

Clinical outcome measures in rheumatoid arthritis

Piet L C M van Riel, Anke M van Gestel

Rheumatoid arthritis (RA) is a chronic systemic inflammatory disease with peripheral synovitis as its main manifestation. The presentation of the disease and the course during time are highly variable both within as well as between individuals. The symptoms and signs of RA may vary from joint complaints like pain, stiffness, swelling and functional impairment, to more constitutional complaints like fatigue and loss of general health. Because of this variety in disease expression a huge number of outcome variables have been used in the past decades to evaluate interventions in clinical trials.¹

Many efforts have been taken in the past years to standardise the assessment of RA aiming at making study results interchangeable. Consensus has been reached about a minimal set of disease activity variables to be measured in clinical trials.²⁻³ As a following step response criteria based on these core set variables have been developed by the European League Against Rheumatism (EULAR),⁴ and the American College of Rheumatology (ACR).⁵ The recent introduction of new, very effective, antirheumatic agents has forced many researchers to modify these response criteria.⁶⁻⁷ The validity of several of them however is questionable and gives rise to as yet unsolved problems, which will be discussed in this paper.

EULAR and ACR response criteria

The two most widely used sets of improvement criteria have been developed following different routes: the preliminary ACR improvement criteria are using all seven core set variables while the EULAR response criteria are based on the Disease Activity Score (DAS), an index using only three or four core set variables. It turned out that including more variables in a combined index did not increase the validity of it.⁸ These two criteria sets also differ in respect to the way they were developed (the ability to discriminate from a placebo response, high versus low disease activity), the implementation of the used variables (reached value, absolute/relative change), and the classification of improvement (two versus three groups) (table 1). The ACR criteria result from a study investigating ease of use, credibility and discriminant validity (versus placebo) of improvement criteria. Treatment response is defined as a 20% change from baseline in core set variables. The EULAR response criteria include not only

change in disease activity but also current disease activity. They are based upon the DAS. To be classified as responders, patients should have a significant change in DAS and also low current disease activity. Three categories are defined: good, moderate, and non-responders. The original DAS was developed based on the in Europe frequently used graded tender joint count, the Ritchie index, and the swollen joint count evaluating 44 joints.

The performance of the EULAR criteria and the ACR improvement criteria has been compared in different clinical trials.⁹ It was shown that they behave similarly with less than 5% discrepancy in responder status. To increase the knowledge about these criteria sets it has been advised to use in a clinical trial setting either of these as the primary and the other as a secondary end point.

Development of 50% and 70% ACR improvement criteria

It has been shown that patients classified as a 20% ACR responder, or a good or moderate responder according to the EULAR criteria do experience a clinical relevant response. Patients fulfilling this classification both show an improvement in functional capacity as measured by the Health Assessment Questionnaire as well as less radiographic progression compared with patients classified as non-responders.⁹ Still questions about the clinical relevance of these improvements have been raised (see Pincus¹⁰). The introduction in the past years of very effective antirheumatic agents like the biological agents is another important reason why researchers were looking for criteria that were more capable to express the major improvements seen by these and other agents. This has led to the use of 50%, 70% and even 90% ACR improvement criteria.

Although it is possible to classify patients according to these high response percentages one should realise that the discriminating power of these criteria is less than with the validated 20% ACR improvement criteria.⁶ One of the reasons for this might be that the ability to reach this response percentage is partly dependent on the disease activity at inclusion. In figure 1 the problems with the percentages improvement is illustrated using the DAS28 as measure of disease activity instead of the core set. The dark grey area is the area of relatively low disease activity (according to rheumatologists treatment

Table 1 Differences between improvement criteria

	ACR	EULAR
1 Developed to distinguish between:	Active and placebo treatment	High and low disease activity
2 Instrument components:	All seven core set variables	Disease Activity Score (index based on 3-4 core set variables)
3 Improvement defined using:	Relative change	Significant change and reached level
4 Classification of improvement:	two categories	three categories

