Defining remission in rheumatoid arthritis

David Felson[1,2]

[1]Clinical Epidemiology Research and Training Unit, Boston University, Boston, Massachusetts, USA

[2]Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester, UK

Abstract

Objective—Given the inconsistency of remission definitions in rheumatoid arthritis (RA) trials, the goal of this American College of Rheumatology/European League Against Rheumatism committee was to define remission.

Methods—The committee instructed a working group that a new remission definition, among other requirements, needed to allow for little, if any, active clinical disease and to be defined using the core set of outcome measures for RA trials and that those in remission at one time needed to have a low risk of later worsening function or radiograph progression. Remission was to be defined using trial data for use in trials but needed to anticipate use in a practice setting.

Results—The working group started by evaluating the thresholds for core set measures compatible with remission and determined that patient-reported outcomes contributed importantly to the ability of outcome assessment to distinguish more from less effective treatments. The group created a candidate group of remission definitions to test, including Boolean versions and widely used indexes. Testing how well these candidate definitions predicted later good outcomes, the group found that Disease Activity Score 28 thresholds for remission performed worse than Simplified Disease Activity Index/Clinical Disease Activity Index or Boolean versions. Also, persons with low Disease Activity Score 28 occasionally had high joint counts, which were incompatible with remission. The parent committee chose two definitions: one Boolean (patient had to have all of the following: tender joint count, swollen joint count 1, C reactive protein 1 mg/dl) and patient global assessment 1 (on a 0–10 scale) and one Simplified Disease Activity Index 3.3.

Conclusion—The American College of Rheumatology/European League Against Rheumatism has promulgated two new similar definitions of remission for RA trials.

The last several years have seen a re-evaluation and formal redefinition of the state of remission in rheumatoid arthritis (RA) carried out by the American College of Rheumatology/European League Against Rheumatism (ACR/EULAR). This effort was led by myself and Professors Joseph Smolen and Maarten Boers.[1]

Remission is a state rather than the description of a transitional state or change. It represents an absence of disease activity, although practically speaking, to require a complete absence of disease activity in RA would define a state that almost no patient with disease could meet. Thus, practically, remission is generally operationalised as either a complete absence of

disease activity or a level of disease activity so low that it is not troublesome to the patient and portends a later good prognosis.

There are many reasons why a new definition of RA remission is needed. The official American Rheumatism Association of RA,[2] now 30 years old, was well validated but contains within it elements that are no longer part of the core set of measures generally used to evaluate disease activity in RA. For example, the ARA remission definition requires no morning stiffness and no soft tissue swelling in joints or tendon sheaths. Because of its lack of sensitivity to change, morning stiffness was not selected as part of the core set of measures, and tendon sheath swelling is not evaluated. In addition, the ARA definition of remission has been so stringent that few patients have had disease activity low enough to meet this definition. A variety of modifications of this definition has been used.

On the other hand, a Disease Activity Score 28 (DAS28)-based definition of remission in RA is thought to be lax, allowing many patients with active disease to be characterised as being in remission. Other definitions of remission have been offered, such as Simplified Disease Activity Index (SDAI) and Clinical Disease Activity Index thresholds.

A single recommended definition is needed, backed up by evidence of its validity. There are other reasons why remission is important to define. Those in remission have better functional outcomes of disease and better radiographic outcomes of disease than those not in remission.[34] Lastly, with the improvement in treatment of RA, remission has become an achievable outcome for many patients, and defining it in a uniform reproducible way will permit the increasing inclusion of a new therapeutic target into trials and even practice.

Recognising these needs, the ACR and EULAR constituted a committee which they tasked with developing a new definition of remission in RA. They created a large parent committee made up of many RA clinical researchers from the international community; there was a core group headed by Drs Felson, Smolen and Boers, who acted on the suggestions of the larger committee carrying out and evaluating data analyses and then presenting results to the committee for decision making. The ACR committee laid out a series of requirements for any new definition of remission.

