Relative Efficiencies of Physician/Assessor Global Estimates and Patient Questionnaire Measures Are Similar to or Greater Than Joint Counts to Distinguish Adalimumab from Control Treatments in Rheumatoid Arthritis Clinical Trials

THEODORE PINCUS, INGRID AMARA, OSCAR G. SEGURADO, MARTIN BERGMAN, and GARY G. KOCH

ABSTRACT. ***Objective.*** To estimate relative efficiencies of the 7 rheumatoid arthritis (RA) Core Data Set measures to distinguish adalimumab from control treatments in 4 clinical trials.

Methods. Four adalimumab clinical trials were analyzed for arithmetic and percentage changes for each Core Data Set measure from baseline to endpoint: 3 assessor/physician measures — swollen joints, tender joints, and global estimate; 1 laboratory test — C-reactive protein; and 3 patient measures — physical function, pain, and global estimate. Relative efficiencies of each measure to distinguish adalimumab from control group responses were assessed, with tender joint count as the referent measure.

Results. Relative efficiencies were in a similar range for physician/assessor, patient, and laboratory measures, with some variation between trials. Among physician/assessor measures, relative efficiencies for global estimates were greater than for swollen and tender joint counts in 8/8 comparisons. Among patient measures, relative efficiencies for global estimates were greater than for physical function and pain scores in at least 6/8 comparisons. Among all measures, relative efficiencies for patient global estimates were greater than for swollen joint counts in 5/8 comparisons, and for tender joint counts in 8/8 comparisons.

Conclusion. Patient and physician/assessor measures distinguished adalimumab from control treatment groups in similar ranges. Among all measures, physician/assessor global estimate was most efficient, and tender joint count least efficient, in all 4 trials. This information suggests that while joint counts are the most specific measure to assess RA, their sensitivity to detect treatment effects in patients with RA is generally no greater, and usually less, than other measures. (First Release Nov 15 2007; J Rheumatol 2008;35:201–5)

Key Indexing Terms:
QUESTIONNAIRES RHEUMATOID ARTHRITIS RANDOMIZED CONTROLLED TRIAL
ADALIMUMAB PATIENT INDEX

No single measure, such as blood pressure or serum cholesterol, can serve as a quantitative "gold standard" to assess individual patients with rheumatoid arthritis (RA). Therefore, pooled indices such as the American College of Rheumatology (ACR) Core Data Set[1], Disease Activity Score (DAS)[2], simplified Disease Activity Index (SDAI)[3], and clinical Disease Activity Index (CDAI)[3] have been developed. All these indices include a formal joint count.

The joint count is the most specific measure to evaluate patients with RA[4], and is regarded by rheumatologists as the most important measure in RA assessment[5]. However, joint counts are poorly reproducible[6-10], and usually are not performed in standard rheumatology care[11]. Further, in analyses of the relative efficiencies of individual measures in the RA Core Data Set to distinguish active from control treatments, only one reported that a tender joint count[12] was the most efficient measure, while 5 other reports indicated that a physician or patient estimate of global status or other patient questionnaire measure was as efficient as or more efficient than swollen and tender joint counts to distinguish active from placebo treatments[13-17].

These observations suggested that further information concerning relative efficiencies of individual measures included in RA indices would be of value to better understand their performance in indices, and their possible use in regular clinical

From New York University–Hospital for Joint Diseases, New York, New York; Quintiles, Inc., Durham; University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Abbott Laboratories, Abbott Park, Illinois; and Arthritis and Rheumatology, Taylor Hospital, Ridley Park, Pennsylvania, USA.

Supported in part by grants from the Arthritis Foundation, the Jack C. Massey Foundation, and Abbott.

T. Pincus, MD, NYU-Hospital for Joint Diseases; I. Amara, DrPH, Quintiles, Inc.; O.G. Segurado, MD, PhD, Abbott; M. Bergman, MD, Arthritis and Rheumatology, Taylor Hospital; G.G. Koch, PhD, University of North Carolina at Chapel Hill.

