS	> 1.2	> 0.6 and ≤ 1.2	≤ 0.6
≤ 3.2	≤ 2.4	Good		
> 3.2 and ≤ 5.1	> 2.4 and ≤ 3.7		Moderate	
> 5.1	> 3.7			Non

DAS: Disease Activity Score; DAS28: Disease Activity Score including 28 joint counts for tenderness and swelling.

Systematic Studies in the Rheumatic Diseases (CSSRD) trials of methotrexate, penicillamine, and injectable gold characterizing the active treatment group in these trials as "strong second-line drugs." We performed one pooled analysis of these 3 trials. Considering the auranofin group as a "weak second-line drug"[13], we looked at the auranofin versus placebo trial of the CSSRD trial that included gold, auranofin, and placebo in a second analysis. We analyzed the other trials separately, the COBRA trial, the methotrexate versus auranofin trial, the combination methotrexate/cyclosporine trial, and the TNFR:Fc trial. Because the CSSRD trials do not include measures of self-reported disability, which are part of the core sets, we substituted grip strength as we have done in previous analyses of these trials[3]. Our analyses have shown that grip strength correlates with self-reported functional disability at values ranging from 0.44 to 0.60.

Statistics. The Mantel-Haenszel chi-squared test permitted us to compare the discriminant validity of these two definitions of response, even though one is dichotomous and the other has 3 levels. To further compare these two measures, we computed a chi-squared ratio, using the ACR chi-squared divided by the EULAR chi-squared. If the chi-squared ratio is close to one, then the two measures have equivalent discriminant validity.

European group. Journal articles (since 1990) and abstracts of all RA clinical trials of at least 24 weeks' duration were screened on the basis of the following criteria:

- report of all components of ACR and EULAR improvement criteria (baseline and endpoint): individual joint counts for tenderness and swelling (at least 28 joint count), ESR, or C-reactive protein (CRP), patient and investigator assessment of global disease activity, patient assessment of pain, and a measure of physical disability;
- patient and investigator assessment of improvement (endpoint);
- radiographic damage score (baseline and endpoint).

If information was available for all these variables, the authors concerned were asked for an extract of their dataset.

For all studies treatment response was calculated per patient using both ACR and EULAR criteria. When possible, two sets of criteria were calculated: (1) with extensive joint counts (ACR) or Ritchie Articular Index and 44 swollen joint count (EULAR), and (2) with 28 joint counts (ACR28 and EULAR28) (Table 1)[14-16]. The classification of treatment response with ACR (ACR28) and EULAR (EULAR28) criteria were compared using cross tables. The following aspects of validity of the improvement criteria were studied: I. the ability to detect differences between treatment groups (discriminant validity); II. the association with patient's and investigator's assessment of improvement (construct validity); and III. the association with radiographic progression (criterion validity).

Statistics. Because of different designs, all trials were analyzed separately. Mantel-Haenszel chi-squared statistics compared the differences in response between treatment groups. Wilcoxon two sample tests (ACR) and Kruskal-Wallis tests (EULAR) tested the association between response and patient and investigator assessment of improvement. Analysis of variance or Kruskal-Wallis (when the progression score could not be transformed to a more normal distribution) tests studied the association between response and radiographic progression.

RESULTS

US Group

Results of our analyses are shown in Table 2. We analyzed a total of 7 data sets. For the COBRA trial, we looked separately at 16 weeks (9 weeks after the end of the oral prednisolone pulse) and at 28 weeks (the last assessment while taking low dose prednisolone). For the other trials, we examined only the end point of the trial. Generally speaking, using the ACR definition of improvement, improvement on active treatment was much more common than improvement on control treatment with the rate of improvement on control being lowest in the Cooperating Clinics Trials, and in the methotrexate/cyclosporine combination trial. Improvement rate was highest in the control treatment arm in the COBRA trial, especially at the end of the trial. The Mantel-Haenszel chi-squared was high using both the ACR and EULAR criteria for almost all of the trials, indicating good discrimination between treatment arms.