1. The definition of RA must be stringent, allowing for little if any residual active disease

2. make use of the ACR/EULAR core set measures[56]

3. include at least the following core set measures: tender and swollen joint counts (SJCs) and acute phase reactant

4. not include measures of physical function, as they are affected by disease duration and would be used as an outcome for validation

5. not include information on the presence or absence of treatment; patients could be in remission either on or off treatment

6. not include a duration, so that remission could occur at one time point or over a period of time

7. predict a later good outcome defined as a later lack of x-ray damage and stable good function

8. be defined for use in trials with a subsequent modification for clinical practice.

To evaluate any candidate definitions of remission, we used data graciously provided to us by industry sources, which gave us access to raw data from large multicentre randomised trials of recently introduced therapies in RA.

Our work was organised into several stages. In Stage 1, we decided with the parent committee what cut points for core set measures were compatible with remission. If a tender joint count (TJC), for example, was too high, then clinicians might suggest that this high TJC was not compatible with the patient being in remission. We surveyed committee members, asking them what threshold was compatible with remission for each core set measure. We created two scenarios. First, we asked what threshold for a core set measure was compatible with remission if all other measures pointed to remission, and second, what threshold was compatible with remission with the use of data from only that measure. While the first survey gave us a variety of thresholds, we focused our attention on the second, in which we asked our respondents to give us information on the thresholds that would be required if that core set measure were the only measure evaluating remission. With the use of that survey, results clustered around the value of 1 for all measures. For example, a TJC of 1 or less was compatible with remission as was an SJC of 1 or less. A patient global assessment of 1 or less (on a 0–10 scale) was also compatible with remission, as was a C-reactive protein (CRP) of 1 mg/dl or less.

In Stage 2 of our work, we asked a specific question raised by the parent committee: whether patient-reported outcomes should be included in the ultimate definition of remission. The patient-reported outcomes in the core set included a patient global assessment and a pain assessment. While the committee felt that assessing patient symptoms was important, we also felt that there needed to be hard statistical evidence that including patient-reported outcomes in a definition of remission would enhance the ability of that definition to discriminate active from less active treatment. We focused therefore on large randomised trials in which generally a combination of biologics and methotrexate was compared with a single drug, a situation in which it has been consistently shown that the combination is more efficacious and which evaluated whether patient-reported outcomes aided in the discrimination between combination therapy and single therapy. If patient-reported outcomes aided in this discrimination, they would, at threshold levels, discriminate well between combination therapy and single therapy and would do so at least as well as non-patient-reported outcomes. In three of the four large randomised trials, patient-reported outcomes were among the best discriminating outcome measures, suggesting that they provided additional information about treatment efficacy that was not provided by the already selected measures, TJC and SJC, and CRP. Thus, based on this statistical evidence, we decided to include one or more patient-reported outcomes in the definition of remission.

Having now identified which type of measures should be included in the candidate definitions of remission and having decided on the threshold for each measure, we turned to Stage 3—the task of creating a candidate list of remission definitions. These included a series of Boolean definitions (which depended on reaching a low level in each core set measure) and a series of indexes such as the DAS28 and SDAI. The specific definitions tested (this is not a comprehensive list) are shown in Table 1, including the DAS28 at two different thresholds. The DAS28 formula, reveals that in the DAS28 the TJC is weighted twice as much as the SJC, suggesting that it could be interpreted as a weighted TJC.

In Stage 4, we turned to selecting among the candidate definitions. Our strategy here was to look at the predictive validity of each definition. We defined predictive validity as how well patients meeting the candidate definition early in a clinical trial had good radiographic and functional results later in that same trial. We defined early as 6 months and later as the interval between 1 and 2 years. For those candidate definitions performing well, a large percentage of patients meeting them would have later stable radiographs and good functional outcomes, whereas those candidate definitions not performing well would not show as high a likelihood of later good outcomes. As shown in table 1, all candidate definitions showed significant discrimination in terms of later good outcome between those

meeting remission at 6 months and those not meeting remission. Likelihood ratios provide a somewhat better gauge of how well these candidate definitions performed. The Boolean definitions with the highest likelihood ratios for later good outcomes were those which had at least four measures, including TJC, SJC, CRP and patient global assessment or patient pain. In terms of indexes, the higher threshold for DAS28 of 2.6 did not perform as well as the more stringent DAS28 threshold or the SDAI 3.3, which had the highest likelihood ratio and was also the most significant in terms of its predictive capability. Although not shown, the DAS28 did not perform well compared with SDAI in predicting later radiographic outcomes (for DAS28 <2.6, the likelihood ratio was 1.0, and for DAS28 <2.0, the likelihood ratio was 1.6; both failed to reach statistical significance as a predictor of later x-ray outcome). The SDAI 3.3, on the other hand, had a positive likelihood of 3.0 for x-ray outcome (p=0.003).