University Medical Centre Nijmegen, Department of Rheumatology, Geert Grooteplein 8, 6525 GA Nijmegen, the Netherlands

Correspondence to: Professor van Riel, (P.vanRiel@reuma.azn.nl)

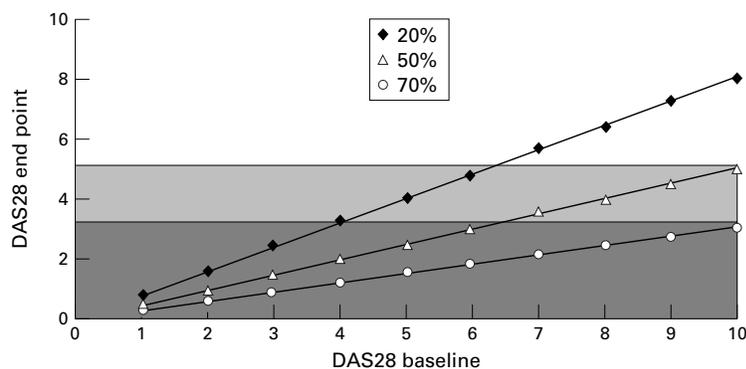


Figure 1 The problem when using a percentage improvement as cut off is illustrated using the DAS28 as single measure of disease activity at baseline and end point of the study. The dark grey area represents low disease activity (according to rheumatologists' treatment decisions), and the lighter grey area represents moderate disease activity. The area of high disease activity is left white. Especially with high disease activity at baseline, a 20% improvement might not be enough, as the reached level of disease activity is still high.

decisions) and the lighter grey area represents moderate disease activity. The area of high disease activity is left blank. Especially with high disease activity at baseline a 20% improvement might not be enough, as the level of disease activity reached is still high. Sharpening the cut offs to 50% or 70% means that in case of a modest or low disease activity at the start of an intervention, the patient should nearly or completely reach a status of remission in order to fulfill such an impressive improvement criterion. Therefore with relative improvement criteria it will be difficult to assess major improvement that is independent of the status of the disease. It has been suggested that the problem should be approached from a different angle: what is really important is for the patient to reach a status of minimal or no disease activity. For this purpose the ACR and EULAR criteria are not applicable, a solution would be to assess the absolute level of disease activity over time (see Remission criteria).

Modifications EULAR response criteria

After it was shown that the 28 non-graded tender and swollen joint count was as valid as the more comprehensive, graded, joint counts,^{11 12} a DAS28 using the 28 tender and swollen joints counts was developed and validated.⁸ The result of the DAS and the DAS28 are not directly interchangeable as the DAS28 has a range varying from about 2 up to 10 and the DAS from about 1 up to 9. However, a transformation formula is available by which one can calculate the DAS28 from the DAS: $DAS28 = (1.072 \times DAS) + 0.938$.

Response criteria using DAS28 were developed and validated against the EULAR criteria using the original DAS and the ACR criteria both using the comprehensive as well as the 28 joint counts. It turned out that the response/improvement criteria using the 28 joint counts were as valid as the criteria using the comprehensive joint counts. It was therefore concluded that for reasons of simplicity the ACR and EULAR criteria using 28 joint counts are preferable.⁹

Remission criteria

Remission criteria define the absence (or a very low level) of disease activity. Any usable criterion should in addition contain a time component as follow up for an indefinite period will not be possible in most clinical settings.

The American Rheumatism Association (ARA) developed preliminary criteria for clinical remission in RA.¹³ The development was based on an optimal discrimination between patients with and without remission according to their rheumatologists. An arbitrary duration of more than two months (the number of assessments was not defined) was chosen because 90% of the patients fulfilled this criterion. There are several problems with this definition that obstruct clinical usefulness. No specifications are given which measurement technique should be used for the different clinical variables, two of the six used variables are not included in the presently accepted and validated core set, and finally the outcome is dichotomous, which implicates that a small change in disease activity may have a great impact on the allocated class.