The chi-squared ratio (see Table 2) for most trials was close to 1, with some trials having ratios above 1 (e.g., methotrexate/cyclosporine combination trial) and others having ratios below 1 (e.g., auranofin vs placebo component of the CSSRD trial). Most trials had ratios close to one (CSSRD stronger trials; the COBRA trial at 16 weeks).

Generally speaking, there was no clear pattern by which one could say that the discriminant validity of the ACR criteria was stronger than the discriminant validity of the EULAR definition response or vice versa.

It should be noted that a chi-squared value of at least 3.84 yields a $p < 0.05$; a chi-squared value of at least 6.64 yields a p value < 0.01, and a chi-squared value of at least 10.83 yields $p < 0.001$. With one exception, either ACR or EULAR criteria would have yielded significant results for these trials, and further, in most cases, a value more stringent than $p < 0.05$ would have been reached.

European Group

Trials. There were only two trials meeting all above men-

Table 2. Discriminant validity of ACR and EULAR improvement criteria. US trials.

Trial	N Stronger/N Control	Percentage Reaching ACR Improvement		Mantel-Haenszel Chi-Squared		Chi-Squared Ratio
		Stronger, %	Control, %	ACR	EULAR	
CSSRD stronger	155/119	40	8	36.8	31.4	1.2
CSSRD AUR	56/38	23	8	3.7	11.5	0.3
COBRA 16 wks	76/79	74	32	25.6	22.8	1.1
COBRA 28 wks	76/79	70	49	8.5	15.0	0.6
MTX vs AUR	119/118	65	29	30.5	16.6	1.8
MTX + Cyclo vs MTX	75/73	44	8	18.1	4.9	3.7
TNFR vs Placebo	44/44	75	14	33.2	18.9	1.8

AUR: auranofin; MTX: methotrexate; Cyclo: cyclosporine; TNFR: tumor necrosis factor receptor.

tioned requirements available for further analysis. Trial 1: A 24-week randomized, double blind, placebo controlled multicenter trial. Patients (n = 472) with early active RA received either placebo or 30 mg, 75 mg, or 150 mg recombinant human interleukin 1 receptor antagonist (rhIL-1ra) by daily subcutaneous injections[17,18]. Trial 2: A 52-week randomized, double blind, placebo controlled trial. Patients (n = 105) with early active RA received either sulfasalazine (2000-3000 mg/day), methotrexate (7.5–15 mg/wk), or the combination of sulfasalazine and methotrexate[19]. Table 3 summarizes for both trials the measurements used for the specified variables.

Response classification. Regardless of treatment group, the ACR criteria classified 159 patients of trial 1 (34%) as responders, and 305 patients (66%) as non-responders (Table 4). The EULAR criteria classified 28 patients of trial 1 (7%) as good responders, 188 (41%) as moderate responders, and 248 (53%) as non-responders. ACR and EULAR had conflicting results in 3% of the patients: 13 patients who were classified as ACR responder but not as EULAR responder had no significant change in DAS and endpoint DAS > 3.7, while of the ACR components, in general the Health Assessment Questionnaire (HAQ) and/or the ESR change was less than 20%. One patient who was classified as a EULAR responder but not as ACR responder did not change in number of swollen joints and HAQ score, which were low at baseline (swollen joint count = 9, HAQ = 0.00). Only 1% of the patients (n = 6) had conflicting results (ACR responders, EULAR non-responders) when using the ACR28 and EULAR28 criteria. In most of these patients the ESR improved less than 20% and the endpoint DAS28 was > 5.1. Between EULAR and EULAR28 criteria there were no real conflicting results, and the percentages of patients in the different classes remained almost the same, although there were 66 patients (14%) shifting between the classes, mainly because of the different joint counts used. The percentage of responders with ACR28 criteria (32%) is comparable with the ACR criteria (34%), with 45 patients (10%) changing classes.

In trial 2 the number of ACR responders is 78 (76%), and the number of EULAR good, moderate, and non-responders is, respectively, 41 (40%), 45 (44%), and 18 (17%) (Table 4). Conflicting results between ACR and EULAR were present in 3 patients (3%). The number of responders with ACR28 and EULAR28 criteria compared with the original criteria comprising extensive joint counts was somewhat lower (ACR28: 70%, EULAR28: 36% good and 46% moderate).