The predictive validity analysis results suggested that Boolean definitions all performed well, with those containing SJCs and TJCs, a CRP and a patient-reported outcome having similar predictive validity. On the other hand, indexes did not perform the same. The DAS28 <2.6 did not predict a later good outcome as well as a lower threshold for the DAS28 (2.0) or the SDAI at the recommended threshold (3.3). The DAS28 <2.0 also did not predict x-ray outcome well and was in fact rarely achieved (<1/3 as often as other index thresholds). One reason for the relatively weaker performance of the DAS28 is that the DAS28, as noted above, is a weighted TJC, and the prediction of later x-ray outcomes may be more dependent on the SJC than the TJC.[7]

For Stage 5, we evaluated face validity—whether candidate definitions produced values that were compatible with remission, especially the index measures. We carried out a series of analyses, asking whether specific core set measure values in those who met remission criteria were incompatible with remission (see figure 1). For DAS28, we found that some patients meeting even stringent remission threshold had SJCs that were incompatible with remission. For example, the highest SJC of those patients meeting DAS28 <2.6 was above 20. Ten per cent of the patients meeting DAS28<2.6 had SJCs ranging from 4 to over 20; all values were incompatible with remission. Even when a lower threshold for DAS28 was used (2.0), SJCs were frequently above levels that are compatible with remission (eg, values of 5 or greater). Boolean definitions required that swollen joints and tender joints and other values be 1 or less. SDAI, by the nature of its computation, also required SJCs that were low.

The ACR/EULAR parent committee reconvened in October of 2009 and achieved consensus on definitions of remission selecting a Boolean definition and an index definition based on the analyses presented above. The definitions selected by the committee are shown in table 2. Subsequent analyses of trials suggested that roughly 9%–12% of patients in the trials we studied would have achieved remission based on these definitions and that those rates were similar for the SDAI and Boolean definitions.

A number of additional issues arose as part of our discussions, of which only some were resolved and constitute issues for further and later evaluation. First, our trials generally used a 28-joint joint count (the 28-joint count leaves out ankles and feet), but clinically, it made little sense to define a patient as in remission when joints outside of the 28 joints evaluated in this count may have active disease. Analyses of two trials that had a fuller joint counts suggested that few, if any patients, would be missed by the prescribed definition—that there were few patients who had active ankles and feet but inactive joints elsewhere and that most of those patients had patient global assessments that were high enough that they would not be eligible to be classified as in remission. However, the committee did not feel that it had sufficient data to dictate practice on this matter. The criteria do not require inclusions of

ankle and feet and can be completed with a 28-joint joint count, but we recommend that the additional joints be also included in examinations and that investigators be clear about which joints are examined and reported.

We note further that some investigators had suggested that an absence of positive findings on ultrasound imaging of the hands might provide useful additional prognostic information to that of the ACR remission threshold.[8] In fact, the data from Brown *et al* were collected in a group of RA patients limited to those in DAS remission; thus, a whole range of activity measures/activity scores were not available, meaning that remission would have less predictive capability than it would in a broadly based data set. Further, even in the data presented by Brown *et al*, the old ARA remission criteria (not the current criteria) had better predictive capability for later x-ray progression than the median ultrasound power Doppler score (OR of 3.0 for progression in those not in ARA remission vs an OR of 1.85 for those not in power Doppler remission). We do not rule out the possibility that adding imaging information will provide complementary data that would better prognosticate a patient's likelihood of subsequent progression and better biologic information on whether they are actually in remission. For now, however, we recommend clinical remission thresholds as suggested by the ACR/EULAR committee.

In conclusion, there is a new consensus ACR/EULAR definition of remission in RA. It is stringent yet achievable and should be a major outcome for trials. Variants of these definitions may be utilised in practice settings, but additional research validating these outcomes in practice settings is needed.

