Evaluation of different methods used to assess disease activity in rheumatoid arthritis: analyses of abatacept clinical trial data

M Dougados,[1] N Schmidely,[2] M Le Bars,[2] C Lafosse,[3] M Schiff,[4] J S Smolen,[5] D Aletaha,[5] P van Riel,[6] G Wells[7]

ABSTRACT

Objectives: To evaluate different methods of reporting response to treatment or disease status for their ability to discriminate between active therapy and placebo, or to reflect structural progression or patient satisfaction with treatment using an exploratory analysis of the Abatacept in Inadequate Responders to Methotrexate (AIM) trial.

Methods: 424 active (abatacept ~10 mg/kg) and 214 placebo-treated patients with rheumatoid arthritis (RA) were evaluated. Methods of reporting included: (1) response (American College of Rheumatology (ACR) criteria) versus state (disease activity score in 28 joints (DAS28) criteria); (2) stringency (ACR20 vs 50 vs 70; moderate disease activity state (MDAS; DAS28 <5.1) vs low disease activity state (LDAS; DAS28 ≤3.2) vs DAS28-defined remission (DAS28 <2.6)); (3) time to onset (time to first ACR50/LDAS) and (4) sustainability of ACR50/LDAS for consecutive visits. Methods were assessed according to: (1) discriminatory capacity (number of patients needed to study (NNS)); (2) structural progression (Genant-modified Sharp score) and (3) patient satisfaction with treatment. Positive likelihood ratios (LR) evaluated the ability of the above methods to reflect structural damage and patient satisfaction.

Results: MDAS and ACR20 had the highest discriminatory capacity (NNS 49 and 69). Sustained LDAS best reflected no radiographic progression (positive LR ≥2). More stringent criteria (at least ACR50/LDAS), faster onset (≤3 months) and sustainability (>3 visits) of ACR50/LDAS best reflected patient satisfaction (positive LR >10).

Conclusions: The optimal method for reporting a measure of disease activity may differ depending on the outcome of interest. Time to onset and sustainability can be important factors when evaluating treatment response and disease status in patients with RA.

The current gold standard composite assessment used in clinical trials of patients with rheumatoid arthritis (RA) is the American College of Rheumatology (ACR) criteria, evaluated at study endpoint. Increasingly, the disease activity score in 28 joints (DAS28) is also used.[1] Different methods can be used to report these composite indices: (1) response to treatment (eg, ACR criteria) versus disease state (DAS28 criteria); (2) "stringent" versus "less stringent" assessment (eg, ACR70 versus ACR20); (3) time to onset of a successful response (eg, time to achievement of ACR50) or (4) sustainability of a response (eg, maintenance of an ACR50 over a given period of time).

To evaluate the performance of these methods, several different perspectives may be considered: those of the clinical trial investigator, the rheumatologist and the patient. For example, the trial investigator seeks to reduce the costs and risks associated with trial design by determining the discriminatory capacity of specific reporting methods. Primarily, this may be achieved by reducing the number of patients needed to study (NNS) to discriminate between active drug and placebo. In order to prevent irreversible loss of physical function, one of the rheumatologist's primary aims is to inhibit structural damage, as assessed by scores including the Sharp score[2] and its modifications.[3] Finally, patients are concerned with the impact of their condition on daily life, including dimensions such as quality of care. All three groups are concerned with efficacy, safety, sustainability of response to treatment or disease activity status.

The performance of different methods of reporting ACR and DAS28-based criteria according to the different viewpoints or perspectives described above has not previously been studied in a single patient cohort. To address this, we used data from the phase III, Abatacept in Inadequate Responders to Methotrexate (AIM) trial in patients with RA. The efficacy and safety results from this trial have been reported elsewhere, using prespecified primary and secondary endpoints.[4] The objective of the present exploratory analysis was to evaluate different methods of reporting ACR and DAS28-based criteria for their ability to discriminate between active and inactive drugs, to reflect the absence of structural damage progression or to reflect whether patients are satisfied with their treatment. Whereas a range of additional methods is also currently used to assess clinical efficacy,[4 5] exhaustive assessment of all of these measures is beyond the scope of this publication and we have focused on the most commonly used composite indices in clinical trials. To simplify the outputs of this analysis further, presentation of data relating to onset and sustainability have been limited to ACR50 and low disease activity state (LDAS; DAS28 ≤3.2).

