

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/23689>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

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

A. M. VAN GESTEL, M. L. L. PREVOO, M. A. VAN 'T HOF, M. H. VAN RIJSWIJK,
L. B. A. VAN DE PUTTE, and P. L. C. M. VAN RIEL

Objective. To validate the European League Against Rheumatism (EULAR), the American College of Rheumatology (ACR), and the World Health Organization (WHO)/International League Against Rheumatism (ILAR) response criteria for rheumatoid arthritis (RA).

Methods. EULAR response criteria were developed combining change from baseline and level of disease activity attained during followup. In a trial comparing hydroxychloroquine and sulfasalazine, we studied construct (radiographic progression), criterion (functional capacity), and discriminant validity.

Results. EULAR response criteria had good construct, criterion, and discriminant validity. ACR and WHO/ILAR criteria showed only good criterion validity.

Conclusion. EULAR response criteria showed better construct and discriminant validity than did the ACR and the WHO/ILAR response criteria for RA.

Rheumatoid arthritis (RA) is a chronic systemic disease, with polyarthritis as its main feature. Chronic inflammation of the joints often leads to joint damage

and functional impairment. Because the pathogenesis of RA is still unknown, antirheumatic therapies are focused on nonspecific suppression of disease activity. Patients with RA have various manifestations of the disease, and therefore, disease activity cannot be expressed by a single parameter of inflammation. An index of disease activity should combine measurements representing several aspects of the disease (1). The Disease Activity Score (DAS) is such a validated index. It combines the Ritchie articular index (2), a count of swollen joints, the erythrocyte sedimentation rate (ESR), and an assessment of the patient's general health (3).

In general, the efficacy of a treatment is determined by comparing group means of changes in disease activity variables. However, a significant difference between groups does not indicate the actual number of patients who responded to treatment. Therefore, in addition to disease activity, the response of individual patients to antirheumatic therapy is an important measurement in clinical trials. Response criteria should include the relevant change in disease activity since the start of treatment, and the level of disease activity attained during followup (4).

Recently, 2 sets of response criteria were proposed: the American College of Rheumatology (ACR) preliminary criteria for improvement (5), and the World Health Organization (WHO)/International League Against Rheumatism (ILAR) criteria for decreased inflammatory synovitis (6). The components of these criteria were selected using a judgmental approach rather than a statistical approach (7). On behalf of the EULAR Standing Committee for International Clin-

Supported by the Het Nationaal Reumafonds (the Dutch League against Rheumatism).

A. M. van Gestel, MSc, M. L. L. Prevo, MSc, L. B. A. van de Putte, MD, PhD, P. L. C. M. van Riel, MD, PhD: University Hospital Nijmegen, Nijmegen, The Netherlands; M. A. van 't Hof, PhD: Nijmegen University, Nijmegen, The Netherlands; M. H. van Rijswijk, MD, PhD: University Hospital Groningen, Groningen, The Netherlands.

Address reprint requests to A. M. van Gestel, MSc, University Hospital Nijmegen, Department of Rheumatology, P. O. Box 9101, 6500 HB Nijmegen, The Netherlands.

Submitted for publication March 14, 1995; accepted in revised form August 4, 1995.

Table 1. Response criteria defined by the WHO/ILAR and the ACR*

| WHO/ILAR response criteria | ACR response criteria |
|--|--|
| <ol style="list-style-type: none"> 1. >20% improvement in swollen joint count 2. >20% improvement in tender joint count, or ≥ 5 if the count is between 16 and 20 3. >20% improvement in at least 2 of the following 3 measures: <ol style="list-style-type: none"> A. Patient's or physician's assessment of global disease activity B. Pain C. ESR | <ol style="list-style-type: none"> 1. $\geq 20\%$ improvement in swollen joint count 2. $\geq 20\%$ improvement in tender joint count 3. $\geq 20\%$ improvement in at least 3 of the following 5 measures: <ol style="list-style-type: none"> A. Patient's global assessment of disease activity B. Physician's global assessment of disease activity C. Patient's assessment of pain D. Acute-phase reactant E. Disability |

* WHO = World Health Organization; ILAR = International League Against Rheumatism; ACR = American College of Rheumatology; ESR = erythrocyte sedimentation rate.

cal Studies including Therapeutic Trials, it was decided to develop response criteria based on a combination of the judgmental and statistical approaches. The validity of the newly developed EULAR criteria and the ACR and WHO/ILAR response criteria was also studied. Our findings are presented herein.

PATIENTS AND METHODS

Development of response criteria. *Patients and measurements.* Response criteria were developed using a cohort of patients with recent-onset (<1 year) definite or classic RA (8) who were attending the outpatient clinic at the University Hospital Nijmegen. Between 1985 and 1994, 227 patients were included in the study.

Slow-acting antirheumatic drugs (SAARDs) were prescribed when treatment with nonsteroidal antiinflammatory drugs (NSAIDs) alone was not sufficiently effective. Sulfasalazine (SSZ), hydroxychloroquine (HCQ), or aurano-fin (AUR) were regarded as first-option SAARDs. In case of treatment failure, aurothioglucose or methotrexate could be prescribed. A third option was treatment with D-penicillamine or azathioprine. Oral prednisone (10 mg) and intra-articular injections of steroid were allowed as adjuvant therapy. Rheumatologists decided about all changes in therapy.

Patients were seen every 3 months by specially trained research nurses. The nurses collected clinical and laboratory data, including the Ritchie articular index (RAI), number of swollen joints, erythrocyte sedimentation rate (ESR, in mm/hour), and general health status (by 100-mm visual analog scale). On the basis of these measurements, we calculated the Disease Activity Score (DAS) (3):

$$\text{DAS} = 0.54(\sqrt{\text{RAI}}) + 0.065(\text{SwJts}) + 0.33(\ln \text{ESR}) + 0.0072(\text{GH})$$

where SwJts = the number of swollen joints and GH = general health status.