Another approach would be to define remission with a continuous variable of disease activity such as the DAS and add the time period that the patient was at a certain level or just calculate the cumulative disease activity over a certain time period. A DAS was calculated for patients who were categorised with the ARA criteria as being in remission or not.¹⁴ At a cut off value of the DAS of 1.6 (DAS28 2.6) the percentage of misclassification for both categories was 10%. However, as it is frequently observed that the disease activity of a patient may fluctuate around the level of "no or minimal" disease activity a better way of expressing the disease status of a patient would be the cumulative amount of disease activity over a certain period of time (area under the curve) or the mean disease activity in a certain period in stead of classifying a patient as in remission.

Numeric ACR and area under the curve ACR criteria

To be able to more accurately express the improvement of patients in clinical trials the numeric ACR criteria (ACR n) have been proposed. Patients are classified according to the least percentage change (from baseline) in the ACR criteria. In this way patients can be classified as percentage responder ranging from 0–100%. Based on the ACR n, area under the curve ACR criteria (ARC AUC) have been developed and used in a clinical study.⁷ The hypothesis of the authors was that this represents better the responder status of a patient over time. However, in this clinical improvement measurement the baseline value plays a very dominating part. Applying this method means that at each time point the disease status is being compared with the disease status at the start of the study. This is illustrated in figure 2. In figure 2A patient A and B follow the same course of disease activity, except for the baseline value. The response

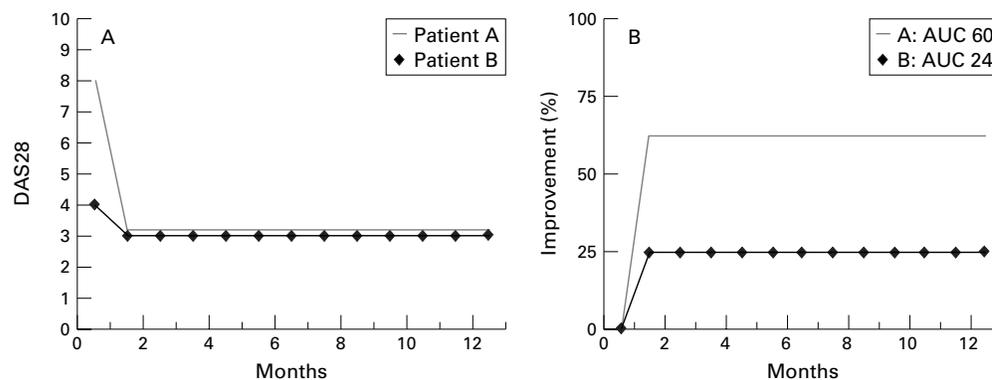


Figure 2 (A) Change in DAS for patients A and B. (B) Calculated ACR AUC for patients A and B.

percentage for patient A is 63% and for patient B this is 25%. The AUC for percentage improvement is shown in figure 2B with the corresponding values. It can be seen that a small difference in the baseline value has a great impact using the ARC AUC.

Finally the meaning of a certain value of the ACR AUC has not been determined: do a group of patients with a ACR AUC that is twice as high as those of another patient group have less radiographic progression or a better functional capacity?

Inclusion criteria

Clinical trials evaluating biological agents often select patients with “active” disease. Most studies use different definitions of the minimal required level of disease activity of patients entering the trial. This results in differences between trial populations, which will hamper the comparison of trial results. Standardised inclusion criteria should be formulated, and might be based upon the core set or an index of disease activity. An index would provide the advantage of a single figure and a continuous scale, so that (as with remission) “active” disease can be seen relative to other levels of disease activity.

When using one of the improvement criteria (EULAR or ACR) trial inclusion criteria are required. With the EULAR criteria a DAS28 > 3.2 at the beginning of the study will be necessary, as a DAS28 < 2.0 indicates the absence of disease activity, and a change of 1.2 should be possible to be able to meet the (good) response criteria. With the ACR criteria all included variables should be at least larger than zero at baseline, as dividing by zero (to calculate relative change) is not possible. Because of this, the HAQ score at baseline will be the bottleneck for calculating the ACR improvement criteria especially in early disease.