METHODS
Database
The analyses reported here were exploratory assessments of a global, phase III, 1-year, multinational, randomised, double-blind, placebo controlled study of abatacept compared with placebo (2 : 1) in combination with methotrexate in

Table 1 Performance of "stringent" versus "less stringent" methods of reporting ACR (treatment response) and DAS28 (disease status) criteria according to their discriminatory capacity and their ability to reflect inhibition of structural damage progression or patient satisfaction

Technique	Less stringent	Intermediate	Stringent
ACR-based techniques at month 6	ACR20	ACR50	ACR70
Overall success, n/N (%)	373/576 (65)	205/576 (36)	98/576 (17)
NNS*, n	69	76	125
Structural damage at year 1†, LR+ (95% CI)	1.11 (0.96 to 1.29)	1.22 (0.93 to 1.60)	1.18 (0.75 to 1.87)
Patient satisfaction at month 6‡, LR+ (95% CI)	2.79 (1.92, 4.05)	Infinite	Infinite
DAS28-based techniques at month 6	MDAS	LDAS	Remission
Overall success, n/N (%)	439/629 (70)	147/629 (23)	68/629 (11)
NNS*, n	49	71	103
Structural damage at year 1§, LR+ (95% CI)	1.20 (1.07 to 1.34)	1.48 (1.06 to 2.06)	1.71 (1.02 to 2.87)
Patient satisfaction at month 6¶, LR+ (95% CI)	2.09 (1.64 to 2.65)	16.82 (4.23 to 66.94)	Infinite

The total number of patients for each analysis was: *n = 636; †n = 551; ‡n = 575; §n = 602; ¶n = 623. A lower number of patients needed to study (NNS) value indicates greater discriminatory capacity; higher positive likelihood ratio (LR+) values indicate a greater probability of observing no structural damage or satisfaction with treatment. Abatacept and placebo treatment groups were pooled for radiographic progression and patient satisfaction and unpooled for NNS. ACR, American College of Rheumatology; DAS28, disease activity score in 28 joints; MDAS, moderate disease activity state (DAS28 (C-reactive protein; CRP) <5.1); LDAS, low disease activity state (DAS28 (CRP) ⩽3.2); remission, DAS28 (CRP) <2.6.

patients with active RA and an inadequate response to methotrexate (clinical trials registration number NCT00048568). The detailed study design of this trial has been reported previously.[6]

Methodology
General considerations
For these exploratory analyses, abatacept was considered the "active" drug and placebo the "inactive" drug. The sample size was based on primary efficacy analyses.[6] The analyses presented here are considered exploratory, as they were not prespecified and the sample size may not be appropriate for statistical testing. Because of the nature of this analysis, the imputation of missing data was not appropriate; all analyses are based on patients with data available at the visit of interest ("as-observed").

Assessments
Disease activity
The following composite indices were assessed on each visit day before study drug administration for a duration of 6 months, at week 2, week 4 and every 4 weeks thereafter: response to treatment, ACR criteria[7 8] (ACR20, 50 and 70) and status of disease, DAS28 criteria (moderate disease activity state (MDAS); DAS28 <5.1), LDAS (DAS28 ⩽3.2) and DAS28-defined remission (DAS28 <2.6)).

Methods of reporting ACR and DAS28 criteria
Response to treatment versus status of disease
To determine the performance of measures that assess response to treatment versus those that assess disease status, the proportion of patients achieving an ACR response or DAS28 status was compared at 6 months.

"Stringent" versus "less stringent" methods
For the ACR criteria, ACR20 was regarded as "less stringent", ACR50 as "intermediate" and ACR70 as "stringent". For DAS28, MDAS was considered as "less stringent", LDAS was considered "intermediate" and DAS28-defined remission was considered "stringent". Assessments were performed at the end of a 6-month study period.

Based on the results obtained in the above analyses (comparing response versus status and "more stringent" versus "less stringent" methods), results for onset and durability were only reported for ACR50 and LDAS, as these measures were of comparable stringency and numbers of responders were high enough to enable meaningful interpretation of the results.

