Origins of Discordant Responses among 3 Rheumatoid Arthritis Improvement Criteria

Michael M. Ward, Lori C. Guthrie, Maria I. Alba, and Abhijit Dasgupta

ABSTRACT. ***Objective.*** We examined agreement between the American College of Rheumatology (ACR), European League Against Rheumatism (EULAR), and Simplified Disease Activity Index (SDAI) response criteria in rheumatoid arthritis (RA) and tested whether discordant responses were associated with patients' baseline characteristics or changes in RA activity encapsulated by the different criteria.
Methods. In a prospective longitudinal study, we examined responses of 243 patients with active RA to escalation of antirheumatic treatment. We computed agreement between pairs of response criteria using κ coefficients and identified patient characteristics associated with unique responses to individual criteria.
Results. We found that 110 patients (45.3%) had an ACR 20% improvement (ACR20) response, 135 (55.5%) had a EULAR moderate/good response, and 83 (34.1%) had an SDAI50 response. Agreement was moderate to good (ACR20/EULAR κ 0.57; ACR20/SDAI50 κ 0.64; EULAR/SDAI50 κ 0.59). All who had SDAI50 response also had a EULAR response. Patient characteristics at baseline generally did not distinguish those who responded to both, 1, or neither criterion. Discordance was most often because of improvements in the erythrocyte sedimentation rate or C-reactive protein level among EULAR and SDAI50 responders, which were not as common among ACR20 responders. Based on receiver-operating characteristic curves, SDAI35 response had a better balance of sensitivity and specificity relative to ACR20 and EULAR moderate/good responses than SDAI50.
Conclusion. Discordant responses to RA improvement criteria are most often because of differences in responses of acute-phase reactants. SDAI35 response had higher sensitivity for improvement, as reflected by other response criteria, than SDAI50 response. (First Release April 1 2018; J Rheumatol 2018;45:745–52; doi:10.3899/jrheum.170788)

Key Indexing Terms:
RHEUMATOID ARTHRITIS   RESPONSE CRITERIA
SIMPLIFIED DISEASE ACTIVITY INDEX

Response criteria are key measures in the evaluation of medication efficacy because they provide benchmarks for patient improvement. In rheumatoid arthritis (RA) clinical trials, the American College of Rheumatology (ACR) and the European League Against Rheumatism (EULAR) response criteria have been widely accepted[1,2]. The ACR20 response criteria require that a threshold of 20% improvement be met in both tender and swollen joint counts and in 3 of 5 additional RA activity measures. The ACR criteria are at once restrictive, because of the joint count requirements, and permissive, allowing some measures to remain unchanged or

even worsen as long as 3 other measures meet the improvement threshold. EULAR response criteria are based on changes in the Disease Activity Score (DAS), and relate to both the degree of improvement in the DAS and the residual level of RA activity. Response criteria have also been proposed for the Simplified Disease Activity Index (SDAI)[3]. These criteria are based on percentage changes in the SDAI and were developed by mapping changes in the SDAI to ACR responses[3].

Given the different components, formats, and requirements of these response criteria, it is possible that patients who are classified as responders by 1 criterion may not be classified as responders by another criterion. For example, patients with large relative changes in RA activity may satisfy the ACR criteria but not the EULAR criteria if the posttreatment requirement of the latter are not met. Alternatively, patients with major changes in pain, joint tenderness, and acute-phase reactants may not meet the ACR criteria if joint swelling did not improve sufficiently. Each criterion set relies to some extent on coordinate changes among different RA activity measures. However, because the ACR criteria consider each activity measure individually, it may be affected more than EULAR and SDAI responses if

M.M. Ward, MD, MPH, Intramural Research Program, NIAMS, NIH; L.C. Guthrie, BSN, Intramural Research Program, NIAMS, NIH; M.I. Alba, MD, Intramural Research Program, NIAMS, NIH; A. Dasgupta, PhD, Intramural Research Program, NIAMS, NIH.

Address correspondence to Dr. M.M. Ward, NIAMS/NIH, Building 10 CRC, Room 4-1339, 10 Center Drive, Bethesda, Maryland 20892, USA. E-mail: wardm1@mail.nih.gov

changes are discordant among measures. Conversely, large improvements in some component measures of the DAS or SDAI can compensate for negligible changes or worsening in other components.

Discordant responses may result not only from differences in the structure of the response criteria, but also from differences in the sensitivity of the criteria to the pretreatment level of RA activity. For example, ACR responses may be more achievable for patients with lower pretreatment levels of RA activity, because small changes (for example, a decrease from 10 to 8 tender joints) might satisfy the percent improvement requirements.

In clinical trials, EULAR moderate/good responses have often been met more commonly than ACR20 responses[4,5,6,7,8,9]. Investigating discordant responses may help explain this pattern of results. Of 4 studies that examined the concordance of ACR and EULAR responses at the individual patient level, agreement as measured by κ coefficients was 0.26, 0.54, 0.61, and 0.63, suggesting substantial agreement in some studies but fair agreement in others[10,11,12,13]. Agreement between ACR and SDAI responses were similar[11,12,13,14]. Previous studies have not compared the composition of concordant and discordant groups to determine whether patients with particular characteristics account for the discrepant responses, or whether differences result solely from the way in which the response criteria are constructed. The aim of our study was to test the concordance of ACR, EULAR, and SDAI response criteria in patients with RA, and to determine whether pretreatment characteristics or the profile of responses to the component RA activity measures favored satisfaction of 1 set of criteria over another.