The measurement error of the DAS was estimated, using interperiod correlation matrix analysis, in patients with ≥ 3 years of followup ($n = 78$). With this method, the assumption was made that the correlation between 2 DAS measurements declines as the interval in between increases. The intercept of the regression line (y axis: interperiod Pearson correlations between DAS measurements; x axis: intermediate time interval) was used for estimating the

measurement-remeasurement correlation r_0 (correlation between DAS measurements with intermediate time interval = 0) (9). From this r_0 , the measurement error (e) can be calculated:

$$e^2/\text{SD}^2 = (1/r_0) - 1$$

where SD = the standard deviation of the DAS.

Definition of response. Response was defined as both: (a) change in disease activity from baseline and (b) the level of disease activity attained during followup. The following criteria for change and attained disease activity were used. (a) For good response, a statistically significant decrease in disease activity from baseline (i.e., more than 2 times the measurement error [$2e$] [95% confidence interval = $\text{DAS} \pm 2e$]), was necessary. (b) For good response, the DAS level attained must correspond to low levels of disease activity.

Periods of low disease activity were defined as either the time during which the rheumatologist recommended that SAARD treatment be stopped because of remission, or periods of at least 1 year during which SAARD treatment was not started or the existing SAARD treatment was not changed. A high level of disease activity was defined as the time at which the rheumatologist decided that the patient should start SAARD treatment, or that the SAARD being taken should be changed (after a washout period of >1 month for SSZ or methotrexate, and >2 months for the remaining SAARDs). Medical records were checked to correct for reasons other than high or low levels of disease activity that could bias the rheumatologist's decision regarding treatment (noncompliance, refusal of therapy, etc.). In the analyses, ≤ 2 periods of high disease activity and ≤ 2 periods of low disease activity per patient, with a time period between high and low activity of >1 year, were randomly chosen.

Validation of response criteria. *Patients and measurements.* In a 48-week double-blind trial comparing SSZ and HCQ in 60 patients with recent-onset RA (10), the EULAR response criteria and 2 other newly developed response criteria, the WHO/ILAR and the ACR criteria (Table 1), were validated. Twenty-five patients in the SSZ-HCQ trial were also included in the open (development) study. However, the overlap existed only during the first year of the open study (<10% of the data); thereafter, no overlap was present.

