RHEUMATOID ARTHRITIS RESPONSE CRITERIA AND PATIENT-REPORTED IMPROVEMENT IN ARTHRITIS ACTIVITY: Is an ACR20 Response Meaningful to Patients?

Michael M. Ward, MD, MPH,
Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Lori C. Guthrie, BSN, and
Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Maria I. Alba, MD
Intramural Research Program, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health

Abstract

Objective—We examined the association of the American College of Rheumatology (ACR) response criteria (ACR20, ACR50, ACR70) and the European League Against Rheumatism (EULAR) response criteria with patient-reported improvement in rheumatoid arthritis (RA) activity.

Methods—We studied 250 patients with active RA before and after escalation of anti-rheumatic treatment in a prospective longitudinal study. We asked patients to report if they subjectively judged if they had experienced important improvement with treatment, and compared the proportion that reported improvement with the proportion that met ACR20, ACR50, ACR70, and EULAR responses.

Results—Improvement in overall arthritis status was reported by 167 patients (66.8%), while 107 patients (42.8%) had an ACR20 response, 52 (20.8%) had an ACR50 response, 24 (9.6%) had an ACR70 response, and 136 (54.4%) had a EULAR moderate/good response. An ACR20 response had a sensitivity of 0.57 and a specificity of 0.85 for clinically important improvement as judged by patients. Sensitivities of the ACR50, ACR70, and EULAR moderate/good responses were 0.30, 0.14, and 0.68, respectively, while their specificities were 0.97, 0.99, and 0.73, respectively. The ACR hybrid score with the highest sensitivity and specificity for important improvement was 19.99%.

Conclusions—Among patients with active RA, ACR20 responses are highly specific measures of improvement as judged by patients, but exclude a substantial proportion of patients who

consider themselves improved. Response criteria are associated with, but not equivalent to, patient-perceived improvement.

The American College of Rheumatology (ACR) response criteria and the European League Against Rheumatism (EULAR) response criteria for rheumatoid arthritis (RA) have been widely adopted as measures of medication efficacy in clinical trials (1,2). Both sets of criteria define hierarchical categories of response. The ACR criteria require meeting a threshold of 20% change in RA activity measures, with 50%, and 70% improvement also commonly reported. The ACR hybrid measure was developed to capture gradations of response between and beyond these specific thresholds, although many studies continue to rely on ACR20, ACR50, and ACR70 responses (3). EULAR response criteria incorporate both the degree of absolute change in the Disease Activity Score (DAS) and the level of disease activity attained with treatment. The ACR20 response has been the preferred endpoint for clinical trials because it is the response shown to discriminate optimally between active treatment and placebo while identifying few placebo-treated patients as improved (4).

RA response criteria were derived and selected based on their discriminative ability, but because they serve as measures of drug efficacy, they have also been interpreted to represent clinically important improvements. Indeed, the first step in the development of the ACR response criteria was the generation of candidate criteria based on ratings of important changes by rheumatologists (1). However, some have questioned whether an ACR20 response represents a clinically important improvement, given that 20% represents a relatively modest change in RA activity measures, and have instead favored the more exacting 50% or 70% responses (5,6). It is not known how well these responses correspond to changes in RA activity that are meaningful to patients. Knowing whether patients who achieve an ACR20 response consider themselves to have had an important improvement in their RA, or whether only larger responses are judged as meaningful by most patients, would help answer this question. We investigated the sensitivity and specificity of ACR and EULAR response criteria for clinically important improvement as judged by patients.

METHODS

Participants

We enrolled patients with active RA who were receiving ongoing care in our community-based and National Institutes of Health-based clinics in a prospective longitudinal study, with the goal to assess clinically important changes in RA activity measures (7). We enrolled adults age 18 or older who fulfilled the 1987 ACR classification criteria for RA (8), who had active RA based on physician clinical judgment and six or more tender joints, and who also had initiation or escalation of disease-modifying anti-rheumatic medications or biologics, or initiation of prednisone to treat active RA, at the baseline visit. Treatment decisions were made by the patient's rheumatologist and not determined by the study. The study was approved by the institutional review board, and all patients provided written informed consent.