MATERIALS AND METHODS

Participants. We enrolled adults with active RA in a prospective observational study, with the main goal of determining clinically important changes in RA activity measures[15]. We defined active RA by the combined presence of physician judgment of active RA, ≥ 6 tender joints, and escalation of RA treatment with disease-modifying medications, prednisone, or biologics because of active synovitis at the baseline visit. The specific treatment was decided by the patient's rheumatologist and not determined by the study. Of 262 patients enrolled, 250 patients completed the study. We excluded 7 patients who were missing values for the erythrocyte sedimentation rate (ESR) at any visit, leaving 243 patients for analysis. The study was approved by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases institutional review board (03-AR-0133), and all participants provided written informed consent.

Study procedures. Participants completed a baseline visit and a followup visit either 4 months later (if treatment escalation involved a disease-modifying medication or biologic) or 1 month later (if escalation was prednisone). The timing of the followup visits differed based on the expected time of responses to these medications. At both visits, we performed joint counts (66 swollen, 68 tender) and scored the physician's global assessment [PGA; on a 0–100 visual analog scale (VAS)]. The same rheumatologist examined the patient and performed the joint counts on both visits. We tested the ESR and C-reactive protein (CRP) level, and administered written questionnaires to obtain the patient's global assessment (PtGA; 0–100 VAS, with higher scores indicating more active RA), pain score (0–100 VAS, with higher scores indicating greater pain severity), and Health Assessment Questionnaire–Disability Index (HAQ-DI; 0–3, with higher scores indicating more functional limitations).

Statistical analysis. We computed ACR, EULAR, and SDAI responses from changes in the relevant measures between the baseline and final visits. Either acute-phase reactant could be used in ACR responses. DAS28-ESR was used for EULAR responses. We examined the concordance between each pair of criteria at 2 levels of response, following previous studies[3,11,13]: ACR20, EULAR moderate/good (hereafter EULAR), and SDAI50 (i.e., 50% decrease in SDAI) as first-level responses; and ACR50, EULAR good, and SDAI70 (i.e., 70% decrease in SDAI) as second-level responses. We measured percent agreement as the percent of patients categorized as responders or nonresponders by both criteria. We measured concordance using κ coefficients. Conventional interpretation of the κ coefficient is < 0, poor agreement; 0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, near perfect agreement[16].

To assess predictors of combined responses to pairs of criteria (e.g., ACR20 and EULAR), we categorized patients in 4 mutually exclusive groups based on whether they met neither, only 1, or both response criteria in the pair. We tested whether demographic features and levels of arthritis activity at baseline differed among the 4 groups using ANOVA. This analysis indicated whether any baseline characteristics suggested a preferential subsequent response to 1 criterion versus another. To examine correlates of discordant responses, we repeated this analysis using changes in the arthritis activity measures during our study as the dependent variable. We used log-transformed values in the analysis of CRP, and converted these back to natural units for presentation. This analysis indicated whether differences in the composition of the response criteria were associated with preferential responses to 1 criterion versus another.

In a more direct assessment of the discordant groups, we computed relative effect sizes between changes among those with responses to only 1 criterion (e.g., ACR20) and those with only responses to the alternative criterion (e.g., EULAR) using Hedges's g, which is the standardized mean difference of the mean change in 1 group versus the mean change in the second group, with standardization based on the pooled SD of the 2 means[17]. This measure is analogous to Cohen's d but is more appropriate for small samples. Hedges's g provides an indication of whether a particular feature was selectively larger in either discordant subgroup. Absolute values of Hedges's g between 0.5–0.8 are considered medium effects, and those > 0.8 as large effects[18]. We considered effects that were at least medium to be clinically meaningful.

We also tested whether values other than a 50% decrease in the SDAI were more closely associated with ACR20 and EULAR moderate/good responses, using receiver-operating characteristic (ROC) curves. The percent change in SDAI that had the highest sensitivity and specificity for either an ACR20 response or a EULAR response (i.e., the point on the ROC curve closest to the 0, 1 corner of the plot) was determined.

Analyses were performed using SAS version 9.3 (SAS Institute) and R programs.

RESULTS

Patient characteristics and treatment responses. Patients were predominantly middle-aged women with a mean DAS28 of 6.1 and mean SDAI of 38.5 at baseline (Supplementary Table 1, available from the authors upon request). Treatment escalation at the baseline visit included an increased dose of disease-modifying medication in 100 patients (41.1%), initiation of a new disease-modifying medication or biologic in 89 patients (36.6%; 60 with methotrexate, 20 with tumor necrosis factor-α inhibitors, and 9 with other medications), and initiation of prednisone in 54 patients (22.2%).