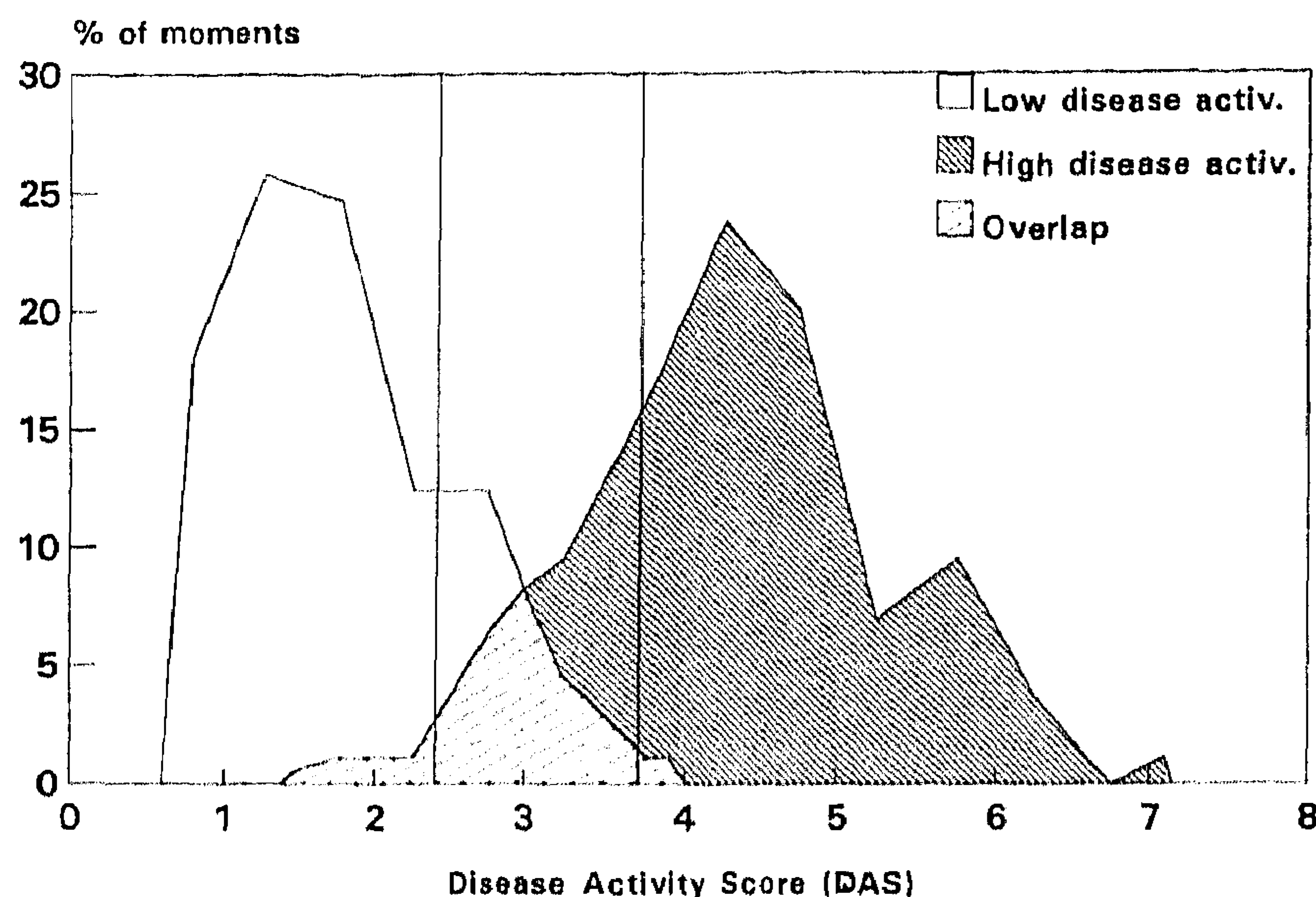


Figure 1. Distribution of Disease Activity Scores at moments of low ($n = 89$) or high ($n = 189$) disease activity, according to treatment decisions made by rheumatologists in the cohort study of patients with rheumatoid arthritis. The vertical lines divide the DAS into low (≤ 2.4), moderate (> 2.4 and ≤ 3.7), and high (> 3.7) levels of disease activity (activ.).

Disease activity variables were measured every 12 weeks. Radiographs of the hands and feet were taken at baseline and at 48 weeks. Films were scored using a modification of the method of Sharp et al (11). At weeks 0, 12, 24, 36, and 48, a Dutch equivalent of the Stanford Health Assessment Questionnaire (HAQ) was used to measure functional capacity in the last 41 consecutive patients of the trial (12).

Response was retrospectively assessed every 12 weeks. Because the physician's global assessment of disease activity was not included in the original trial, we modified the ACR criteria, such that patients had to fulfill 3 of the 4 remaining measures as well as have improvement in the tender and swollen joint counts to be considered a responder.

Validation procedures (7). Criterion validity tests the accuracy of the criteria. Because no gold standard for response was available, we compared the criteria with each other and with the true clinical status, as determined by the functional capacity (from the HAQ). To correct for baseline values, the relative change in the HAQ score was evaluated at weeks 12, 24, 36, and 48.

Construct validity is an aspect of validity that investigates the association of the criteria with the expected results (the outcome). Radiographic progression represents an observable biologic end point resulting from inflammation and enzymatic degradation of cartilage and subchondral bone (13). Therefore the association between response (every 12 weeks) and radiographic progression (total number of new erosions and joint space narrowing at week 48 compared to baseline) was analyzed.

Discriminant validity refers to the ability of the criteria to detect clinically important differences. Therefore, the proportion of responders in both treatment categories (HCQ and SSZ) was compared every 12 weeks.

Statistical analysis. Missing disease activity data were supplied by interpolation when measurements were available within 2 weeks from the missing moment. Tests were performed using patient moments of response: response at weeks 12, 24, 36, and 48 for each of the 60 patients. To equalize the number of patient moments available for the analyses of each set of response criteria, we used 186 moments of the original 240 (4×60), excluding moments with missing EULAR, modified ACR, or WHO/ILAR responses. Radiographic progression was transformed (square root) to "normality." The association between response (week 12, 24, 36, 48) and radiographic progression (week 48) was tested with analysis of variance. The association between response (week 12, 24, 36, 48) and relative change in HAQ score (week 12, 24, 36, 48) was analyzed with Kruskal-Wallis tests. Differences in response in both treatment groups were tested with Wilcoxon rank sum tests (EULAR response criteria: good response = 1, moderate response = 2, no response = 3; WHO/ILAR and modified ACR response criteria: good response = 1, insufficient response = 0).

RESULTS

Development of the EULAR criteria. The estimated Pearson's correlation coefficient for remeasurements was 0.80, and the standard deviation of the DAS was 1.17, leading to a measurement error of 0.6. For good response, a change from baseline that exceeded 1.2 (or 2×0.6) DAS points was necessary.

Disease Activity Scores during moments of high and low disease activity, according to the rheumatologists, were calculated. In 142 patients, 189 moments of high disease activity were defined. The DAS ranged from 1.38 to 7.15, with a mean of 4.32. Eighty-nine moments of low disease activity in 56 patients were defined. The DAS ranged from 0.59 to 4.01, with a mean of 1.77.

To minimize the overlap between high and low disease activity ($\leq 5\%$), 2 limits were chosen: one at the 75th percentile of low disease activity (DAS = 2.4), and the other at the 25th percentile of high disease activity (DAS = 3.7) (Figure 1). With these

Table 2. Means (ranges) of the components of the DAS for low, moderate, and high levels of disease activity*

| | Ritchie index | No. of swollen joints | ESR (mm/hour) | General health |
|-----------------------------|---------------|-----------------------|---------------|----------------|
| DAS ≤ 2.4 | 1 (0-5) | 4 (0-18) | 13 (1-54) | 16 (0-53) |
| $2.4 < \text{DAS} \leq 3.7$ | 5 (0-15) | 10 (0-25) | 29 (3-99) | 33 (0-76) |
| DAS > 3.7 | 14 (1-37) | 18 (7-35) | 45 (3-130) | 51 (1-99) |

* General health was quantified on a 100-mm visual analog scale. DAS = Disease Activity Score; ESR = erythrocyte sedimentation rate.