Study procedures

Participants completed a baseline visit and a follow-up visit four months later (1 month later for those treated with prednisone). At both visits, we performed joint counts (66 swollen, 68 tender) and scored the physician global assessment (on a 0 – 100 visual analog scale), administered questionnaires to obtain the patient global assessment (0 – 100 visual analog scale), pain score (0 – 100 visual analog scale), and Health Assessment Questionnaire Disability Index, and tested the erythrocyte sedimentation rate (ESR).

At the follow-up visit, participants were asked to judge if their arthritis overall had improved, worsened, or was unchanged since the baseline visit, and to rate the importance of any change (7-category scale ranging from "hardly important at all" to "extremely important") (9). The item specifically asked "Since the start of the study, overall my arthritis has: improved, stayed the same, gotten worse." This item has been used extensively in research on clinically important changes (10). Construct validity was demonstrated by significant associations between responses on this question and measured changes in RA activity (7).

Statistical analysis

From the changes in the clinical measures, we computed whether each patient had an ACR20, ACR50, or ACR70 response, and computed their ACR hybrid score. The ACR hybrid is an officially endorsed modification of the ACR response criteria that merges ACR20/50/70 responses with mean responses on the ACR core set measures for those patients who do not achieve an ACR20, ACR50 or ACR70 threshold (3). If a patient does not meet an ACR20 response, their ACR hybrid score is the mean of their core set responses if the mean is less than 20%, and is set to 19.99% if the mean of their core set responses is 20% or higher. Similar rules apply to thresholds surrounding ACR50 and ACR70 responses. We also computed EULAR moderate/good and EULAR good responses for each patient based on changes in their DAS28.

We categorized patients as improved or not based on their judgment of change in overall arthritis status, and computed the sensitivity, specificity, positive predictive value, and negative predictive value for each response criterion in its association with patient-reported improvement. For the ACR hybrid, we used a receiver operating characteristic (ROC) curve to determine the threshold of change in the ACR hybrid that had the highest sensitivity and specificity for patient-reported improvement, as determined by the point on the curve that was closest to the upper left corner [0,1] of the ROC plot. The [0,1] corner represents a threshold with both a sensitivity and specificity of 1.0. As alternatives, we also examined the change in ACR hybrid associated with the Youden index, which is the point on the ROC curve that maximizes the difference between the proportion of true positives and false positives, and the change in ACR hybrid score that had a specificity of 0.80 for patient-reported improvement. We used SAS programs version 9.3 (SAS Institute, Cary, NC) for analysis.

RESULTS

Of 262 patients enrolled, 250 completed the study and were included in the analysis. Ten patients were either lost to follow-up or withdrew, one died prior to the second visit, and one had an incomplete follow-up assessment. Most patients were middle-aged women with seropositive erosive RA (Table 1). Patients had active RA, with a mean of 16 swollen joints (of 66) and 25 tender joints (of 68), and a mean DAS28 of 6.16. At the baseline visit, 104 patients (41.6%) had escalation of their current DMARD, 90 patients (36%) started a new DMARD or biologic (methotrexate in 60 patients and tumor necrosis factor-alpha inhibitor in 20 patients), and 56 patients (22.4%) were treated with prednisone.

At the follow-up visit, 107 patients (42.8%) had an ACR20 response (Table 2). Among the components of the ACR response, 20% responses were most frequent for the physician global assessment (73.1%) and least frequent for the HAQ (53.6%). ACR50 and ACR70 responses were present in 20.8% and 9.6% respectively, while EULAR moderate/good responses were present in 54.4% and EULAR good responses were present in 14.8%. At the follow-up visit, 167 patients (66.8%) reported that their global arthritis status had improved, and 68% rated this improvement as either very important or extremely important.