Table 1. Concordance of responses to the ACR, EULAR, and SDAI improvement criteria. Values are n (%) unless otherwise specified.

Measure A	Measure B	No Response on Measure A or B	Response on Measure A Only	Response on Measure B Only	Responses on Both Measures A and B	Agreement, %	κ
First-level responses							
ACR20	EULAR	94 (38.7)	14 (5.8)	39 (16)	96 (39.5)	78.2	0.57
ACR20	SDAI50	125 (51.4)	35 (14.4)	8 (3.3)	75 (30.9)	82.3	0.64
EULAR	SDAI50	108 (44.4)	52 (21.4)	0	83 (34.2)	78.6	0.59
Second-level responses							
ACR50	EULAR Good	183 (75.3)	32 (13.2)	5 (2.1)	23 (9.5)	84.8	0.47
ACR50	SDAI70	179 (74.9)	17 (7.1)	6 (2.5)	39 (15.5)	90.4	0.70
EULAR Good	SDAI70	194 (80.3)	4 (1.7)	20 (8.4)	23 (9.6)	90.0	0.60

ACR: American College of Rheumatology; EULAR: European League Against Rheumatism; SDAI: Simplified Disease Activity Index.

At the followup visit, 45.3% had an ACR20 response, 55.5% had a EULAR response, and 34.1% had an SDAI50 response. Additionally, 22.6% of subjects had an ACR50 response, 11.1% had a EULAR good response, and 17.8% had an SDAI70 response. All patients were eligible to register a EULAR good or moderate response, based on their baseline DAS28 scores.

Concordance of responses. Of 149 patients who had either an ACR20 or EULAR response, 96 patients had both responses, 14 patients had only an ACR20 response, and 39 patients had only a EULAR response (Table 1). Agreement between ACR20 and EULAR responses was moderate, with κ of 0.57. Agreement was slightly higher between ACR20 and SDAI50 responses. In this cohort, there were no patients with an SDAI50 response who did not also have a EULAR response. Isolated EULAR responses were present in higher proportions of patients than isolated ACR20 or SDAI50 responses, suggesting that the EULAR response was somewhat more permissive. Seventy-four patients (30.5%) met all 3 response criteria.

Concordance was moderate to substantial between second-level responses (Table 1). Because of the low frequency, we did not analyze predictors of differential second-level responses.

Predictors and correlates of ACR20/EULAR discordance. Patients who met both criteria tended to be younger and have shorter durations of RA than nonresponders (Table 2). RA activity at baseline was comparable among response groups, except for slightly higher levels of pain among those who responded to both criteria. There were graded responses in all RA activity measures, with the largest responses in those who met both criteria, and intermediate responses among those who met only 1 criterion.

Those with EULAR-only responses had substantially larger improvements in both the ESR and CRP than those with ACR20-only responses, based on Hedges's g measures of relative effect size (Table 2). Mean ESR and CRP worsened in the ACR20-only responders. In addition, improvement in the patient PtGA was somewhat higher among EULAR-only responders.

To identify further the role of individual measures in the discordance, we examined the proportion of patients with 20% responses in the ACR core measures among EULAR responders (Table 3). Among EULAR responders who also met ACR20 criteria (n = 96), 99% had ≥ 20% improvement in the PGA, and 91% had 20% improvement in pain. The measures least likely satisfied were improvements in the ESR (74%) and CRP (65%). In other words, 26% of patients met ACR20 criteria despite lacking a 20% improvement in ESR, and 35% did so despite lacking a 20% improvement in CRP. Among EULAR responders who did not meet the ACR20 criteria (n = 39), the pain score, HAQ, and CRP were the measures with the lowest response frequencies (Table 3).

Eight of the 14 patients who had an ACR20 response but not a EULAR response had a decrease in DAS28 < 0.6, while 4 patients had a decrease in DAS28 > 0.6 but did not meet the final DAS28 state criterion. These 14 patients most commonly had improvement of 20% or more in the PGA (75%) and pain score (92%), and least commonly in the ESR (42%) or CRP (33%).

Predictors and correlates of ACR20/SDAI50 discordance. This comparison showed associations similar to the ACR20/EULAR comparison: shorter RA duration was associated with responses to both criteria; little difference among response groups in baseline RA activity (except for CRP), and graded responses in RA measures among those with neither, 1, or both responses (Table 4). Among discordant responders, those with SDAI50-only responses had substantially larger improvements in the tender joint count and CRP than ACR20-only responders, while ACR20-only responders had substantially larger improvements in pain scores, based on Hedges's g.

Among SDAI50 responders who also met the ACR20 criterion (n = 75), all had ≥ 20% improvement in the PGA, while 20% improvement in the PtGA, pain score, and HAQ were seen in 88%, 91%, and 88%, respectively (Table 3). Twenty-seven percent of patients met the ACR20 response criteria without achieving a 28% improvement in ESR, while 32% did so without achieving a 20% improvement in CRP. Among SDAI50 responders who did not meet the ACR20

Table 2. Characteristics at study entry and absolute changes in RA activity measures among patients by response to both the ACR20 improvement criterion and the EULAR moderate/good response criterion. Values are mean (SE) unless otherwise specified.