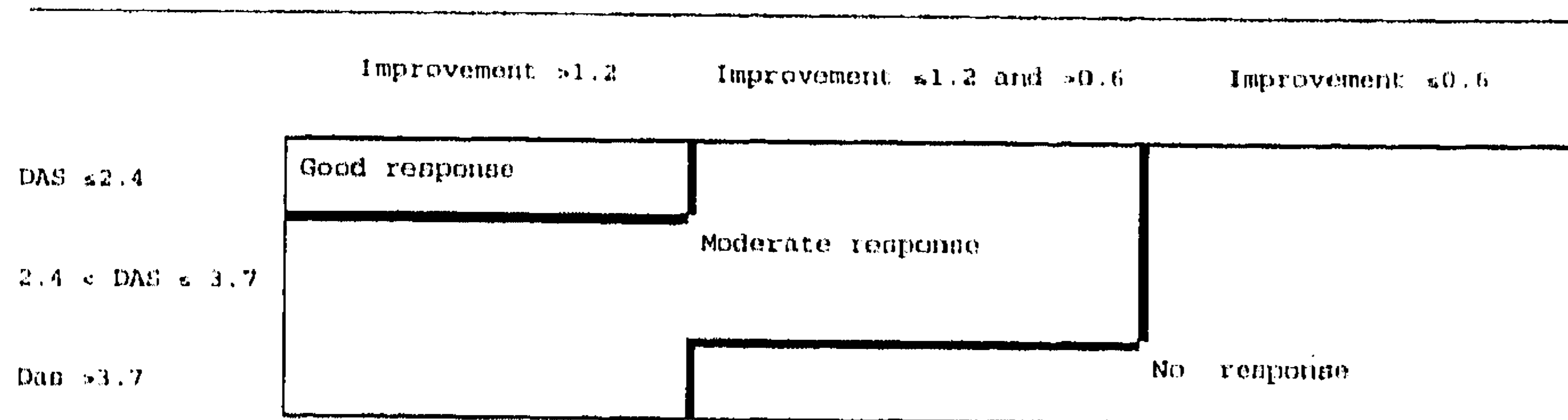


Figure 2. European League Against Rheumatism (EULAR) response criteria based on the Disease Activity Score (DAS) in patients with rheumatoid arthritis. Improvement in the DAS was compared with baseline; categories to the left represent the level of disease activity attained during followup.

limits, the DAS was divided into 3 categories: ≤ 2.4 (low disease activity), >2.4 and ≤ 3.7 (moderate disease activity), and >3.7 (high disease activity). The means and ranges of the components of the DAS are indicated in Table 2. Good response was defined as >1.2 improvement in the DAS from baseline, and a DAS attained during followup of ≤ 2.4 . Nonresponders were patients with an improvement of ≤ 0.6 (1e) or patients with an improvement of >0.6 but ≤ 1.2 , and a DAS attained during followup of >3.7 . The remaining patients would be classified as moderate responders (Figure 2).

Validation of the criteria. Response criteria were validated in a trial comparing SSZ and HCQ. No significant differences in patient characteristics were present at baseline (Table 3).

Criterion validity. Figure 3 shows responders and nonresponders according to the modified ACR criteria, plotted within the 3 classes of EULAR response. The WHO/ILAR criteria showed similar results, with only 3 patients classified differently than