An ACR20 response was a highly specific marker of patient-reported improvement (specificity 0.85), but was only modestly sensitive, indicating that many patients who judged themselves as improved did not have an ACR20 response (Table 2). Ninety-six of 107 patients with an ACR20 response reported improvement, compared to 72 of 143 patients who did not have an ACR20 response. The positive predictive value of an ACR20 response for patient-reported improvement was 0.89. ACR50 and ACR70 responses were even more specific but much less sensitive than the ACR20 response. EULAR moderate/good responses were somewhat more sensitive but less specific than ACR20 responses. One hundred fourteen of 136 patients with a EULAR moderate/good response reported improvement, compared to 53 of 114 patients who did not have a EULAR moderate/good response. EULAR good responses were both less sensitive and less specific than ACR50 responses.

Of the 72 patients who judged themselves to be improved but did not have an ACR20 response, 11 patients (15.3%) had 20% improvement in both the tender joint count and swollen joint count but did not have 20% improvement in three of the five remaining RA core set measures, 34 patients (47.2%) had 20% improvement in three of the five non-joint count measures but did not meet the joint count requirement, and 27 patients (37.5%) did not meet either the joint count requirement or the remaining measures requirement.

We also examined associations of the ACR20 and patient-reported improvement in patient subgroups defined by age, sex, and ethnicity (Supplemental table 1). Specificities and positive predictive values were similar among these subgroups, but sensitivity tended to be lower for older patients, men, and for blacks and whites (relative to Hispanics).

The median ACR hybrid response was 19.99 (25th, 75th percentile 12.0, 49.99). Values were higher among patients who judged themselves improved compared to those not improved (median 28.9 versus 14.6; p < .0001). We used ROC curve analysis to determine the ACR

hybrid response that best discriminated patients who reported improvement versus no improvement (ROC curve area = 0.78). An ACR hybrid change of 19.99 was identified as the optimal threshold for discrimination, with a specificity of 0.81 and sensitivity of 0.62 (Figure 1). A change of 19.99 was also identified as the optimal threshold based on the maximum Youden index and when a specificity of 0.80 was used as the criterion for discrimination.

DISCUSSION

Because response criteria were developed as measures of treatment efficacy, they have been interpreted as thresholds of important improvement (11). Although some have questioned the importance of an ACR20 response, our results indicate that these responses are very specific markers of improvement in RA activity as judged by patients. Additionally, ACR20 responses had a very high positive predictive value for patient-reported improvement. ACR50 and ACR70 responses were even more specific than ACR20 responses. Results were similar for EULAR responses, even though these criteria use a continuous measure rather than a combination of dichotomous measures and include the state of RA activity achieved. This similarity suggests that the format of the response criteria did not influence the results.

Despite its high specificity, an ACR20 response had only modest sensitivity for patient-reported improvement, indicating that many patients who judged themselves to be improved did not have an ACR20 response. When the components of the ACR response criteria were analyzed, failure to have a 20% improvement in both the tender and swollen joint counts was the main reason that patients with subjective improvement did not meet the ACR20 response criterion. Failure to have 20% improvement in the remaining RA activity measures, which include pain, patient and physician global assessments, and HAQ, were less often the limiting factor. This may not be surprising, as pain and functioning are likely more readily appreciated by patients, and limited changes in the joint count measures may be less relevant to patients' judgments provided their pain and functioning improves.

The ACR hybrid was developed to provide additional grading of responses beyond the 20%, 50%, and 70% thresholds. We tested this measure to determine if there was any degree of response that had a higher sensitivity for patient-reported improvement than the ACR20 criterion. The optimal threshold in this analysis was 19.99%, which had a sensitivity of 0.62, compared to a sensitivity of 0.57 for the ACR20. An ACR hybrid score of 19.99% results when both joint count measures fail to have a 20% improvement but the average improvement of the measures exceeds 20%. This pattern reflects the central role of the joint count measures in determining the association between the ACR response criteria and patient-reported improvement (12).

We studied patients' judgments of improvement but not physicians' judgments, based on the premise that physician interpretations of response criteria should be informed by what patients think. Studies that included a large number of physicians could compare both patient and physician judgments of improvement to the response criteria to provide another perspective. We also examined associations with improvement rated dichotomously, rather

than by degree of importance, because most patients reported that their improvement was very important or extremely important. It is important to stress that these associations apply to patients with active RA. Similar associations between RA response criteria and patient-reported improvement may not be present among patients with low or more moderate RA activity.