Variables	Neither Response	ACR20 Only	EULAR Only	Both Responses	p	Hedges's g*
N	94	14	39	96		
Age, yrs	52.0 (1.4)	53.7 (3.6)	53.3 (2.2)	47.6 (1.4)	0.05	0.037
Women, %	74.5 (4.2)	71.4 (11.0)	76.9 (6.6)	84.4 (4.2)	0.34	–0.118
Duration, yrs	11.9 (1.0)	10.6 (2.6)	10.9 (1.6)	6.5 (1.0)	0.002	–0.038
Measures at study entry						
TJC, 0–68	23.5 (1.6)	26.4 (4.0)	27.3 (2.4)	26.6 (1.5)	0.42	–0.052
SJC, 0–66	14.8 (0.9)	15.4 (2.4)	16.5 (1.4)	17.1 (0.9)	0.33	–0.140
PGA, 0–100	45.6 (1.8)	48.1 (4.7)	48.2 (2.8)	51.0 (1.8)	0.22	–0.008
PtGA, 0–100	52.4 (2.6)	50.4 (6.7)	58.2 (4.0)	58.1 (2.6)	0.33	–0.354
Pain, 0–100	59.3 (2.6)	54.2 (6.6)	54.2 (4.0)	65.8 (2.5)	0.06	0
HAQ, 0–3	1.4 (0.1)	1.3 (0.2)	1.5 (0.1)	1.4 (0.1)	0.88	–0.266
ESR	37.0 (2.9)	28.7 (7.4)	41.5 (4.5)	42.9 (2.8)	0.21	–0.513
CRP, mg/l	14.5 (3.4)	8.3 (8.8)	19.0 (5.3)	25.1 (3.4)	0.10	–0.808
Absolute changes during the study†						
TJC	0.69 (1.2)	10.1 (3.0)	10.7 (1.8)	19.8 (1.2)	< 0.0001	–0.069
SJC	1.1 (0.8)	6.1 (2.0)	5.3 (1.2)	11.4 (0.8)	< 0.0001	0.148
PGA	9.6 (1.6)	22.7 (4.2)	17.2 (2.5)	35.1 (1.6)	< 0.0001	0.276
PtGA	1.8 (2.3)	8.4 (6.0)	22.5 (3.6)	33.8 (2.3)	< 0.0001	–0.584
Pain	7.0 (2.6)	14.7 (6.8)	10.8 (4.1)	40.9 (2.6)	< 0.0001	0.207
HAQ	0.1 (0.1)	0.3 (0.1)	0.3 (0.1)	0.8 (0.1)	< 0.0001	0.187
ESR	3.5 (2.1)	–14.4 (5.4)	14.1 (3.3)	17.6 (2.1)	< 0.0001	–0.946
CRP, mg/l	2.0 (3.0)	–11.7 (7.8)	9.2 (4.7)	18.7 (3.0)	< 0.0001	–0.734

* Hedges's g is a measure of the relative effect size of changes in the ACR20 only group and the EULAR moderate/good only group. Negatively signed values indicate greater changes among ACR20-only responders, while positively signed values indicate greater changes among EULAR-only responders. Values < –0.5 or > 0.5 were considered clinically meaningful. † Positive values indicate a decrease from baseline (improvement); negative values indicate an increase from baseline (worsening). RA: rheumatoid arthritis; ACR20: American College of Rheumatology 20% improvement; EULAR: European League Against Rheumatism; TJC: tender joint count; SJC: swollen joint count; PGA: physician's global assessment; PtGA: patient's global assessment; HAQ: Health Assessment Questionnaire; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein.

Table 3. Proportion of patients with at least 20% improvement in RA activity measures among EULAR moderate/good responders by ACR20 response status, and among SDAI50 responders by ACR20 response status. Values are %.

Variables	EULAR Responders		SDAI50 Responders	
	ACR20 Responders, n = 96	ACR20 Nonresponders, n = 39	ACR20 Responders, n = 75	ACR20 Nonresponders, n = 8
TJC	100	69	100	100
SJC	100	58	100	62
PGA	99	72	100	75
PtGA	85	67	88	50
Pain	91	49	91	25
HAQ	84	46	88	50
ESR	74	55	72	50
CRP	65	42	68	50

RA: rheumatoid arthritis; EULAR: European League Against Rheumatism; ACR20: American College of Rheumatology 20% improvement; SDAI50: Simplified Disease Activity Index 50% decrease; TJC: tender joint count; SJC: swollen joint count; PGA: physician's global assessment; PtGA: patient's global assessment; HAQ: Health Assessment Questionnaire; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein.

criteria (n = 8), one-half or fewer had 20% responses in the pain score, HAQ, and acute-phase reactants (Table 3).

Among the 35 patients who had an ACR20 response but not a SDAI50 response, 23 had a decrease in SDAI of ≥ 10 points, and 5 had a decrease of more than 20 points. The ACR20 criteria were met in these patients through responses to the PGA in 88% of patients, and to the PtGA, pain score, HAQ, ESR, and CRP in 70%, 91%, 64%, 64%, and 48%, respectively.