Treatment groups. According to both ACR and EULAR criteria, the treatment response in trial 1 was not very impres-

Table 3. Methods of assessment. European trials.

	Trial 1	Trial 2
Joint counts: tender (swollen)	68 (66)	53 (44)
Patient/investigator global*	0–4 Likert scale	1–5 Likert scale
Pain	100 mm VAS	100 mm VAS
General health	—	100 mm VAS
ESR	mm/h	mm/h
HAQ	Several versions	Dutch
Patient/investigator overall**	1–7 Likert scale	1–5 Likert scale
Radiographs	Larsen	Modified Sharp

VAS: visual analog scale; *patient's and investigator's assessment of global disease activity; **patient's and investigator's assessment of improvement.

Table 4. Response classification (N). European trials.

		Trial 1							
		EULAR			EULAR 26			ACR	
		G	M	N	G	M	N	G	N
EU28	G	21	4	0					
	M	7	152	23					
	N	0	32	225					
ACR	G	27	118	13	25	119	14		
	N	1	70	233	0	63	241		
AC28	G	26	109	11	24	116	6	131	17
	N	1	78	236	0	66	249	28	288
		Trial 2							
EU28	G	36	1	0					
	M	5	39	3					
	N	0	5	15					
ACR	G	39	36	2	36	36	5		
	N	1	8	16	1	9	15		
AC28	G	38	31	2	36	33	2	71	1
	N	2	13	16	1	12	18	7	24

G: good; M: moderate; N: non-responder.

sive, with the majority of patients classified as non-responder. Also there was no clear dose-response relation between the 30, 75, and 150 mg recombinant human interleukin 1 receptor antagonist (rhIL-1ra) groups (Table 5). However, there was a small difference in treatment response between the groups, as expressed by the p values of the Mantel-Haenszel chi-squared statistics for the ACR criteria, p = 0.02 (ACR28 p = 0.04). Using the EULAR criteria the p value was somewhat higher, p = 0.07 (EULAR28 p = 0.02).

In trial 2 the majority of patients were classified as responder with both criteria, but there were no significant differences between the 3 treatment groups (ACR p = 0.77, ACR28 p = 0.50, EULAR p = 0.73, EULAR28 p = 0.15) (Table 5).

Patient's and investigator's overall improvement.

Overall improvement was assessed by both patient and investigator on a scale from 1 to 7 (1 = marked improvement, 7 = markedly worse) in trial 1, and on a scale from 1 to 5 (1 = very much improved, 5 = very much deteriorated) in trial 2. In both trials treatment response as assessed with ACR and EULAR criteria was linearly associated with the overall assessments of both patients and investigators (p = 0.0001 for all 4 response criteria). At the extremes, there were no EULAR good and 1 EULAR moderate responders

Table 5. Response per treatment group (N). European trials.

	Trial 1									
	EULAR			EULAR28			ACR		ACR28	
	G	M	N	G	M	N	G	N	G	N
0 mg	4	44	70	3	39	76	31	88	29	90
30	10	43	64	8	45	64	44	75	41	78
75	10	41	64	9	42	64	36	79	34	80
150	4	60	50	5	56	53	49	64	44	69
	p = 0.07			p = 0.02			p = 0.02		p = 0.04	
	Trial 2									
SSZ	13	13	8	11	12	11	25	9	22	12
MTX	13	17	4	11	19	4	25	9	23	11
S+M	15	15	6	15	16	5	28	7	27	8
	p = 0.73			p = 0.15			p = 0.77		p = 0.50	

SSZ: sulfasalazine; MTX: methotrexate; S+M: combined.