Address reprint requests to Dr. T. Pincus, NYU-Hospital for Joint Diseases, 301 East 17 Street, New York, NY 10003.

Accepted for publication August 9, 2007.

care. Therefore, we analyzed 4 adalimumab clinical trials for relative efficiencies of the 7 RA Core Data Set measures to distinguish active from control treatments, as presented in this report.

MATERIALS AND METHODS

Clinical data from the adalimumab and control arms of 4 clinical trials were analyzed: ARMADA (DE009)[18] and DE019[19] — adalimumab plus methotrexate (MTX) versus placebo plus MTX; DE011[20] — adalimumab monotherapy versus placebo; and STAR (DE031)[21] — adalimumab plus other traditional disease modifying antirheumatic drugs (DMARD) versus placebo plus traditional DMARD. Analyses of all 4 trials were performed at 24 weeks, although 2 trials were conducted over longer periods, so that comparisons of relative efficiencies would be based on similar periods. If 24-week data were not available, the last observation that was not beyond the latest visit during the 24-week time period with sufficient information to determine ACR 20% response (ACR20) was analyzed. Patients were excluded if any Core Data Set measure was missing at baseline or if no observation was available during the 24-week time period post-baseline — seen in fewer than 2% of instances.

Relative efficiencies of each of the 7 RA Core Data Set measures were calculated using the methods described by Tugwell, *et al*[15]. Arithmetic changes were calculated as differences between mean values for each measure at baseline and endpoint for each treatment group. The standard effect size was the ratio of the difference between the means for the adalimumab and control arms (numerator) to the corresponding calculated standard deviation from analysis of variance for the 2 treatment groups (denominator). Relative efficiency for each Core Data Set measure was obtained by dividing the square of its standard effect size by the square of the standard effect size of the tender joint count as the referent measure, as described by Tugwell, *et al*[15].

In addition to arithmetic changes described by Tugwell, *et al*[15], percentage changes for each Core Data Set measure were computed for each individual patient from baseline to endpoint. Percentage change had 0% assigned for measures with "0" at baseline; "–100%" was assigned for measures with worsening more extensive than 100% — also observed in fewer than 2% of instances. Percentage changes may provide further clarification of the relative efficiencies of each measure, as the percentage changes have the same range of –100% to 100% for all Core Data Set measures. Further, ACR improvement criteria[22] are based on percentage improvement, so that computation of relative efficiencies for percentage changes of the individual measures may be more pertinent to interpretation and reporting of clinical trial results than relative efficiencies for arithmetic changes. Baseline adjusted mean values for percentage changes for each measure were calculated for each treatment group through the use of analysis of covariance (ANCOVA). For percentage change, the standard effect size was computed as the ratio of the baseline-adjusted difference between the mean percentage changes for the adalimumab and control arms from ANCOVA (numerator) to the corresponding estimated standard deviation from ANCOVA (denominator). Relative efficiencies were obtained from the standard effect sizes in the same manner as for arithmetic improvement.

RESULTS

Each ACR Core Data Set measure was improved significantly more from baseline to endpoint in the adalimumab versus control groups in all trials (Table 1). Relative efficiencies were in similar ranges for physician/assessor, patient, and laboratory measures in all 4 trials (Table 2A). Differences were seen from one trial to another, but no single RA Core Data Set measure stood out as substantially more efficient than the others. This finding supports the use of indices to assess RA.

Relative efficiencies according to arithmetic changes were greatest for physician/assessor global estimate in all 4 trials (Table 2A, Figure 1). Among physician/assessor measures, relative efficiencies for global estimates were greater than for tender or swollen joint counts in all 4 trials. Among patient measures, relative efficiencies were greater for global estimates than for physical function or pain in 3 of the 4 trials. Relative efficiencies for patient global estimates were also greater than for swollen joint counts in 2 of 4 trials, and for tender joint counts in all 4 trials.