Acknowledgments

REFERENCES

1. Felson DT, Smolen JS, Wells G, et al. American College of Rheumatology/European League Against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. Arthritis Rheum. 2011; 63:573–86. [PubMed: 21294106]

2. Pinals RS, Masi AT, Larsen RA. Preliminary criteria for clinical remission in rheumatoid arthritis. Arthritis Rheum. 1981; 24:1308–15. [PubMed: 7306232]

3. Aletaha D, Ward MM, Machold KP, et al. Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states. Arthritis Rheum. 2005; 52:2625–36. [PubMed: 16142705]

4. Koevoets R, Machado P, Bombardier C, et al. The value of conventional radiographs in undifferentiated arthritis: a systematic review. J Rheumatol Suppl. 2011; 87:26–30. [PubMed: 21364053]

5. 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. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Arthritis Rheum. 1993; 36:729–40. [PubMed: 8507213]

6. Boers M, Tugwell P, Felson DT, et al. World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. J Rheumatol Suppl. 1994; 41:86–9. [PubMed: 7799394]

7. Kaarela K. Prognostic factors and diagnostic criteria in early rheumatoid arthritis. Scand J Rheumatol Suppl. 1985; 57:1–54. [PubMed: 3863243]

8. Brown AK, Conaghan PG, Karim Z, et al. An explanation for the apparent dissociation between clinical remission and continued structural deterioration in rheumatoid arthritis. Arthritis Rheum. 2008; 58:2958–67. [PubMed: 18821687]
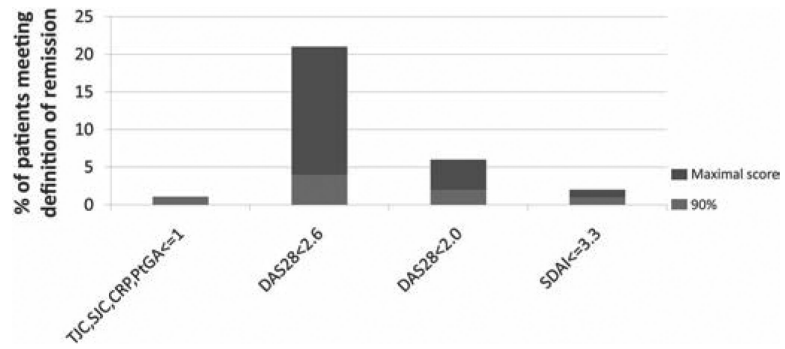
Figure 1.
Face validity: if you meet the remission definition, do you always have a low swollen joint count?

Table 1

Candidate remission definitions and their predictive validity: predicting a good outcome for both x-ray and HAQ

	Remission with good outcome (%)	NOT in remission with good outcome (%)	Positive likelihood ratio	p Value
TJC28, SJC28, CRP 1	46	17	3.2	<0.0001
+ PtGA 1	66	17	7.2	<0.0001
+ Pain 1	60	17	5.7	<0.0001
+ PhGA and PtGA 1	68	17	8.0	<0.0001
+ PhGA and pain 1	64	18	6.7	<0.0001
+ PtGA and pain 1	64	17	6.8	<0.0001
+ PhGA, PtGA and pain 1	67	18	7.5	<0.0001
INDEXES				
DAS28<2.6	38	18	2.2	0.01
DAS28<2.0	56	20	4.5	0.01
SDAI 3.3	56	17	4.8	<0.0001

CRP, C-reactive protein; DAS, Disease Activity Score; SDAI, Simplified Disease Activity Index; SJC, swollen joint count; TJC, tender joint count.

Table 2

ACR/EULAR provisional definitions of remission for clinical trials

Boolean-based definition
At any time point, a patient must satisfy all of the following:
TJC   1
SJC   1
CRP   1 mg/dl
Patient Global Assessment   1 (on a 0–10 scale) ('all the ways your arthritis affects you')
Index-based definition
At any time point, a patient must have SDAI   3.3

ACR/EULAR, American College of Rheumatology/European League Against Rheumatism; CRP, C reactive protein; SDAI, Simplified Disease Activity Index; SJC, swollen joint count; TJC, tender joint count.