Onset of action
To determine the importance of onset of action, the proportion of patients achieving a first ACR50 response or LDAS within 1 month, or within the first 2, 3, 4, 5 or 6 months of the evaluation period was assessed.

Sustainability of response or status of disease
To determine the importance of sustainability of a response/status, the proportion of patients experiencing ACR50 or LDAS for at least one, two, three, four, five or six consecutive visits over 6 months of the evaluation period was calculated. Equal weighting was applied for each visit.

Assessment of the different methods of reporting ACR response and DAS28-derived criteria
Discriminatory capacity as assessed by NNS
Discriminatory capacity was calculated based on the number of patients required per treatment arm to perform a two-arm 1 : 1 randomised study comparing active treatment with placebo, based on a difference similar to that observed in the AIM study. The number of patients required was calculated with the appropriate basic testing procedure (with $\alpha = 0.05$ (two-tailed), $\beta = 0.20$, χ^2 test for binary variables and Student's t test for continuous variables). The lowest NNS indicates the greatest discriminatory capacity.

Structural damage and patient satisfaction
Structural damage progression in the hands and feet was assessed as radiographic changes from baseline to year 1, using Genant-modified Sharp scores.[3 9] The maximum possible normalised total score (TS) was 290. Data were dichotomised as the percentage of progressors (TS ⩾0) versus non-progressors (TS <0).

Patient satisfaction with treatment was assessed at month 6 or at early termination[10] using the following question on a five-point scale: "how would you rate your satisfaction with the treatment you received?": excellent, 1; very good, 2; good, 3; fair, 4; or poor, 5. Responses were dichotomisd as follows: 1, 2, 3 (favourable) versus 4, 5 (not favourable). A sensitivity analysis was performed using different cut-offs for dichotomization

Table 2 Performance of methods of reporting ACR and DAS28 criteria based on onset of action according to their discriminatory capacity and their ability to reflect inhibition of structural damage progression or patient satisfaction

First ACR50 achieved in	First month	First 2 months	First 3 months	First 4 months	First 5 months	First 6 months
Overall success, n/N (%)	45/636 (7.08)	117/636 (18.40)	199/636 (31.29)	234/636 (36.79)	279/636 (43.87)	302/636 (47.48)
NNS*, n	555	161	59	59	71	58
Structural damage at year 1†, LR+ (95% CI)	2.05 (1.09 to 3.84)	1.60 (1.09 to 2.34)	1.58 (1.22 to 2.05)	1.47 (1.17 to 1.85)	1.32 (1.08 to 1.62)	1.24 (1.02 to 1.51)
Patient satisfaction at month 6‡, LR+ (95% CI)	10.31 (1.44 to 74.10)	6.62 (2.49 to 17.59)	4.43 (2.42 to 8.10)	3.98 (2.36 to 6.72)	4.80 (2.85 to 8.07)	5.21 (3.10 to 8.75)
First LDAS achieved in	**First month**	**First 2 months**	**First 3 months**	**First 4 months**	**First 5 months**	**First 6 months**
Overall success, n/N (%)	26/636 (4)	65/636 (10)	117/636 (18)	154/636 (24)	193/636 (30)	218/636 (34)
NNS*, n	257	207	242	157	116	104
Structural damage at year 1†, LR+ (95% CI)	1.59 (0.65 to 3.85)	1.40 (0.79 to 2.49)	1.62 (1.10 to 2.37)	1.52 (1.10 to 2.09)	1.36 (1.02 to 1.80)	1.31 (1.01 to 1.70)
Patient satisfaction at month 6‡, LR+ (95% CI)	Infinite	7.27 (1.80 to 29.29)	3.16 (1.59 to 6.31)	2.16 (1.32 to 3.54)	2.27 (1.46 to 3.53)	2.59 (1.67 to 4.02)

The total number of patients for each analysis was: *n = 638; †n = 610; ‡n = 632; categories overlap, so patients experiencing a response in the first month will also be counted in all other categories. A lower number of patients needed to study (NNS) value indicates greater discriminatory capacity; higher positive likelihood ratio (LR+) values indicate a greater probability of observing no structural damage or satisfaction with treatment. Abatacept and placebo treatment groups were pooled for radiographic progression and patient satisfaction and unpooled for NNS. ACR, American College of Rheumatology; LDAS, low disease activity state (disease activity score on 28 joints (DAS28) (C-reactive protein) ≤3.2).