Relative efficiencies according to percentage changes, adjusted for baseline values, were greatest for C-reactive protein (CRP) in 3 trials, and for patient global estimate in the STAR trial (Table 2B, Figure 1). However, relative efficiencies for CRP were not substantially different from other measures, except in ARMADA (Table 2B, Figure 1). Relative efficiencies for physician/assessor global estimates again were greater than for tender or swollen joint counts in all 4 trials. Relative efficiencies for patient global estimates were greater than for pain in all 4 trials, for physical function and swollen joint count in 3 of 4 trials, and for tender joint count in all 4 trials.

As illustrated in Table 2, the 3 highest relative efficiencies among the 7 RA Core Data Set measures were seen for assessor global estimate in 8 of 8 comparisons, patient global estimate and CRP in 4 of 8, physical function and pain in 3 of 8, swollen joint count in 2 of 8, and tender joint count in 0 of 8 comparisons. Relative efficiencies compared to the tender joint count were greater for assessor global estimate and patient global estimate in 8 of 8 comparisons, swollen joint count in 6 of 8 comparisons, and CRP, physical function and pain in 5 of 8 comparisons. Relative efficiencies compared to the swollen joint count were greater for assessor global estimate in 8 of 8, physical function in 6 of 8, pain and patient global estimate in 5 of 8, CRP in 3 of 8, and tender joint count in 2 of 8 comparisons.

DISCUSSION

The observations presented in this report appear similar to those of previous reports of relative efficiencies in RA clinical trials[13-17] in 2 respects: (1) No single core data set measure was substantially more efficient than any other measure. (2) In particular, joint count measures, the most specific measure in RA[4], appear to be no more sensitive and generally less efficient to detect differences in responses to active versus control treatments. Several differences seen in the 4 different adalimumab clinical trials and in previous reports of other trials appear instructive regarding the rationale for an index of multiple measures to assess patients with RA.

Each individual RA Core Data Set measure distinguished adalimumab from control treatment groups in similar ranges for both arithmetic and percentage differences. The details concerning variation are far less important than the evidence reported here and by others[12-17] that joint count measures are

Table 1. Baseline and endpoint means for adalimumab and control treatment with treatment differences for arithmetic-change means and percentage-change adjusted means (and their standard errors) in 7 ACR Core Data Set measures in 4 clinical trials.