Predictors and correlates of EULAR/SDAI50 discordance.

Table 4. Characteristics at study entry and absolute changes in RA activity measures among patients by response to both the ACR20 improvement criterion and the SDAI50 response criterion. Values are mean (SE) unless otherwise specified.

Characteristics	Neither Response	ACR20 Only	SDAI50 Only	Both Responses	p	Hedges's g*
N	125	35	8	75		
Age, yrs	52.6 (1.2)	49.3 (2.3)	48.2 (4.8)	48.0 (1.6)	0.10	0.074
Women, %	74.4 (3.7)	82.9 (7.0)	87.5 (14.5)	82.7 (4.8)	0.43	−0.127
Duration, yrs	11.8 (0.9)	8.8 (1.7)	9.4 (3.5)	6.1 (1.1)	0.002	−0.049
Measures at study entry						
TJC, 0–68	24.4 (1.4)	25.0 (2.6)	28.0 (5.3)	27.3 (1.7)	0.56	−0.189
SJC, 0–66	15.2 (0.8)	15.9 (1.5)	17.0 (3.2)	17.4 (1.0)	0.40	−0.115
PGA, 0–100	46.3 (1.6)	47.5 (3.0)	48.4 (6.2)	52.0 (2.0)	0.16	−0.071
PtGA, 0–100	54.0 (2.3)	58.4 (4.3)	55.3 (9.0)	56.6 (2.9)	0.80	0.156
Pain, 0–100	57.8 (2.2)	59.3 (4.2)	58.9 (8.8)	66.6 (2.9)	0.11	0.017
HAQ, 0–3	1.4 (0.1)	1.4 (0.1)	1.5 (0.3)	1.4 (0.1)	0.98	−0.213
ESR	38.1 (2.5)	33.0 (4.7)	41.3 (9.8)	44.9 (3.2)	0.17	−0.297
CRP, mg/l	15.2 (2.9)	10.3 (5.5)	26.5 (11.5)	28.8 (3.8)	0.02	−0.529
Absolute changes during the study†						
TJC	2.5 (1.0)	12.0 (1.9)	20.4 (3.9)	21.7 (1.3)	< 0.0001	−0.618
SJC	1.9 (0.7)	7.2 (1.2)	8.9 (2.6)	12.3 (0.8)	< 0.0001	−0.175
PGA	10.9 (1.3)	22.6 (2.5)	26.5 (5.3)	38.6 (1.7)	< 0.0001	−0.223
PtGA	7.4 (2.1)	17.9 (3.9)	14.8 (8.3)	36.4 (2.7)	< 0.0001	0.109
Pain	8.5 (2.2)	19.6 (4.2)	2.6 (8.8)	45.9 (2.9)	< 0.0001	0.554
HAQ	0.1 (0.1)	0.4 (0.1)	0.5 (0.2)	0.9 (0.1)	< 0.0001	−0.116
ESR	6.1 (1.9)	2.2 (3.6)	14.0 (7.5)	18.9 (2.4)	< 0.0001	−0.375
CRP, mg/l	3.2 (2.6)	−0.5 (5.0)	17.6 (10.3)	21.8 (3.4)	< 0.0001	−0.656

*Hedges's g is a measure of the relative effect size of changes in the ACR20 only group and the SDAI50 only group. Negatively signed values indicate greater changes among ACR20 only responders, while positively signed values indicate greater changes among SDAI50 only responders. Values < −0.5 or > 0.5 were considered clinically meaningful. † Positive values indicate a decrease from baseline (improvement); negative values indicate an increase from baseline (worsening). RA: rheumatoid arthritis; ACR20: American College of Rheumatology 20% improvement; SDAI50: Simplified Disease Activity Index 50% decrease; TJC: tender joint count; SJC: swollen joint count; PGA: physician's global assessment; PtGA: patient's global assessment; HAQ: Health Assessment Questionnaire; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein.

Because all patients with an SDAI50 response also had a EULAR response, we were unable to identify baseline characteristics or changes in RA activity measures that were unique to each criterion. Compared to nonresponders, patients who met both criteria were younger and had a shorter duration of RA, following the pattern of other responses (Table 5). These patients also had larger mean improvements in RA activity measures than those who did not meet either criterion or those who only met the EULAR criterion. These results indicate that the EULAR criterion was more permissive than SDAI50 in categorizing patients as improved. The SDAI improved > 40% in 23 of the 52 sole EULAR responders, and > 45% in 9 sole EULAR responders, indicating that a number of these patients were close to meeting the SDAI50 criterion.