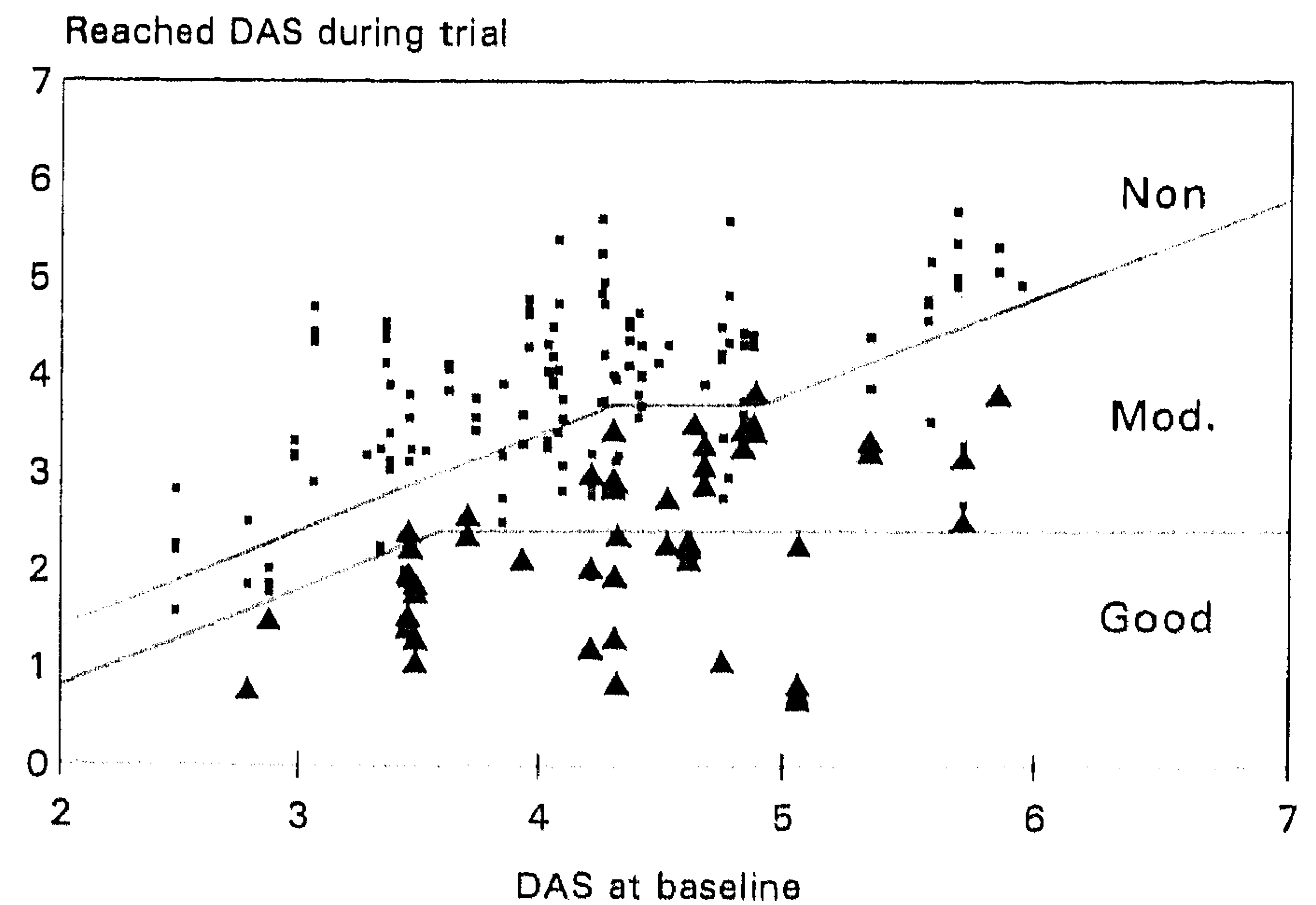


Figure 3. Distribution of the modified American College of Rheumatology (ACR) responses over the 3 European League Against Rheumatism (EULAR) response categories (56 patients, 186 moments). Triangles indicate good response according to the modified ACR criteria, squares indicate no response according to the modified ACR criteria. Responders according to the EULAR criteria are shown in the upper (no response [Non]), middle (moderate response [Mod.]), and lower (good response [Good]) areas of the figure. The results with the World Health Organization/International League Against Rheumatism (WHO/ILAR) criteria were comparable with those of the ACR criteria.

with the modified ACR criteria. Responders as defined by the modified ACR (and WHO/ILAR) criteria were classified with the EULAR response criteria as good 58% (56%), moderate 40% (42%), and no response 2% (2%). Nonresponders as defined by the modified ACR (and WHO/ILAR) response criteria were classified with the EULAR response criteria as good 1% (1%), moderate 26% (25%), and no response 72% (74%).

Table 3. Patient characteristics at baseline (treatment trial)*

| | Hydroxychloroquine (n = 30) | Sulfasalazine (n = 30) |
|---|--------------------------------|---------------------------|
| Sex, % female | 57 | 68 |
| Age, mean (range) years | 53.1 (22-72) | 53.5 (22-75) |
| Disease duration, median (range) months | 8.0 (3-120) | 8.5 (3-165) |
| IgM-RF, % >10 IU | 96 | 91 |
| DAS, mean (range) | 4.32 (3.07-5.87) | 4.37 (2.48-6.77) |
| RAI, median (range) | 14.0 (3-34) | 11.5 (1-34) |
| No. of swollen joints, mean (range) | 14.7 (5-26) | 14.4 (5-32) |
| ESR, median (range) mm/hour | 39.5 (5-110) | 35.5 (35-120) |
| General health, mean (range) | 28.7 (0-85) | 39.0 (0-100) |
| No. of erosions, median (range) | 1 (0-18) | 0 (0-17) |
| No. with narrowing, median (range) | 2 (0-20) | 0 (0-10) |

* There were no significant differences in characteristics between the 2 groups. IgM-RF = IgM rheumatoid factor; DAS = Disease Activity Score; RAI = Ritchie articular index; ESR = erythrocyte sedimentation rate.

At weeks 12, 24, 36, and 48, treatment response was compared with the relative change in functional capacity (the HAQ) ($n = 102$). All 3 response criteria showed a good correlation with functional capacity ($P = 0.0001$). Patients with good response had significantly more improvement in functional capacity than did those with moderate response and those with no response (Figure 4).

Discriminant validity. The discriminating capacity of the response criteria was studied comparing the numbers of responders who were taking SSZ and HCQ ($n = 186$). With the EULAR response criteria, patients in the SSZ-treated group showed a significantly better response (23% good, 36% moderate, and 41% no response; $n = 86$) compared with the HCQ-treated group (12% good, 25% moderate, 63% no response; $n = 100$) ($P = 0.002$). With the modified ACR criteria (SSZ 34% good, HCQ 23% good) and the WHO/ILAR criteria (SSZ 35% good, HCQ 25% good), no significant differences could be demonstrated ($P = 0.11$ and $P = 0.14$, respectively).

Construct validity. Response according to the EULAR criteria was significantly associated with radiographic progression of disease. Patients showing no response had significantly more progression in joint destruction ($n = 169$; $P = 0.0001$). With the modified ACR response criteria, a significant association was also found ($P = 0.03$). The association with the WHO/