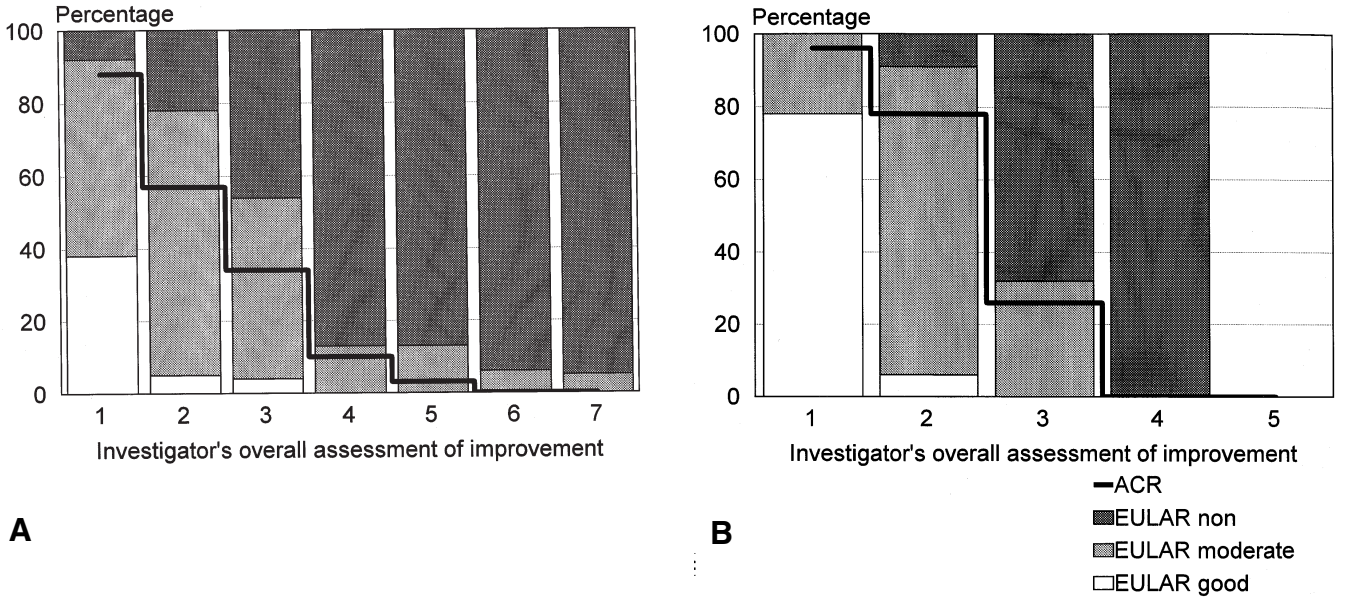

Figure 1. ACR and EULAR improvement classification compared with the investigator overall assessment of improvement. The bars indicate the percentage of EULAR response (good, moderate, non), and the line indicates the percentage of ACR responders, per category of overall improvement. A. European trial 1, *x* axis: 1 = marked improvement, 2 = moderate improvement, 3 = slight improvement, 4 = no change, 5 = slightly worse, 6 = moderately worse, 7 = markedly worse. B. European trial 2, *x* axis: 1 = very much improved, 2 = moderately improved, 3 = no change, 4 = moderately deteriorated, 5 = much deteriorated.

in class 7, i.e. "markedly worse," and 4 non-responders in class 1, i.e. "marked improvement" (Figure 1a). Similarly, at the extremes there were 0 ACR responders in the worst class and 6 non-responders in the best class. In trial 2 there were no cases with score 5 (Figure 1b). Zero EULAR good and moderate responders were in class 4, and 0 non-responders in class 1. Similarly, class 4 comprised no ACR responders, and class 1 comprised 2 non-responders.

Radiographic progression. The interquartile range of radiographic progression per response class is visualized in Figure 2. The response criteria including 28 joint counts gave results comparable to the results using the original response criteria. Disregarding treatment groups, in trial 1 there was no significant association between radiographic progression and treatment response. The median radiographic progression was 3 for almost all response groups (EULAR p = 0.57, EULAR28 p = 0.35, ACR p = 0.40, ACR28 p = 0.76). Trial 2 did show a significant association, with more progression in patients without a good treatment response (EULAR p = 0.0001, EULAR28 p = 0.0001, ACR p = 0.03, ACR28 p = 0.01).