Association of alternative SDAI improvements with ACR20 and EULAR responses. We tested whether improvements in the SDAI at levels other than 50% would have closer agreement with ACR20 and EULAR moderate/good responses, using ROC curves (Figure 1). In this analysis, the change in SDAI with the optimal combination of sensitivity and specificity for association with the ACR20 was 35%, with a sensitivity of 0.87 and specificity of 0.82 (compared to 0.68 and 0.94 for SDAI50). The corresponding change in SDAI for association with the EULAR moderate/good

response was 34%, which had a sensitivity of 0.87 and specificity of 0.93 (compared to 0.61 and 1.00 for SDAI50). Areas under the ROC curves were high, indicating good discrimination. The κ between SDAI35 and ACR20 responses was 0.67, and between SDAI35 and EULAR moderate/good responses was 0.79.

DISCUSSION

Our study has 3 main findings. First, the pattern of discordant responses revealed a hierarchy of stringency among the 3 RA response criteria, with the EULAR moderate/good criterion being the most permissive, SDAI50 the most strict, and ACR20 intermediate. Second, patient characteristics at baseline were largely not associated with differential responses. Third, when EULAR and SDAI50 responses were present in the absence of ACR20 responses, it was often because of greater responses in the acute-phase reactants.

More patients had a EULAR moderate/good response than an ACR20 or SDAI50 response, corroborating previous studies that reported higher frequencies of EULAR responses than ACR20 responses[4-13]. Discordance was commonly because of patients who had EULAR responses without an ACR20 or SDAI50 response. Sole SDAI50 responders were also uncommon when compared with ACR20 responses, indicating that the SDAI50 was stricter in categorizing

Table 5. Characteristics at study entry and absolute changes in rheumatoid arthritis activity measures among patients by response to both the EULAR moderate/good response criterion and the SDAI50 response criterion. Values are mean (SE) unless otherwise specified.

Characteristics	Neither Response	EULAR Only	Both Responses	p
N	108	52	83	
Age, yrs	52.2 (1.3)	51.2 (1.9)	48.0 (1.5)	0.10
Women, %	74.1 (4.0)	80.8 (5.7)	83.1 (4.5)	0.29
Duration, yrs	11.7 (1.0)	9.9 (1.4)	6.4 (1.1)	0.002
Measures at study entry				
TJC, 0–68	23.8 (1.5)	25.9 (2.1)	27.4 (1.7)	0.27
SJC, 0–66	14.9 (0.9)	16.3 (1.2)	17.4 (1.0)	0.16
PGA, 0–100	45.9 (1.7)	47.8 (2.4)	51.7 (1.9)	0.08
PtGA, 0–100	52.2 (2.4)	60.8 (3.5)	56.4 (2.8)	0.11
Pain, 0–100	58.7 (2.4)	57.0 (3.5)	65.9 (2.7)	0.07
HAQ, 0–3	1.4 (0.1)	1.5 (0.1)	1.4 (0.1)	0.84
ESR	36.0 (2.7)	39.2 (3.9)	44.6 (3.0)	0.11
CRP, mg/l	13.7 (3.2)	14.8 (4.5)	28.6 (3.6)	0.005
Absolute changes during the study*				
TJC	1.9 (1.1)	10.2 (1.5)	21.6 (1.2)	< 0.0001
SJC	1.7 (0.7)	5.8 (1.0)	12.0 (0.8)	< 0.0001
PGA	11.3 (1.5)	17.9 (2.1)	37.4 (1.7)	< 0.0001
PtGA	2.7 (2.2)	24.4 (3.1)	34.4 (2.5)	< 0.0001
Pain	8.0 (2.5)	17.0 (3.6)	41.8 (2.8)	< 0.0001
HAQ	0.1 (0.1)	0.3 (0.1)	0.8 (0.1)	< 0.0001
ESR	1.1 (2.0)	13.8 (2.9)	18.4 (2.3)	< 0.0001
CRP, mg/l	0.2 (2.8)	7.1 (4.1)	21.4 (3.2)	< 0.0001