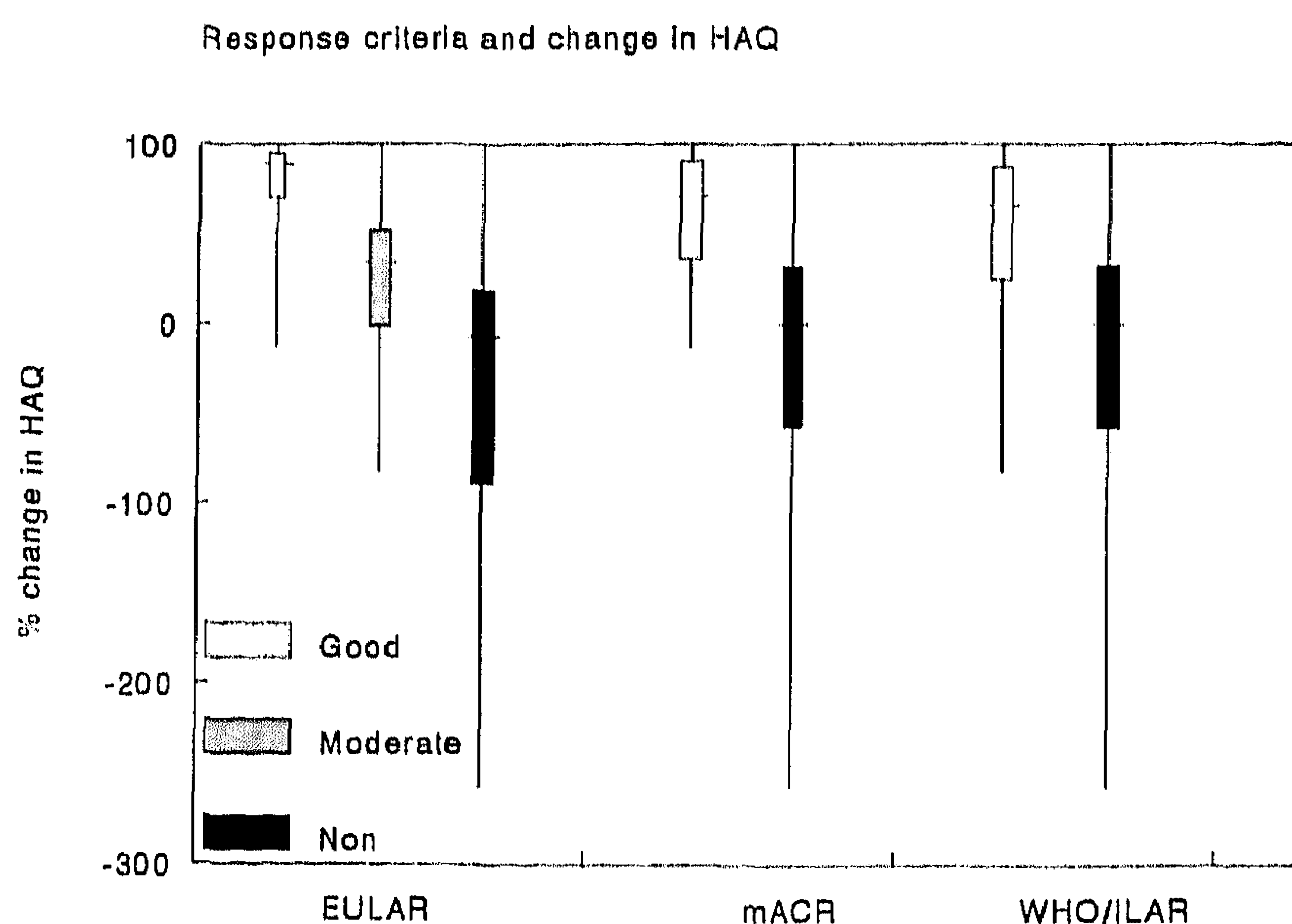


Figure 4. Relative change in scores on the Health Assessment Questionnaire (HAQ) for categories of responders, as defined by the EULAR, modified ACR (mACR), and WHO/ILAR response criteria ($n = 102$). Boxes show the interquartile range; horizontal lines show the median; vertical lines show the range. See Figure 3 for other definitions.