The high specificity of ACR and EULAR responses for patient-reported improvement aids interpretation of studies that use these criteria to test treatments. Based on our findings, achievement of an ACR20 response is considered important by the majority of patients with active RA who experience such a change. However, the limited sensitivity of these response criteria suggests a role for reporting the proportion of participants who have clinically important improvement in patient-reported measures in clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum. 1995; 38:727–35. [PubMed: 7779114]

2. Van Gestel AM, Prevoo ML, van't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. 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. 1996; 39:34–40. [PubMed: 8546736]

3. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. Arthritis Rheum. 2007; 57:193–202. [PubMed: 17330293]

4. Felson DT, Anderson JJ, Lange ML, Wells G, LaValley MP. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? Arthritis Rheum. 1998; 41:1564–70. [PubMed: 9751088]

5. Pincus T, Stein CM. ACR20: Clinical or statistical significance? Arthritis Rheum. 1999; 42:1572–6. [PubMed: 10446853]

6. Chung CP, Thompson JL, Koch GG, Amara I, Strand V, Pincus T. Are American College of Rheumatology 50% response criteria superior to 20% criteria in distinguishing active aggressive treatment in rheumatoid arthritis trials reported since 1997? A meta-analysis of discriminant capacities. Ann Rheum Dis. 2006; 65:1602–7. [PubMed: 16504992]

7. Ward MM, Guthrie LC, Alba MI. Clinically important changes in individual and composite measures of rheumatoid arthritis activity. Thresholds applicable in clinical trials. Ann Rheum Dis. 2014 in press.

8. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 1988; 31:315–324. [PubMed: 3358796]

9. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989; 10:407–15. [PubMed: 2691207]

10. Revicki D, Hays RD, Cella D, Sloan J. Recommend methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008; 61:102–9. [PubMed: 18177782]

11. Ward MM. Response criteria and criteria for clinically important improvement: separate and equal? Arthritis Rheum. 2001; 44:1728–9. [PubMed: 11508421]

12. van Vollenhoven RF, Felson DT, Strand V, Weinblatt ME, Luijtens K, Keystone EC. American College of Rheumatology hybrid analysis of certolizumab pegol plus methotrexate in patients with active rheumatoid arthritis: Data from a 52-week phase III trial. Arthritis Care Res. 2011; 63:128–34.
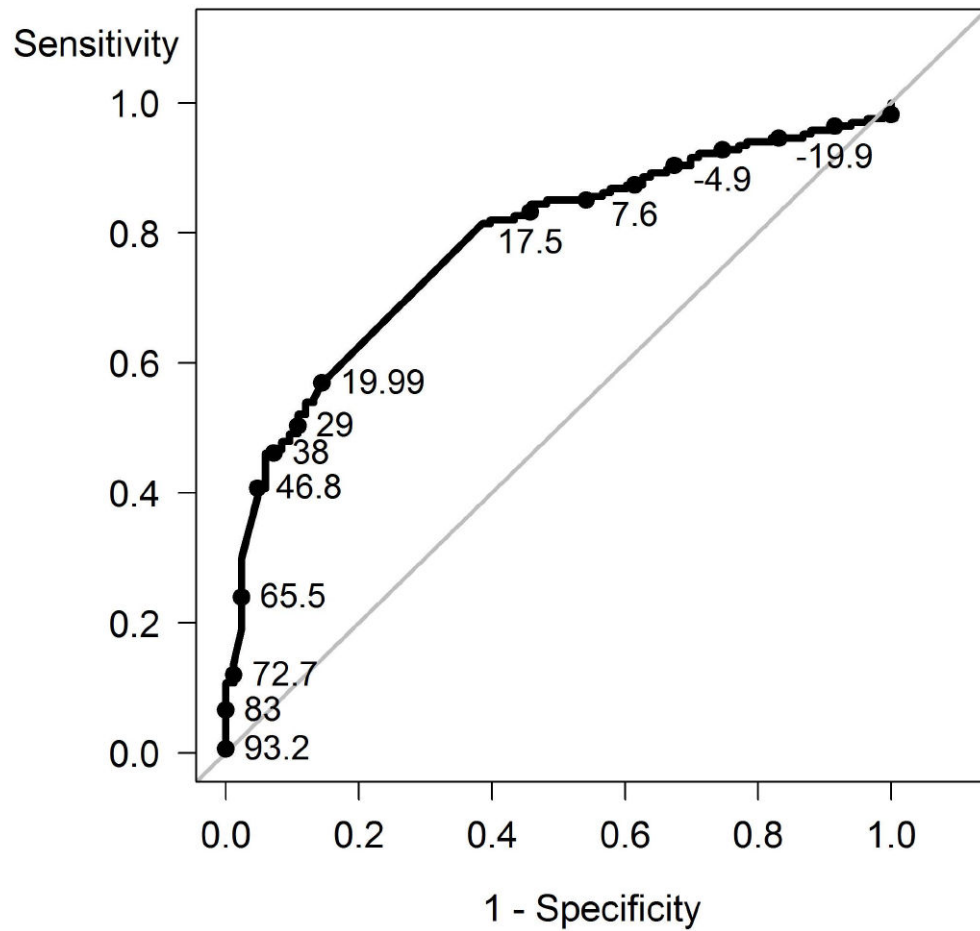
Figure 1.
Receiver operating characteristic curve for the association of the American College of Rheumatology hybrid improvement score with patient's judgment of improvement in overall arthritis status. Data labels refer to the ACR hybrid score.