* Positive values indicate a decrease from baseline (improvement); negative values indicate an increase from baseline (worsening). EULAR: European League Against Rheumatism; SDAI50: Simplified Disease Activity Index 50% decrease; TJC: tender joint count; SJC: swollen joint count; PGA: physician's global assessment; PtGA: patient's global assessment; HAQ: Health Assessment Questionnaire; ESR: erythrocyte sedimentation rate; CRP: C-reactive protein.
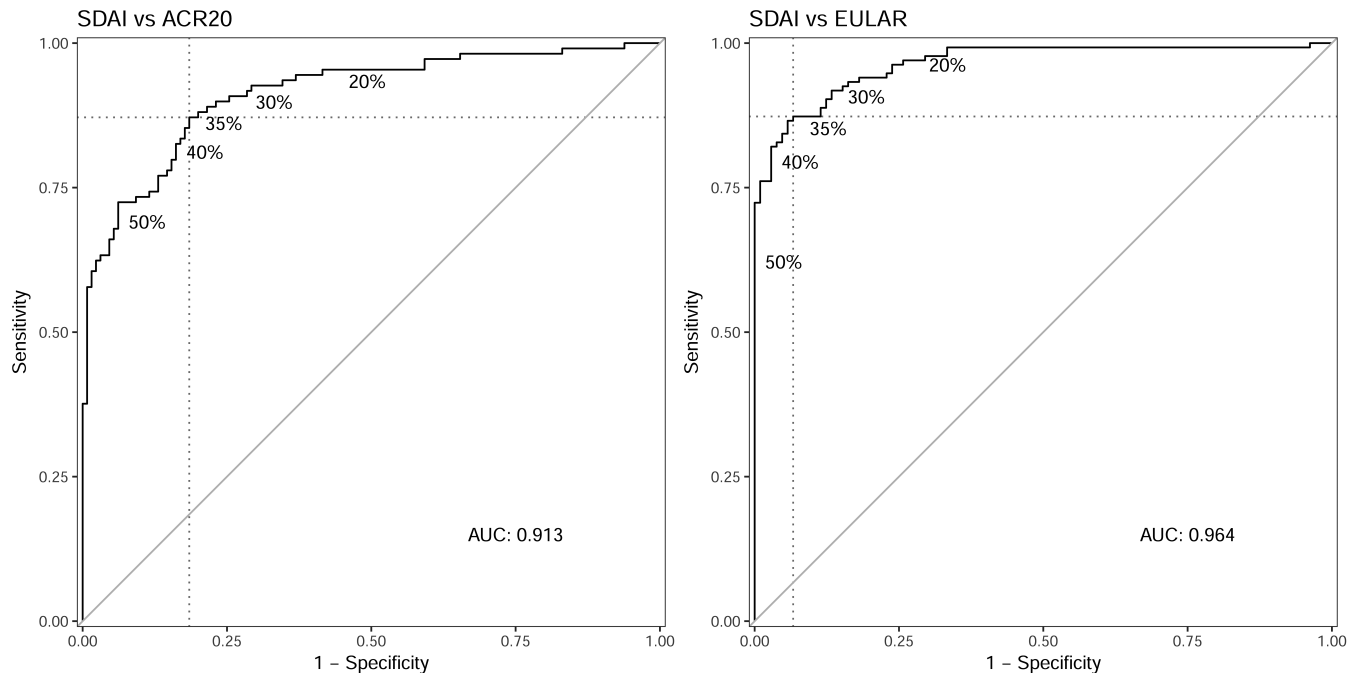

Figure 1. ROC curve of the association between different percent improvements in the SDAI using either the ACR20 or the EULAR moderate/good response, as the standard for judging response. Various cutoff points of percent improvement in the SDAI were tested for their sensitivity and specificity in association with whether the patient was also an ACR20 or EULAR responder. ROC: receiver-operating characteristics; SDAI: Simplified Disease Activity Index; ACR20: American College of Rheumatology 20% improvement response; EULAR: European League Against Rheumatism; AUC: area under the curve.

improvement than the other criteria. These results indicate that the response criteria, at least for first-level responses, can be ranked in order of permissiveness as EULAR moderate/good, ACR20, and SDAI50.

Agreement between ACR20 and EULAR moderate/good responses in our study was comparable to that in prior studies[10,11,13]. The 1 study that reported poorer agreement (κ = 0.27) did not examine patients at the start of a new treatment and included worsening as a response category, which may have altered the estimate of agreement[12]. Our results on ACR20 and SDAI50 agreement were also comparable to prior studies, which reported κ of 0.67, 0.37, and 0.54[11,12,13]. Similarly, our agreement between EULAR and SDAI50 responses matched prior studies, in which κ ranged from 0.54 to 0.71[11,12,13,14]. It is natural to compare these particular responses because they represent the minimum definition of improvement for each measure. However, this should not necessarily imply that the same patients should be identified as improved by these criteria.

Our first consideration in investigating the origin of discordant responses was whether particular types of patients were more likely to register responses to 1 criterion versus another. Nonresponders were older and had longer RA duration than responders, perhaps because they received less aggressive treatment escalation[19]. However, baseline features generally did not distinguish patients who were more likely to respond to 1 criterion versus another.

The second, and the anticipated more likely, explanation for discordant responses was the structure of the response criteria. The main difference was related to acute-phase reactants, and secondarily to patient-reported outcomes. Improvement in acute-phase reactants was less likely than improvement in other measures among ACR20 responders. Responses of acute-phase reactants were the least commonly satisfied elements among EULAR and SDAI50 responders who were not also ACR20 responders. Lack of sufficient improvement in the PtGA, pain score, and HAQ was common among patients who were categorized as improved by the EULAR or SDAI50 criteria but not by the ACR20. Conversely, these were among the elements most highly satisfied in ACR20 responders. Together, these results suggest that some patients achieve an ACR20 response because responses on the patient-reported measures compensate for the lower likelihood of an acute-phase reactant response.

Evaluation of changes in the sole EULAR responders may provide some insight into whether the EULAR criterion is too permissive or the SDAI50 criterion is too strict. These patients had, on average, > 30% improvement in the tender and swollen joint counts, PGA, and PtGA. Forty-three percent of these patients had improvement in the SDAI of ≥ 40%. These findings suggest that many patients with substantial clinical improvement were not categorized as improved by the SDAI50. In our cohort, a 35% improvement in the SDAI was more comparable to both the ACR20 and EULAR moderate/good responses than a 50% improvement. SDAI35 had lower specificity for improvement relative to ACR20 and EULAR moderate/good criteria than the SDAI50, but much higher sensitivity.