DISCUSSION

This transatlantic study shows that ACR and EULAR definitions of response in RA have equivalent discriminant validity. The combined analysis was based on 9 well done clinical trials, and covered a range of response and differences in response between treatment groups. Both criteria sets performed similarly in differentiating active or experimental treatment from placebo or control treatment. In addition, the European analysis indicated comparable construct and criterion validity.

Although the number of published RA clinical trials is substantial, it was difficult to find trials in which both the ACR and the EULAR improvement criteria could be calculated, and in which radiographs (European analyses) were taken and scored. Only recently has the complete core set of disease activity measures been included in trials, and even within these trials there is a lot of discrepancy in the measurement techniques used. Joint counts differ with respect to the number of joints included, and the way they are examined, i.e., one joint at a time, or a complete joint unit at a time. Most trials did not include both the original Ritchie Articular Index and the full joint count necessary to calculate the 2 improvement criteria. In future the standard assessment of the 28 joint count will make it possible to calculate both ACR28 and EULAR28 improvement. The availability of radiographic scores appeared to be the main limitation to include trials in the European study.

In the European setting of 2 trials there was only 3% real discrepancy between ACR and EULAR improvement despite readily apparent differences in the definition of response. The main difference is determined by the number of response classes per criteria. The EULAR moderate group comprises more ACR good than non-responders. On an increasing level of response, patients will generally first classify as EULAR moderate responders, then as ACR responder, and finally as EULAR good responders. Response criteria with extensive and with 28 joint counts yielded similar percentages of patients per class, although there were some shifts between the classes.

Both ACR and EULAR criteria showed a high association with patient and investigator overall impression of

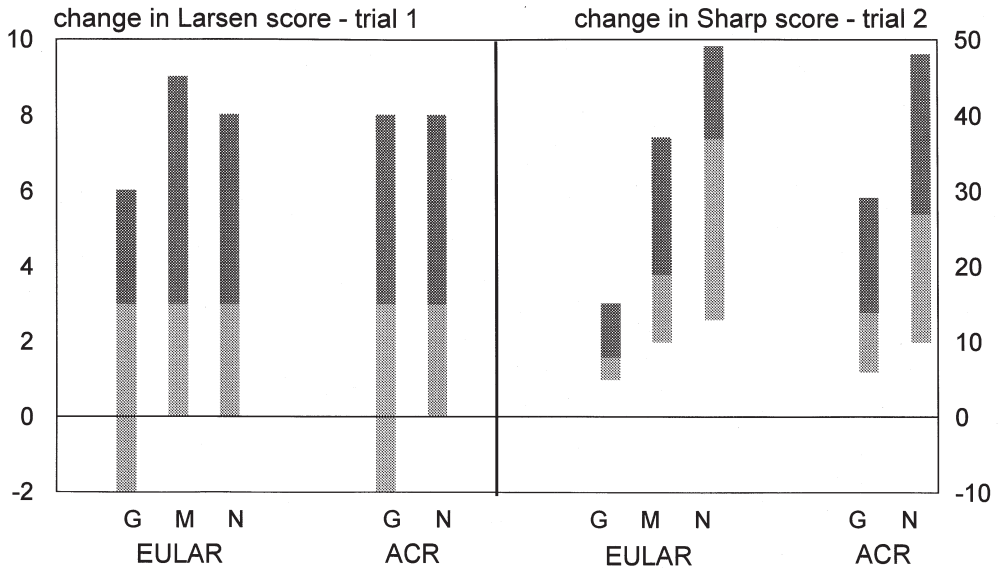

Figure 2. ACR and EULAR improvement classification compared with the progression of joint damage per European trial. Bars indicate the interquartile range of radiographic progression per response class. The median is the demarcation between the two shades of grey of the bars. The left *y* axis indicates the progression in Larsen score after 24 weeks for European trial 1, and the right *y* axis indicates the progression in modified Sharp score after 52 weeks for European trial 2.

improvement in the European trials. There were some discrepancies, especially where ACR and EULAR "moderate" appeared to overestimate the response detected by patients and investigators.