(1, 2, (favourable) vs 3, 4, 5 (not favourable)); results were not affected by cut-off choice (data not shown).

To evaluate the relevance of the different reporting techniques for ability to reflect the inhibition of structural damage progression and patient satisfaction, positive likelihood ratios (LR)[11] were calculated. Available abatacept and placebo data for patients in the AIM trial were pooled at each time point (as-observed analysis). An LR determines the likelihood of a given clinical finding in a patient with a studied disorder compared with that in a patient without the studied disorder.[12] The LR, which combines information on sensitivity and specificity, is used to select the best and most appropriate diagnostic tests. LR may range from 0 to infinity. An LR greater than 1 indicates an increased probability that the target disorder is present and an LR less than 1 indicates a decreased probability that the target disorder is present. Likelihood ratios of 2, 5 and 10 increase the probability of the studied disorder by approximately 15%, 30% and 45%, respectively.[12][13] Based on the literature, a positive LR greater than 2 may be considered of relevant prognostic value and results will, therefore, be presented using this cut-off.[14] In our study, the use of LR was transposed to express performance of reporting techniques in reflecting structural damage or patient satisfaction. Higher values are indicative of better techniques. Although an infinite positive LR generally indicates that a technique has good prognostic value, results should be interpreted with caution when the proportion of patients achieving success is low.

RESULTS

Patient disposition

In the 1-year AIM study, 433 and 219 patients were randomly assigned and treated with abatacept or placebo, respectively, on background methotrexate. In total, 385 (88.9%) and 162 (74.0%) patients in the abatacept and placebo groups, respectively, completed 6 months of treatment and 375 (86.6%) and 158 (72.1%) patients were ongoing at 1 year. Baseline demographics and clinical characteristics, including disease activity, were comparable between groups and are described elsewhere.[6] Patients from one site were excluded from efficacy analyses due to compliance issues; 424 and 214 patients from the abatacept and placebo groups, respectively, were included in the primary efficacy analyses and were considered for this exploratory analysis.

Treatment response and disease activity assessments

Overall, at month 6, 512 (80%) patients in the pooled abatacept and placebo population reported satisfaction with treatment as at least "good". At year 1, 97 (15%) patients in the pooled abatacept and placebo population demonstrated no radiographic damage progression from baseline (as defined by a change in the TS <0).

Response to treatment versus status of disease

The ability of ACR (response to treatment) versus DAS28 (disease status) either to detect a treatment effect (NNS) or to reflect inhibition of structural damage progression or patient satisfaction was generally comparable, when comparing methods of reporting with similar levels of stringency (table 1).

"Stringent" versus "less stringent" methods of reporting

Table 1 presents the overall percentage of responders, NNS and positive LR values for "less stringent" (ACR20 and MDAS), "intermediate" (LDAS or ACR50) and "stringent" (ACR70 or

Table 3 Performance of methods of reporting ACR and DAS28 criteria based on sustainability of response according to their discriminatory capacity and their ability to reflect inhibition of structural damage progression or patient satisfaction

ACR50 achieved for	⩾1 Visits	⩾2 Visits*	⩾3 Visits*	⩾4 Visits*	⩾5 Visits*	⩾6 Visits*
Overall success, n/N (%)	302/636 (47)	206/636 (32)	135/636 (21)	93/636 (15)	43/636 (7)	20/636 (3)
NNS†, n	58	51	79	81	149	286
Structural damage at year 1‡, LR+ (95% CI)	1.24 (1.02 to 1.51)	1.46 (1.13 to 1.89)	1.52 (1.07 to 2.16)	1.63 (1.05 to 2.53)	1.82 (0.95 to 3.48)	1.76 (0.66 to 4.74)
Patient satisfaction at month 6§, LR+ (95% CI)	5.21 (3.1 to 8.75)	15.86 (5.16 to 48.73)	31.41 (4.44 to 222.34)	Infinite (N/A)	Infinite (N/A)	Infinite (N/A)
LDAS achieved for	**⩾1 Visits**	**⩾2 Visits***	**⩾3 Visits***	**⩾4 Visits***	**⩾5 Visits***	**⩾6 Visits***
Overall success, n/N (%)	218/636 (34)	132/636 (21)	77/636 (12)	44/636 (7)	22/636 (4)	10/636 (2)
NNS†, n	104	109	134	119	184	406
Structural damage at year 1‡, LR+ (95% CI)	1.31 (1.01 to 1.70)	1.38 (0.95 to 2.00)	2.28 (1.45 to 3.58)	3.05 (1.68 to 5.55)	3.53 (1.48 to 8.40)	2.27 (0.60 to 8.61)
Patient satisfaction at month 6§, LR+ (95% CI)	2.59 (1.67 to 4.02)	5.91 (2.47 to 14.12)	17.58 (2.47 to 125.16)	Infinite	Infinite	Infinite