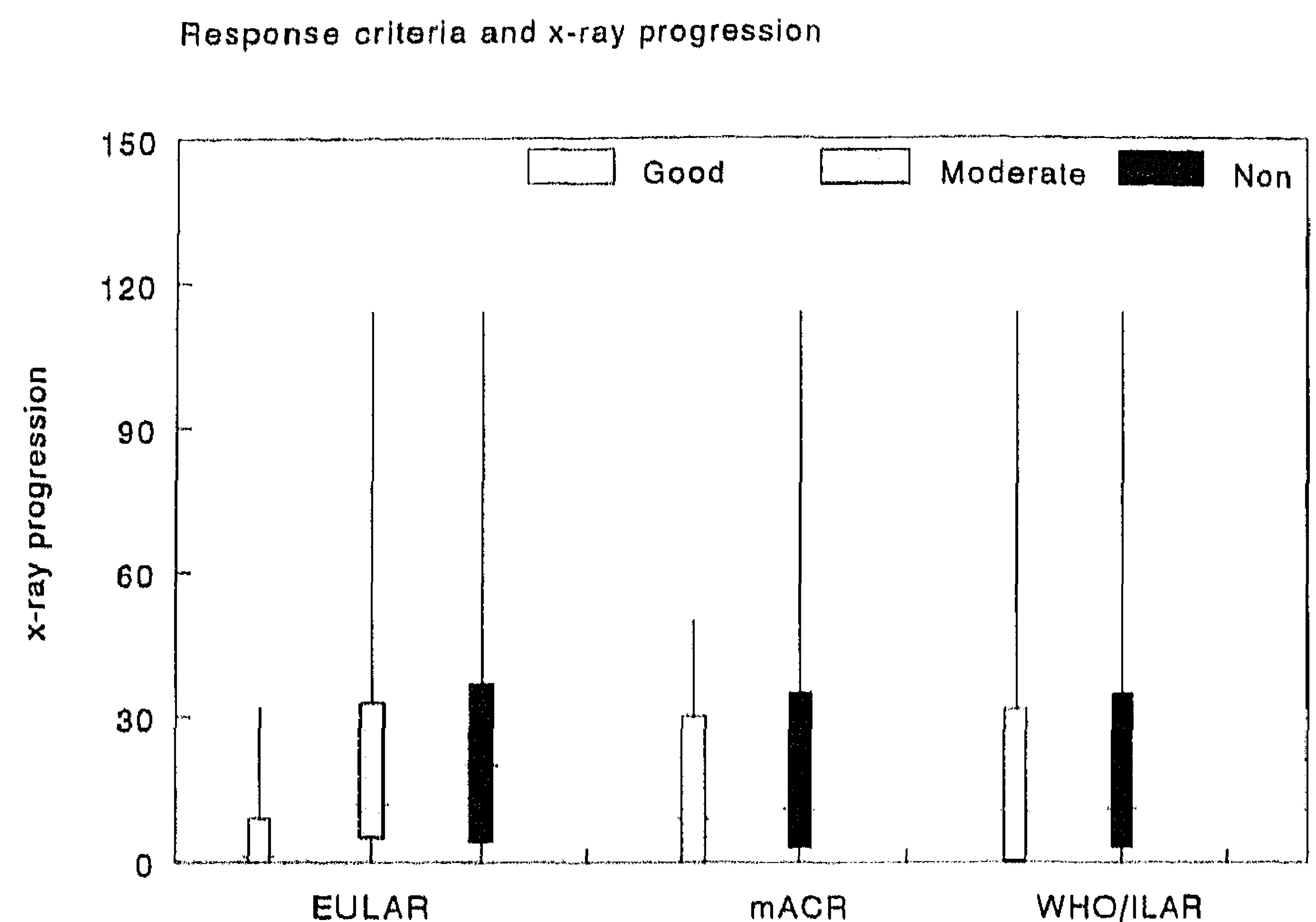


Figure 5. Radiographic progression of disease over 48 weeks, for categories of responders as defined by the EULAR, modified ACR (mACR), and WHO/ILAR response criteria ($n = 169$). Boxes show the interquartile range; horizontal lines show the median; vertical lines show the range. See Figure 3 for other definitions.

ILAR response criteria was not significant ($P = 0.07$) (Figure 5).

DISCUSSION

Although it has been stated that individual treatment response is an important outcome-of-interest in clinical trials, it is seldom assessed (1). Only mean changes in disease activity variables within and between groups are tested. This is partly due to the absence of validated, commonly accepted response criteria. In the past, several groups have made attempts to develop response measures. These measures, comprising a set of often arbitrarily selected variables (14–16) and arbitrarily defined class limits (15,16), have sometimes not been validated (16). The selection of components of the ACR and WHO/ILAR response criteria is based on consensus, although there is great disagreement between the purported and actual judgment of disease activity by rheumatologists (17). These and other response definitions are based solely on changes from baseline (5,6,14). Whether this change is significant or relevant (resulting in low disease activity) is not included in the definition. Also, they assume a certain percentage of improvement (20%) to be an equally relevant change in all variables (18). The ACR response criteria are based on the capacity of the criteria to discriminate between active treatment and placebo. An aspect which might have an impact on these criteria is the bias which may occur by

including patients being treated with "stable" doses of corticosteroids (19).

The Disease Activity Score is a composite index of disease activity based on rheumatologists' opinions of disease activity. Rheumatologists were unaware of the ESR, and therefore of the DAS, when treatment decisions were made. The DAS showed good correlational, criterion, and construct validity (3). Components of the DAS (articular index, ESR), as well as the DAS itself (20), were found to be sensitive to change (21,22).

The EULAR response criteria as presented here are based on change from baseline (or, measurement error, which is a statistical approach) and level of DAS reached during followup (low, moderate, or high, based on treatment decisions, which is a judgmental approach) (4). They show good construct, criterion, and discriminant validity: they are associated with progression of joint damage and relative change in functional capacity, and they discriminate between therapies. The median change in the HAQ score for both moderate and good responders was of the same magnitude as has been declared to be a clinically important difference by Redelmeier and Lorig (23). The WHO/ILAR criteria show a marked association with functional capacity, but no association with radiographic progression, and they are not sensitive to discriminating between SSZ and HCQ treatments. Apart from the component of functional capacity, the ACR response criteria are almost identical to the WHO/ILAR criteria. Because the physician's global assessment was not measured in the original trial, only a global impression of the performance of the ACR criteria could be given in the present study, however. The association with radiographic progression was less strong than that found with the EULAR criteria. The association with functional capacity is not surprising, since disability is part of this response definition. The modified ACR criteria were not able to discriminate between therapies.

We compared patient responses according to the EULAR criteria with the responses defined by the WHO/ILAR and the modified ACR criteria. With the latter 2 sets of criteria, about 25% of the nonresponders and about 40% of the responders were classified as having a moderate response by the EULAR criteria. We studied whether the difference in results (construct/discriminant validity) between the 3 criteria could be explained by the number of categories. Therefore, the performance of 2-group EULAR response criteria (good/moderate versus no response)

was studied. The results were equal to those with our original 3-group response criteria, and were better than those with the modified ACR and the WHO/ILAR criteria. Redefining the modified ACR criteria into 3 groups (good = >50%, moderate = >20%, no response = \leq 20% improvement from baseline) showed similar associations as found with the original 2 categories. EULAR response criteria based solely on change from baseline (good >1.2, moderate >0.6, no response \leq 0.6) showed results that were comparable with those of the modified ACR and the WHO/ILAR criteria and were worse than those of the EULAR response criteria that included level of disease activity attained during followup (data not shown).

Tugwell and Bombardier describe 2 additional types of validity apart from criterion, discriminant, and construct validity, namely content validity and face validity (7). Content validity, or comprehensiveness of response criteria, means that all important components of disease activity are included in the criteria. The Disease Activity Score was based on treatment decisions made by rheumatologists. The disease activity variables best explaining these decisions were selected using a statistical approach (3). The inclusion of other disease activity variables in the DAS had no additional discriminating capacity between high and low disease activity (24). The ACR and WHO/ILAR components were selected using a judgmental approach. A disadvantage of such a method is that some components may be duplicative; for example, there is a high correlation between the results of patient's assessment of global disease activity and patient's assessment of pain, and no additional information is gained by measuring both. The face validity of the response criteria, the credibility, depends on the willingness to accept the method and on the extent to which the results are interpretable. The assessment of response with all 3 criteria is complicated. The interpretation of the results is less difficult because there are only 2 or 3 categories of response.