Trial Measures	Adalimumab		Control		Differences in Groups in Change from Baseline	
	Baseline	Endpoint	Baseline	Endpoint	Arithmetic Diff+	Percentage Diff+
ARMADA						
Tender joint count (0–28)	13.9	6.4	14.6	11.8	5.0 (1.5)	42.6 (9.7)
Swollen joint count (0–28)	11.8	4.7	11.0	9.1	5.5 (1.1)	41.2 (8.3)
Assessor global estimate (0–10 VAS)	5.9	2.8	5.9	5.2	2.5 (0.4)	41.9 (7.0)
C-reactive protein (mg/dl)	21.3	5.7	31.0	31.0	16.5 (3.5)	79.7 (9.3)
Physical function (HAQ)	1.6	0.9	1.6	1.4	0.4 (0.1)	25.0 (7.7)
Pain (0–10 VAS)	5.3	2.8	5.7	4.9	1.7 (0.5)	28.5 (9.0)
Patient global estimate (0–10 VAS)	5.7	2.7	5.8	4.9	2.1 (0.5)	37.7 (8.5)
DE011						
Tender joint count	17.7	10.5	17.8	14.2	3.7 (1.2)	22.5 (6.6)
Swollen joint count	14.3	9.6	14.1	12.6	3.2 (0.8)	20.0 (6.4)
Assessor global estimate	6.7	4.0	6.8	5.8	1.8 (0.4)	29.3 (5.6)
C-reactive protein	52.9	30.6	57.5	54.3	18.7 (5.4)	39.2 (7.3)
Physical function (HAQ)	1.8	1.5	1.9	1.8	0.3 (0.1)	18.9 (4.5)
Pain	7.0	4.2	7.0	5.9	1.7 (0.4)	27.6 (5.7)
Patient global estimate	7.3	4.4	7.2	6.0	1.8 (0.4)	27.8 (5.6)
DE019						
Tender joint count	14.2	5.2	15.3	9.7	3.4 (0.7)	28.3 (4.3)
Swollen joint count	13.0	5.2	13.5	9.5	3.6 (0.6)	30.0 (4.5)
Assessor global estimate	6.2	2.4	6.1	4.0	1.6 (0.2)	26.5 (3.6)
C-reactive protein	18.2	8.3	17.3	13.5	8.9 (2.4)	38.1 (4.7)
Physical function (HAQ)	1.5	0.9	1.5	1.2	0.3 (0.1)	27.2 (3.9)
Pain	5.6	2.7	5.6	4.3	1.5 (0.3)	29.1 (4.6)
Patient global estimate	5.2	2.5	5.4	4.2	1.5 (0.3)	33.1 (4.7)
STAR						
Tender joint count	14.8	6.7	14.8	9.8	3.2 (0.6)	22.8 (3.7)
Swollen joint count	13.7	7.4	14.0	10.3	2.6 (0.5)	19.0 (3.5)
Assessor global estimate	6.0	2.9	6.0	4.1	1.3 (0.2)	21.4 (3.1)
C-reactive protein	15.5	9.9	15.4	13.6	3.5 (1.4)	19.5 (3.7)
Physical function (HAQ)	1.4	0.9	1.4	1.2	0.3 (0.04)	19.6 (3.3)
Pain	5.5	3.2	5.6	4.6	1.3 (0.2)	25.0 (3.9)
Patient global estimate	5.4	3.0	5.3	4.3	1.4 (0.2)	27.6 (3.9)
Total						
Tender joint count	15.0	6.8	15.4	10.6	3.5 (0.4)	26.2 (2.5)
Swollen joint count	13.4	6.9	13.6	10.3	3.3 (0.3)	24.6 (2.4)
Assessor global estimate	6.2	2.9	6.1	4.5	1.6 (0.1)	26.0 (2.1)
C-reactive protein	22.9	12.5	24.1	22.1	8.7 (1.3)	33.9 (2.7)
Physical function (HAQ)	1.5	1.0	1.5	1.3	0.3 (0.03)	22.1 (2.1)
Pain	5.8	3.2	5.8	4.8	1.5 (0.2)	26.9 (2.5)
Patient global estimate	5.7	3.1	5.7	4.6	1.6 (0.2)	30.1 (2.5)

ACR: American College of Rheumatology; HAQ: Health Assessment Questionnaire; VAS: visual analog scale.

not associated with greater relative efficiencies compared to the other Core Data Set measures. Indeed, in our study, relative efficiencies of the tender joint counts were lowest in all 4 trials, and of the swollen joint counts in the range of patient questionnaire scores.

Several limitations are seen in these studies. First, the data were derived from clinical trials, which involve only a small minority of all patients, on the basis of inclusion and exclusion criteria[23,24]. For example, the erythrocyte sedimentation rate or CRP is normal in 40% of patients with RA[25,26], many of whom are excluded from clinical trials. Second, in clinical trials, patients are informed that they will be randomized to active or control treatment regimens, rather than assigned an efficacious therapy chosen by a physician. Third, data concerning relative efficiencies remain available for only a few clinical trials; similar analyses of additional trials would appear to be of value.

The swollen and tender joint counts remain the most specific measures to depict the status of a patient with RA[4,5]. The lower relative efficiency of joint count measures compared to other measures in no way suggests that a careful joint examination should not be included as a primary activity in any encounter of a patient with RA and a rheumatologist. However, quantitative joint counts are not performed by most

Table 2. Relative efficiencies (t-statistic) of measures to distinguish adalimumab from control treatment in 4 clinical trials based on analysis of variance for arithmetic change from baseline to endpoint and analysis of covariance for percentage change from baseline to endpoint.