*Consecutive visits: equal weighting was applied. The total number of patients for each analysis was: †n = 638; ‡n = 638; §n = 610; ¶n = 632. A lower number of patients needed to study (NNS) value indicates greater discriminatory capacity; higher positive likelihood ratio (LR+) values indicate greater probability of observing no structural damage or satisfaction with treatment. Abatacept and placebo treatment groups were pooled for radiographic progression and patient satisfaction and unpooled for NNS. ACR, American College of Rheumatology; LDAS, low disease activity state (disease activity score in 28 joints (DAS28) (C-reactive protein) ⩽3.2).

remission) criteria each evaluated with respect to the discriminatory capacity (at month 6), radiographic progression (at year 1) and patient satisfaction (at month 6).

Using the NNS to determine discriminatory capacity, there was a trend towards greater diagnostic ability with "less stringent" versus "stringent" criteria (table 1). However, interestingly, the NNS for ACR20 and ACR50 were similar.

No difference was observed between "less stringent", "intermediate" and "stringent" methods of reporting ACR and DAS28-based criteria in their ability to reflect inhibition of structural progression (table 1) and all methods of reporting had a positive LR of less than 2.

For patient satisfaction, positive LR were higher with the "stringent" and "intermediate" compared with "less stringent" criteria, suggesting that more stringent criteria better reflected patient satisfaction with treatment (table 1).

Similar analyses were also performed using the OMERACT definition of DAS28 minimal disease activity[4] and DAS28 European League Against Rheumatism (EULAR) good/moderate responders[1] (supplemental table 1, available online only). The results for DAS28 minimal disease activity were comparable to those observed for MDAS and the results observed for EULAR good/moderate responders were similar to those observed for LDAS.

Onset of action

Table 2 presents the overall percentage success rate, NNS and positive LR values for the onset of an ACR50 response compared with achievement of an LDAS according to discriminatory capacity or their ability to reflect inhibition of radiographic progression or patient satisfaction. Analysis of the impact of onset on ACR20 and ACR70 and MDAS and remission is presented in supplemental table 2 (available online only).

For discriminatory capacity, the NNS was lowest when the first ACR50 response occurred any time during the 6-month period compared with occurrence in the first month; a similar trend was observed for LDAS (table 2).

For structural damage progression at year 1, positive LR were less than 2.05 for ACR50 and LDAS (table 2) criteria and were comparable regardless of when the first response was achieved.

For patient satisfaction with treatment, positive LR for time to onset of ACR50 was approximately 4 or more for all onset time points explored and there was a trend towards an increased positive LR with earlier onset (table 2). An earlier onset of action was an important factor in the ability of LDAS to reflect patient satisfaction (table 2). The positive LR of greater than 5 for onset of ACR50 or LDAS in the first 1 or 2 months suggests not only prognostic, but also strong diagnostic evidence (table 2).[14]

Further analyses were performed on the subgroup of patients who achieved an ACR20 or LDAS, examining the impact of whether achievement of this response/status within the first 3 months reflected radiographic progression or patient satisfaction: positive LR were less than 2 for both ACR20 and LDAS (supplemental table 3, available online only).

Sustainability of response or disease status

Table 3 presents the overall percentage success rate, NNS and positive LR values for sustainability of ACR50 versus LDAS according to discriminatory capacity, inhibition of radiographic progression and patient satisfaction. As expected, the overall response rate for ACR50 and LDAS progressively decreased with increasing numbers of required consecutive visits.