In European trial 1 there was no association between response and radiographic progression. This might be explained by the short treatment/followup period of the study, 24 weeks, but perhaps the agent tested, rhIL-1ra, was responsible for a dissociation between clinical response and radiographic progression of the disease. In European trial 2 there was a clear association between response and radiographic progression for both criteria. The p values suggest a stronger association between EULAR response and radiographic progression, but this might be a result of the higher number of response classes in comparison with the ACR response criteria.

In conclusion, there is a high level of agreement between ACR and EULAR improvement classification, and their validity is equivalent. The discriminating potential of the criteria between treatment groups is comparable.

It is recommended for future clinical trials to assess all components of both criteria, but to choose in advance which criteria will be used as primary, and which as secondary endpoint measure. Further, it will be necessary to define inclusion criteria based on the components of the primary endpoint measure to standardize the baseline disease activity of the trial population, and to make sure that response can be assessed.

REFERENCES

1. Boers M, Tugwell P, Felson DT, et al. WHO and ILAR core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol 1994; 21 Suppl 41:86-9.
2. Felson DT, Anderson JJ, Boers M, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. Arthritis Rheum 1993;36:729-40.
3. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
4. van Gestel AM, Prevoo MLL, van't Hof MA, et al. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Arthritis Rheum 1996;39:34-40.
5. van der Heijde DMFM, van't Hof MA, van Riel PLCM, van de Putte LBA. Development of a disease activity score based on judgment in clinical practice by rheumatologists. J Rheumatol 1993;20:579-81.
6. Williams HJ, Willkens RF, Samuelson CO Jr, et al. Comparison of low-dose oral pulse methotrexate and placebo in the treatment of rheumatoid arthritis. Arthritis Rheum 1985;28:721-30.
7. Williams HJ, Ward JR, Reading JC, et al. Low-dose D-penicillamine therapy in rheumatoid arthritis. Arthritis Rheum 1983;26:581-92.
8. Ward JR, Williams HJ, Egger MJ, et al. Comparison of auranofin, gold sodium thiomalate, and placebo in the treatment of rheumatoid arthritis. Arthritis Rheum 1983;26:1303-15.
9. Boers M, Verhoeven AC, Markusse HM, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine alone in early rheumatoid arthritis. Lancet 1997;330:309-18.
10. Weinblatt ME, Kaplan H, Germain BF, et al. Low-dose methotrexate compared with auranofin in adult rheumatoid arthritis. Arthritis Rheum 1990;33:330-8.
11. Tugwell P, Pincus T, Yocum D, et al. Combination therapy with cyclosporine and methotrexate in severe rheumatoid arthritis. N Engl J Med 1995;333:137-41.
12. Moreland LW, Baumgartner SW, Schiff MH, et al. Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. N Engl J Med 1997;337:141-7.
13. Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis: results of two metaanalyses. Arthritis Rheum 1990;33:1449-61.
14. Fuchs HA, Pincus T. Reduced joint counts in controlled clinical trials in rheumatoid arthritis. Arthritis Rheum 1994; 37:470-5.
15. American College of Rheumatology Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Reduced joint counts in rheumatoid arthritis clinical trials. Arthritis Rheum 1994;37:463-4.
16. van Gestel AM, Haagsma CJ, van Riel PLCM. Validation of rheumatoid arthritis improvement criteria including simplified joint counts. Arthritis Rheum 1998;41:1845-50.
17. Bresnihan B, Lookabaugh J, Witt K, Musikic P. Treatment with recombinant human interleukin-1 receptor antagonist in rheumatoid arthritis: results of a randomized double-blind, placebo-controlled multicenter trial [abstract]. Arthritis Rheum 1996;39 Suppl:S73.
18. Watt I, Cobby M, Amgen rhIL-1ra Clinical Research Product Team. Recombinant human IL-1 receptor antagonist (rhIL-1ra) reduces the rate of joint erosion in rheumatoid arthritis [abstract]. Arthritis Rheum 1996;39 Suppl:S123.
19. Haagsma CJ, van Riel PL, de Jong AJ, van de Putte LB. Combination of sulphasalazine and methotrexate versus the single components in early rheumatoid arthritis: a randomized, controlled, double-blind, 52 week clinical trial. Br J Rheumatol 1997; 36:1082-8.