A. Arithmetic Change	ARMADA	DE011	DE019	STAR	No. of Relative Efficiencies Greater than Tender Joint Count	No. of Relative Efficiencies Greater than Swollen Joint Count
Tender joint count	1.00 (3.43)	1.00 (3.04)	1.00 (4.74)	1.00 (5.38)	NA	0/4
Swollen joint count	**2.20 (5.09)**	1.55 (3.78)	1.42 (5.65)	1.12 (5.70)	4/4	NA
Assessor global estimate	**2.72 (5.66)**	**2.65 (4.95)**	**2.06 (6.81)**	**1.66 (6.93)**	4/4	4/4
C-reactive protein	**1.86 (4.68)**	1.30 (3.46)	0.60 (3.66)	0.22 (2.51)	2/4	0/4
Physical function (HAQ)	0.94 (3.34)	1.60 (3.84)	**1.52 (5.85)**	**1.27 (6.05)**	3/4	3/4
Pain	0.92 (3.28)	**2.12 (4.42)**	**1.48 (5.77)**	1.17 (5.82)	3/4	3/4
Patient global estimate	1.48 (4.17)	**2.14 (4.44)**	1.36 (5.54)	**1.43 (6.44)**	4/4	2/4
B. Percentage Change	**ARMADA**	**DE011**	**DE019**	**STAR**		
Tender joint count	1.00 (4.40)	1.00 (3.42)	1.00 (6.60)	1.00 (6.15)	NA	2/4
Swollen joint count	**1.27 (4.96)**	0.83 (3.12)	1.02 (6.68)	0.80 (5.50)	2/4	NA
Assessor global estimate	**1.84 (5.98)**	**2.38 (5.27)**	**1.26 (7.42)**	**1.26 (6.89)**	4/4	4/4
C-reactive protein	**3.89 (8.59)**	**2.46 (5.35)**	**1.50 (8.12)**	0.72 (5.22)	3/4	3/4
Physical function (HAQ)	0.55 (3.25)	1.53 (4.22)	**1.13 (7.04)**	0.94 (5.96)	2/4	3/4
Pain	0.52 (3.17)	2.01 (4.84)	0.92 (6.35)	**1.11 (6.49)**	2/4	2/4
Patient global estimate	1.02 (4.45)	**2.14 (5.00)**	1.13 (7.02)	**1.33 (7.08)**	4/4	3/4

The 3 measures showing the highest relative efficiencies are shown in bold type. NA: not applicable.