The strengths of our study include examination of a cohort with active RA undergoing treatment escalation, detailed examination of discordant responses among the 3 improvement criteria, and testing concordance with different cutoff points of SDAI improvement. Our study is limited in having a relatively small number of patients in the discordant groups. However, we were able to detect features that differentiated among responders.

These results can aid in the interpretation of RA clinical trials. ACR20 responses are somewhat more sensitive to improvement in patient-reported measures such as pain score and HAQ, while EULAR responses tend to be more broad-based, likely because improvement in some features of the DAS can compensate for other features that do not improve as much. SDAI50 was the most stringent response criterion. Modification of this criterion to a 35% improvement in SDAI aligned better with the ACR20 and EULAR moderate/good criteria, and this should be tested in other cohorts.

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.
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.
3. Aletaha D, Martinez-Avila J, Kvien TK, Smolen JS. Definition of treatment response in rheumatoid arthritis based on the simplified and the clinical disease activity index. Ann Rheum Dis 2012;71:1190-6.
4. van de Putte LB, Atkins C, Malaise M, Sany J, Russell AS, van Riel PL, et al. Efficacy and safety of adalimumab as monotherapy in patients with rheumatoid arthritis for whom previous disease modifying antirheumatic drug treatment has failed. Ann Rheum Dis 2004;63:508-16.
5. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, et al; REFLEX Trial Group. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. Arthritis Rheum 2006;54:2793-806.
6. Burmester GR, Mariette X, Montecucco C, Monteagudo-Saez I, Malaise M, Tzioufas AG, et al. Adalimumab alone and in combination with disease-modifying antirheumatic drugs for the treatment of rheumatoid arthritis in clinical practice: the Research in Active Rheumatoid Arthritis (ReAct) trial. Ann Rheum Dis 2007;66:732-9.
7. Bombardieri S, Ruiz AA, Fardellone P, Geusens P, McKenna F, Unnebrink K, et al; Research in Active Rheumatoid Arthritis

(ReAct) Study Group. Effectiveness of adalimumab for rheumatoid arthritis in patients with a history of TNF-antagonist therapy in clinical practice. Rheumatology 2007;46:1191-9.

8. Emery P, Fleischmann R, Filipowicz-Sosnowska A, Schechtman J, Szczepanski L, Kavanaugh A, et al; DANCER Study Group. The efficacy and safety of rituximab in patients with active rheumatoid arthritis despite methotrexate treatment. Results of a Phase IIb randomized, double-blind, placebo-controlled, dose-ranging trial. Arthritis Rheum 2006;54:1390-400.

9. Weinblatt ME, Bingham CO 3rd, Mendelsohn AM, Kim L, Mack M, Lu J, et al. Intravenous golimumab is effective in patients with active rheumatoid arthritis despite methotrexate therapy with responses as early as week 2: results of the phase 3, randomised, multicentre, double-blind, placebo-controlled GO-FURTHER trial. Ann Rheum Dis 2013;72:381-9.

10. van Gestel AM, Anderson JJ, van Riel PLCM, Boers M, Haagsma CJ, Rich B, et al. ACR and EULAR Improvement criteria have comparable validity in rheumatoid arthritis trials. J Rheumatol 1999;26:705-11.

11. Gülfe A, Geborek P, Saxne T. Response criteria for rheumatoid arthritis in clinical practice: how useful are they? Ann Rheum Dis 2005;64:1186-9.

12. Michaud K, Mikuls TR, Call SE, Reimold AM, Hooker R, Kerr GS, et al. Poor to modest agreement between rheumatoid arthritis response measures in clinical practice. Clin Exp Rheumatol 2009;27:633-40.

13. Gülfe A, Aletaha D, Saxne T, Geborek P. Disease activity level, remission and response in established rheumatoid arthritis: Performance of various criteria sets in an observational cohort, treated with anti-TNF agents. BMC Musculoskeletal Disord 2009;10:41.

14. Ranganath VK, Yoon J, Khanna D, Park GS, Furst DE, Elashoff DA, et al; Western Consortium of Practicing Rheumatologists. Comparison of composite measures of disease activity in an early seropositive rheumatoid arthritis cohort. Ann Rheum Dis 2007;66:1633-40.

15. Ward MM, Guthrie LC, Alba M. Clinically important changes in individual and composite measures of rheumatoid arthritis activity: Thresholds applicable to clinical trials. Ann Rheum Dis 2015;74:1691-6.

16. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

17. Rosenthal R, Rosnow RL. Essentials of behavioral research: methods and data analysis. 2nd ed. New York: McGraw Hill; 1991.

18. Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012;141:2-18.

19. Aletaha D, Alasti F, Smolen JS. Chronicity of rheumatoid arthritis affects the responsiveness of physical function, but not of disease activity measures in rheumatoid arthritis clinical trials. Ann Rheum Dis 2015;74:532-7.