Daily clinical practice

In daily clinical practice most rheumatologists do not consistently monitor the disease course as is the case in clinical trials; often even a global way of assessing disease activity is being performed. One of the reasons for this may be the lack of time during the outpatient visits and the lack of a simple instrument to perform this assessment. The recently introduced very effective but also expensive agents like the bio-

logical agents forces the rheumatological society to more accurately assess patients with RA. To optimise the treatment with these agents it is necessary to monitor the disease as accurately as possible to titrate these treatments according to the level of disease activity. An index expressing disease activity as a single continuous variable, will be the most helpful measure to follow up the course of the disease. The DAS28 including two joint counts, an acute phase reactant, and a general health assessment is a valid, easy to use instrument for this purpose. Although the formula to calculate the DAS28 is rather complicated, by using a simple programmable calculator (or a personal computer), it only takes a few seconds.

As outcome variables the Health Assessment Questionnaire (HAQ) at 6 or 12 months intervals, and even at longer intervals radiographs from hands and feet scored either by the Larsen method or the Sharp method should be taken.

Conclusion

Many measures of outcome are being used in the assessment of RA. Core sets of valid outcome measures have been defined to be used in clinical trials. Two criteria sets for assessing improvement in clinical trials, ACR improvement criteria and EULAR response criteria have been developed and showed comparable validity. The 50% and 70% ACR improvement criteria showed less discriminating capacity than the original 20% criteria, major improvement should therefore be assessed with an absolute level of disease activity like the remission criteria based for instance on the composite index of disease activity the DAS. The ACR AUC is not a useful instrument in the evaluation of responders status over time.

- 1 van Riel PLCM, van de Putte LBA. Clinical assessment and clinical trials in rheumatoid arthritis. *Curr Opin Rheumatol* 1994;6:132-9.
- 2 Van Riel PLCM. Provisional guidelines for measuring disease activity in clinical trials on rheumatoid arthritis. *Br J Rheumatol* 1992;31:793-6.
- 3 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.
- 4 van Gestel AM, Prevoo MLL, van't Hof MA, *et al.* Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. *Arthritis Rheum* 1996;39:34-40.
- 5 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.

- 6 Felson DT, Anderson JJ, Lange ML, *et al*. Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent? *Arthritis Rheum* 1998;41:1564–70.
- 7 Schiff M, Weaver A, Keystone E, *et al*. Comparison of ACR response, Numeric ACR, and ACR AUC as measures of clinical improvement in RA clinical trials. *Arthritis Rheum* 1999;42 (suppl):S81.
- 8 Prevo MLL, van't Hof MA, Kuper HH, *et al*. Modified disease activity scores that include twenty-eight-joint counts. *Arthritis Rheum* 1995;38:44–8.
- 9 van Gestel AM, Anderson JJ, van Riel PLCM, *et al*. ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. *J Rheumatol* 1999;26:705–11.
- 10 Pincus T, Stein CM. ACR20: clinical or statistical significance? *Arthritis Rheum* 1999;42:1572–6.
- 11 Prevo MLL, van Riel PLCM, van't Hof MA, *et al*. Validity and reliability of joint indices. A longitudinal study in patients with recent onset rheumatoid arthritis. *Br J Rheumatol* 1993;32:589–94.
- 12 Smolen JS, Breedveld FC, Eberl G, *et al*. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. *Arthritis Rheum* 1995;38:38–43.
- 13 Pinals RS, Masi AT, Larsen RA. Preliminary criteria for remission in rheumatoid arthritis. *Arthritis Rheum* 1981;24:1308–15.
- 14 Prevo MLL, van Gestel AM, van't Hof MA, *et al*. Remission in a prospective study of patients with rheumatoid arthritis. American Rheumatism Association preliminary remission criteria in relation to the disease activity score. *Br J Rheumatol* 1996;35:1101–5.