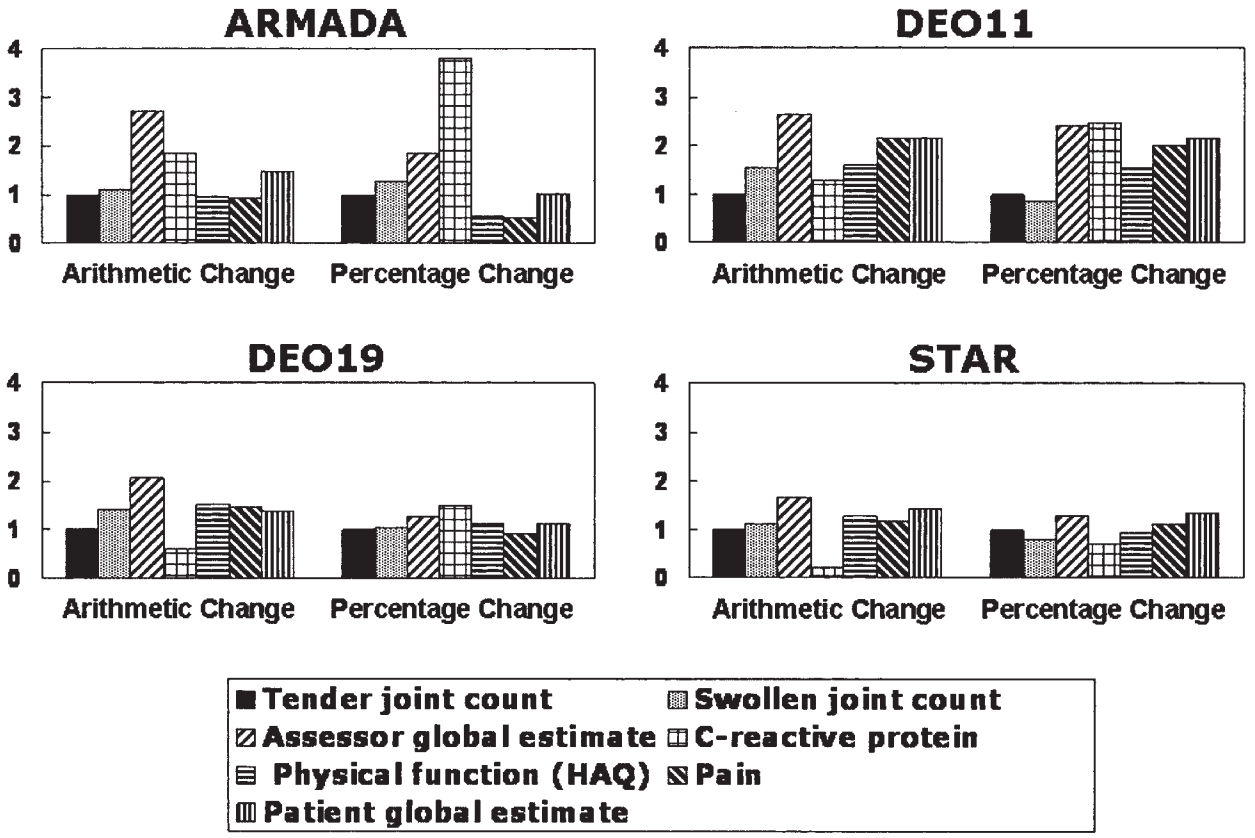

Figure 1. Relative efficiencies of 7 Core Data Set measures to distinguish adalimumab from control treatment in 4 clinical trials according to arithmetic and percentage changes.

rheumatologists at most visits[11], and formal nonquantitative joint assessments may be adequate to monitor patients in a busy clinical care setting. Quantitative monitoring may be accomplished more easily, reliably and efficiently in standard care using patient questionnaires to assess physical function, pain, and patient estimate of global status. Patient questionnaires are easily completed, are at least as sensitive to treatment effects as formal joint counts (and often more sensitive), add minimal burden to the rheumatologist, and might be considered for the infrastructure of a standard care assessment[27].