For both ACR50 and LDAS, NNS was substantially higher for six consecutive visits compared with one or more consecutive visits (table 3).

For inhibition of structural progression, positive LR was less than 2 for all ACR50 methods of reporting, regardless of how long the response was maintained (table 3). For LDAS, there was a slight trend towards better reflection of inhibition of structural damage (TS <0) with increasing sustainability of response. Sustainability for three or more consecutive visits demonstrated positive LR of 2 or greater (table 3).

Sustainability was an important factor in the ability to reflect patient treatment satisfaction for both ACR50 and LDAS (table 3). LR increased progressively with both sustainability of ACR50 and LDAS.

Analysis of the impact of sustainability on ACR20 and ACR70 and MDAS and remission is presented in supplemental table 4 (available online only).

DISCUSSION

This exploratory analysis of data from the AIM trial strongly supports the concept that the performance of different methods of reporting ACR or DAS28-based criteria is dependent on the desired outcome (eg, a better discriminatory capacity, inhibition of radiographic progression or patient satisfaction with treatment). This is the first study to evaluate the performance of some of these methods in relation to outcomes deemed to reflect the perspectives of the clinical trialist, rheumatologist and patient.

From the perspective of researchers designing clinical trials to test the efficacy and safety of new compounds (eg, phase II trials), our data suggest that the "less stringent" ACR20 and MDAS criteria assessed at the end of the trial achieved the highest discriminatory capacity, allowing detection of a treatment effect using fewer patients. When onset of action and sustainability were considered, an increase was observed in the NNS required to detect a treatment effect, suggesting that when designing clinical trials, it may not be beneficial to take these aspects into account for ACR and DAS28-based criteria.

For the criteria examined in this analysis, sustainability of good disease status (LDAS) for 3 months or more during the first 6 months of the study was the only method of reporting that reflected the absence of radiographic progression at year 1 (using the positive LR cut-off of 2). Previously reported data support this finding. In a longitudinal study including patients who were followed for up to 9 years, fluctuations in disease activity (compared with sustained LDAS or high DAS) were predictive of more severe radiographic progression.[15] All other techniques of reporting assessed here were poor predictors of the inhibition of radiographic progression.

The methods that best reflected patient satisfaction with treatment were "stringent" reporting techniques for both response to treatment (ACR70) and status of disease (DAS28-defined remission). A faster onset (within the first 3 months) and the sustainability of a response/status of disease were important factors in the ability to reflect patient satisfaction. These results would be expected on the basis that a relatively early and sustainable improvement in terms of symptoms, pain, disability and fatigue is likely to be a primary concern for the patient in terms of treatment outcome and quality of life.[16 17]

These findings highlight the importance of the onset and sustainability of a treatment response or disease activity status and support recent EULAR/ACR recommendations that propose that the reporting of clinical trials should include both the time to onset and the sustainability of the primary outcome.[18 19]

Interpretation of our findings should be made in the context of the study limitations. Results were obtained in an exploratory analysis of a single clinical trial evaluating a single compound (abatacept); before proposing firm recommendations, similar analyses should be conducted on data from different trials evaluating alternative compounds. Moreover, similar assessments using other criteria, such as LDAS by DAS28 (using the erythrocyte sedimentation rate) or by the simplified disease activity index or the clinical disease activity index need to be evaluated. For most of the analyses, the observed values of positive LR were not conclusive as the values were below 10 or even 5, the thresholds often used to reflect a "relevant" value for diagnostic purposes.[14] However, a clear, accepted definition of relevant positive LR thresholds does not exist. For example, in the case of evaluation of the risk of toxic events while taking non-steroidal anti-inflammatory drugs, a positive LR value of 1.4 has been considered unacceptable.[20] Conversely, other studies, including one evaluating the ability to predict persistent (erosive) arthritis, have demonstrated that positive LR values of more than 2 may be considered a relevant prognostic value,[21 22] whereas more than 10 is considered a diagnostic value.[12]

Finally, the relatively short duration of follow-up presented here could impact the results. Observations were limited to this time period because the number of patients achieving the more stringent criteria (eg, ACR70 or remission), or onset or sustainability at later time points, was too low to make valid comparisons. However, as most clinicians would expect to observe a treatment effect within 6 months of therapy initiation, this time frame is probably an acceptable period of assessment.