In agreement with the WHO/ILAR response criteria (10), the EULAR response criteria assume a minimum level of baseline disease activity, that is, patients must have the potential to improve significantly (>1.2) to become good responders. A DAS < 1.0 indicates the absence of disease activity; therefore, the baseline value of the DAS has to be >2.2 (1.0 + 1.2) to define response. This could also serve as an inclusion term for clinical trials.

The DAS is a continuous variable that was developed to describe the disease process. EULAR

response criteria give an interpretation of the (change in) DAS for individual patients. We conclude that EULAR response criteria have construct validity, criterion validity, and discriminant validity. However, the performance of these criteria in comparison with others has to be studied further in future clinical trials comparing different treatment regimens in different patient populations. Also the performance of response criteria based on the modified DAS (including 28-joint counts) will be evaluated.

ACKNOWLEDGMENTS

We thank Drs. Gerold Stucki, Daniel Furst, and Josef Smolen for their critical comments and suggestions, and Dr. Désirée van der Heijde for the radiographic data for the validation study.

REFERENCES

1. Felson DT, Anderson JJ, Meenan RF: Time for changes in the design, analysis, and reporting of rheumatoid arthritis clinical trials. *Arthritis Rheum* 33:140-149, 1990
2. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakas TG, Grieveson P, Buchanan WW: Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. *Q J Med* 37:393-406, 1968
3. Van der Heijde DMFM, van 't Hof MA, van Riel PLCM, Theunisse LAM, Lubberts EW, van Leeuwen MA, van Rijswijk MH, van de Putte LBA: Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 49:916-920, 1990
4. Van Riel PLCM, van de Putte LBA: DC-ART: what proportion of response constitutes a positive response? *J Rheumatol Suppl* 21:54-56, 1994
5. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, Katz LM, Lightfoot R Jr, Paulus H, Strand V, Tubwell P, Weinblatt M, Williams HJ, Wolfe F, Kieszak S: American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 38:727-735, 1995
6. Furst DE: Considerations on measuring decreased inflammatory synovitis. *ILAR Bull* 2:17-21, 1994
7. Tugwell P, Bombardier C: A methodologic framework for developing and selecting endpoints in clinical trials. *J Rheumatol* 9:758-762, 1982
8. Ropes MW, Bennett GA, Cobb S, Jacox R, Jessar RA: 1958 revision of diagnostic criteria for rheumatoid arthritis. *Bull Rheum Dis* 9:175-176, 1958
9. Van 't Hof MA, Kowalski CJ: In, *A Mixed Longitudinal Interdisciplinary Study of Growth and Development*. Edited by B Prahl Anderson, CH Kowalski, P Heyendaal. New York, Academic Press, 1979
10. Nuver-Zwart HH, van Riel PLCM, van de Putte LBA, Gribnau FWJ: A double blind comparative study of sulphasalazine and hydroxychloroquine in rheumatoid arthritis. *Ann Rheum Dis* 48:389-395, 1989
11. Van der Heijde DMFM, van Riel PLCM, Nuver-Zwart HH, Gribnau FWJ, van de Putte LBA: Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* 1:1036-1038, 1989
12. Van der Heijde DMFM, van Riel PLCM, van de Putte LBA: Sensitivity of a Dutch Health Assessment Questionnaire in a trial comparing hydroxychloroquine and sulphasalazine. *Scand J Rheumatol* 19:407-412, 1990
13. Fries JF, Bloch DA, Sharp JT, McShane DJ, Spitz P, Bluhm GB, Forrester D, Genant H, Gofton P, Richman S, Weissman B, Wolfe F: Assessment of radiologic progression in rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum* 29:1-9, 1986
14. Paulus HE, Egger MJ, Ward JR, Williams HJ, and the Cooperative Systematic Studies of Rheumatic Disease 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
15. Davis MJ, Dawes PT: A disease activity index: its use in clinical trials and disease assessment in patients with rheumatoid arthritis. *Semin Arthritis Rheum* 23 (Suppl 1):50-56, 1993
16. Scott DL: A simple index to assess disease activity in rheumatoid arthritis. *J Rheumatol* 20:582-584, 1993
17. Kirwan JR: Outcome measures in rheumatoid arthritis clinical trials: assessing improvement. *J Rheumatol* 20:543-545, 1993
18. Dixon JS: Response criteria for slow acting antirheumatic drugs (letter). *Ann Rheum Dis* 49:819-820, 1990
19. Boers M: Low-dose prednisone in rheumatoid arthritis patients: placebo treatment? (Letter) *Arthritis Rheum* 34:501-502, 1991
20. Fuchs HA: The use of the disease activity score in the analysis of clinical trials in rheumatoid arthritis. *J Rheumatol* 20:1863-1866, 1993
21. Anderson JJ: Sensitivity to change of rheumatoid arthritis clinical trial outcome measures. *J Rheumatol* 20:535-537, 1993
22. Dixon JS, Hayes S, Constable PDL, Bird HA: What are the "best" measurements for monitoring patients during short-term second-line therapy? *Br J Rheumatol* 27:37-43, 1988
23. Redelmeier DA, Lorig K: Assessing the clinical importance of symptomatic improvements. *Arch Intern Med* 153:1337-1342, 1993
24. Prevo MLL, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LBA, van Riel PLCM: Modified disease activity scores that include twenty-eight-joint counts: development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 38:44-48, 1995