REFERENCES

1. 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.
2. van der Heijde DMFM, van't Hof M, 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.
3. Aletaha D, Smolen J. The simplified Disease Activity Index (SDAI) and the clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. Clin Exp Rheumatol 2005;23:S100-8.
4. Pincus T. The DAS is the most specific measure, but a patient questionnaire is the most informative measure to assess rheumatoid arthritis. J Rheumatol 2006;33:834-7.
5. Wolfe F, Pincus T, Thompson AK, Doyle J. The assessment of rheumatoid arthritis and the acceptability of self-report questionnaires in clinical practice. Arthritis Care Res 2003;49:59-63.
6. Hart LE, Tugwell P, Buchanan WW, Norman GR, Grace EM, Southwell D. Grading of tenderness as a source of interrater error in the Ritchie articular index. J Rheumatol 1985;12:716-7.
7. Lewis PA, O'Sullivan MM, Rumfeld WR, Coles EC, Jessop JD. Significant changes in Ritchie scores. Br J Rheumatol 1988;27:32-6.
8. Klinkhoff AV, Bellamy N, Bombardier C, et al. An experiment in reducing interobserver variability of the examination for joint tenderness. J Rheumatol 1988;15:492-4.
9. Thompson PW, Hart LE, Goldsmith CH, Spector TD, Bell MJ, Ramsden MF. Comparison of four articular indices for use in clinical trials in rheumatoid arthritis: patient, order and observer variation. J Rheumatol 1991;18:661-5.
10. Scott DL, Choy EHS, Greeves A, et al. Standardising joint assessment in rheumatoid arthritis. Clin Rheumatol 1996;15:579-82.
11. Pincus T, Segurado OG. Most visits of most patients with rheumatoid arthritis to most rheumatologists do not include a formal quantitative joint count. Ann Rheum Dis 2006;65:820-2.
12. Anderson JJ, Felson DT, Meenan RF, Williams HJ. Which traditional measures should be used in rheumatoid arthritis clinical trials? Arthritis Rheum 1989;32:1093-9.
13. Gotzsche PC. Sensitivity of effect variables in rheumatoid arthritis: a meta-analysis of 130 placebo controlled NSAID trials. J Clin Epidemiol 1990;43:1313-8.
14. Bombardier C, Raboud J, The Auranofin Cooperating Group. A comparison of health-related quality-of-life measures for rheumatoid arthritis research. Control Clin Trials 1991;12:243S-56S.
15. Tugwell P, Wells G, Strand V, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: Sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 2000;43:506-14.
16. Cohen SB, Strand V, Aguilar D, Ofman JJ. Patient- versus physician-reported outcomes in rheumatoid arthritis patients treated with recombinant interleukin-1 receptor antagonist (anakinra) therapy. Rheumatology Oxford 2004;43:704-11.
17. Strand V, Cohen S, Crawford B, Smolen JS, Scott DL. Patient-reported outcomes better discriminate active treatment from placebo in randomized controlled trials in rheumatoid arthritis. Rheumatology Oxford 2004;43:640-7.
18. Weinblatt ME, Keystone EC, Furst DE, et al. Adalimumab, a fully human anti-tumor necrosis factor a monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: the ARMADA trial. Arthritis Rheum 2003;48:35-45.
19. Keystone EC, Kavanaugh AF, Sharp JT, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. Arthritis Rheum 2004;50:1400-11.
20. van de Putte LBA, Rau R, Breedveld FC, et al. Efficacy and safety of fully human anti-tumour necrosis factor a monoclonal antibody adalimumab (D2E7) in DMARD refractory patients with rheumatoid arthritis: a 12 week, phase II study. Ann Rheum Dis 2003;62:1168-77.
21. Furst DE, Schiff MH, Fleischmann RM, et al. Adalimumab, a fully human anti-tumor necrosis factor-a monoclonal antibody, and concomitant standard antirheumatic therapy for the treatment of rheumatoid arthritis: results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). J Rheumatol 2003;30:2563-71.
22. 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.
23. Sokka T, Pincus T. Most patients receiving routine care for rheumatoid arthritis in 2001 did not meet inclusion criteria for most recent clinical trials or American College of Rheumatology criteria for remission. J Rheumatol 2003;30:1138-46.
24. Gogus F, Yazici Y, Yazici H. Inclusion criteria as widely used for rheumatoid arthritis clinical trials: patient eligibility in a Turkish cohort. Clin Exp Rheumatol 2005;23:681-4.
25. Wolfe F, Michaud K. The clinical and research significance of the erythrocyte sedimentation rate. J Rheumatol 1994;21:1227-37.
26. Pincus T, Sokka T. Prevalence of normal erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) on presentation of patients with rheumatoid arthritis (RA) at two rheumatology settings, one in the US and the other in Finland: Is a patient questionnaire a better quantitative measure of clinical severity? [abstract]. Arthritis Rheum 2005;52 Suppl:S127.
27. Pincus T, Yazici Y, Bergman M. Saving time and improving care with a multidimensional Health Assessment Questionnaire: 10 practical considerations. J Rheumatol 2006;33:448-54.