Considering these limitations, our analyses demonstrate that the optimal method of assessment can depend on the outcome of interest, and that onset and sustainability of success may be important factors to consider when assessing the efficacy of therapies for patients with RA. A potential "optimal" technique could be the life-table analyses technique, in which the event is defined by the time taken to reach an acceptable sustained status. Future studies are required to confirm and extend the findings presented here using other RA patient databases, longer study durations and similar analyses in other disease areas.

REFERENCES

1. **Fransen J,** van Riel PL. The Disease Activity Score and the EULAR response criteria. *Clin Exp Rheumatol* 2005;**23**(Suppl 39):S93–9.
2. **Sharp JT,** Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. *Arthritis Rheum* 1971;**14**:706–20.
3. **Genant HK.** Methods of assessing radiographic change in rheumatoid arthritis. *Am J Med* 1983;**75**:35–47.

4. **Wells GA,** Boers M, Shea B, Brooks PM, Simon LS, Strand CV, *et al*. Minimal disease activity for rheumatoid arthritis: a preliminary definition. *J Rheumatol* 2005;**32**:2016–24.
5. **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**(5 Suppl 39):S100–8.
6. **Kremer J,** Genant H, Moreland L, Russell A, Emery P, Abud-Mendoza C, *et al*. Effects of abatacept in patients with active rheumatoid arthritis despite methotrexate: a randomized trial. *Ann Intern Med* 2006;**144**:865–76.
7. **Felson DT,** Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, *et al*. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;**36**:729–40.
8. **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.
9. **Genant HK,** Jiang Y, Peterfy C, Lu Y, Redei J, Countryman PJ. Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs. *Arthritis Rheum* 1998;**41**:1583–90.
10. **Samsa G,** Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the Health Utilities Index Mark II. *Pharmacoeconomics* 1999;**15**:141–55.
11. **Perera R,** Heneghan C. Making sense of diagnostic test likelihood ratios. *ACP Journal Club* 2007;**146**:A8–9.
12. **McGee S.** Simplifying likelihood ratios. *J Gen Intern Med* 2002;**17**:646–9.
13. **Gallagher EJ.** Clinical utility of likelihood ratios. *Ann Emerg Med* 1998;**31**:391–7.
14. **Deeks JJ.** Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ (Clin Res ed)* 2001;**323**:157–62.
15. **Welsing PM,** Landewe RB, van Riel PL, Boers M, van Gestel AM, van der Linden S, *et al*. The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis. *Arthritis Rheum* 2004;**50**:2082–93.
16. **Lee YH,** Ji JD, Song GG. Can we combine patient's and doctor's perspective when assessing rheumatoid arthritis disease activity? *Rheumatol Int* 2007;**27**:1949–52.
17. **Kirwan JR,** Hewlett SE, Heiberg T, Hughes RA, Carr M, Hehir M, *et al*. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis—progress at OMERACT 7. *J Rheumatol* 2005;**32**:2250–6.
18. **Aletaha D,** Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, *et al*. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Ann Rheum Dis* 2008;**67**:1360–4.
19. **Karonitsch T,** Aletaha D, Boers M, Bombardieri S, Combe B, Dougados M, *et al*. Methods of deriving EULAR/ACR recommendations on reporting disease activity in clinical trials of patients with rheumatoid arthritis. *Ann Rheum Dis* 2008;**67**:1365–73.
20. **Lee YH,** Ji JD, Song GG. Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk. *Rheumatol Int* 2007;**27**:477–82.
21. **van Gaalen FA,** Linn-Rasker SP, van Venrooij WJ, de Jong BA, Breedveld FC, Verweij CL, *et al*. Autoantibodies to cyclic citrullinated peptides predict progression to rheumatoid arthritis in patients with undifferentiated arthritis: a prospective cohort study. *Arthritis Rheum* 2004;**50**:709–15.
22. **Visser H,** le Cessie S, Vos K, Breedveld FC, Hazes JM. How to diagnose rheumatoid arthritis early: a prediction model for persistent (erosive) arthritis. *Arthritis Rheum* 2002;**46**:357–65.