Table 1

Patient characteristics at study entry (N = 250).[*]

Age, years	51.0 ± 13.7
Women	195 (78%)
White, non-Hispanic	102 (418%)
Hispanic	73 (29%)
Black	56 (22%)
Asian	17 (7%)
Multi-ethnic	2 (1%)
Duration of RA, years	9.6 ± 10.0
Seropositive	188 (75%)
Erosive	141 (64%)[†]
Swollen joint count (0 – 66)	16 ± 9
Tender joint count (0 – 68)	25 ± 15
Physician global assessment (0 – 100)	48.2 ± 17.5
Erythrocyte sedimentation rate (mm/hr)	40 ± 27
Medications at entry	
Methotrexate	87 (35%)
Hydroxychloroquine	64 (26%)
Sulfasalazine	24 (10%)
Leflunomide	16 (6%)
Prednisone	88 (35%)
Tumor necrosis factor-alpha inhibitors	28 (11%)
Other biologics	2 (1%)

[*] Plus-minus values are mean ± standard deviation. All other values are number (percentage).

[†] Of 221 patients with radiographs.

Table 2

Sensitivity, specificity, and predictive values of rheumatoid arthritis response critieria.[*]

Response Criterion	Proportion meeting criterion	Sensitivity	Specificity	Positive predictive value	Negative predictive value
ACR20	42.8	0.57 (0.49, 0.65)	0.85 (0.75, 0.92)	0.89 (0.80, 0.94)	0.49 (0.41, 0.58)
ACR50	20.8	0.30 (0.23, 0.38)	0.97 (0.90, 0.99)	0.96 (0.85, 0.99)	0.41 (0.34, 0.48)
ACR70	9.6	0.14 (0.09, 0.20)	0.99 (0.92, 0.99)	0.96 (0.76, 0.99)	0.36 (0.30, 0.43)
EULAR Moderate/Good	54.4	0.68 (0.60, 0.72)	0.73 (0.62, 0.83)	0.84 (0.76, 0.90)	0.53 (0.43, 0.63)
EULAR Good	14.8	0.15 (0.10, 0.23)	0.87 (0.77, 0.93)	0.70 (0.52, 0.84)	0.34 (0.27, 0.41)

[*] ACR = American College of Rheumatology; EULAR = European League Against Rheumatism.

Values in parentheses are 95% confidence intervals.