

Measuring Flares in Rheumatoid Arthritis. (Why) Do We Need Validated Criteria?



In the treatment of rheumatoid arthritis (RA) it is of growing importance to measure disease activity both in clinical practice as well as in research. The last 20 years have brought us several well-validated disease activity indices: for example, the Disease Activity Score (DAS), DAS28 for 28 joints, the Clinical Disease Activity Index, and the Simplified Disease Activity Index are currently being used, and validated cutoff points to determine disease activity states as well as change criteria to indicate improvement to therapy have been developed^{1,2,3,4,5,6}. However, in addition to measuring absolute disease activity states and improvement, there is an increasing need for assessing RA flare or worsening. Therefore, at the OMERACT 9 (Outcomes in Rheumatology) meeting a working definition of RA flare was proposed: flare occurs with any worsening of (or return of) disease activity that would, if persistent, lead to (re)initiation, increase or/and change of therapy; a flare represents a cluster of symptoms of sufficient duration intensity to require (re)initiation, change, or increase in therapy¹. Although this working definition was an essential first step, research is needed on validated flare criteria, and the work of Bykerk, *et al* in this issue of *The Journal* represents an important contribution in the field⁷. Here, we would like to discuss several aspects of development and use of RA flare criteria.

First, why do we need thoroughly validated flare criteria? The first scenario that exemplifies the need for a flare criterion is the use of fire-and-forget type of treatments such as rituximab, in which the timing of retreatment is often based on occurrence of a worsening in disease activity. Flare criteria are also essential in down-titration and discontinuation studies as well as tapering of medication in daily clinical practice to determine clinically relevant worsening to guide reinstating therapy or increasing dose. Finally, in comparing (biologic) disease-modifying antirheumatic drugs (bDMARD) it could be of interest to see which treatment has the lowest risk of occurrence of in-between

flares, because primary outcome measures such as percentage of low disease activity or percentage of American College of Rheumatology (ACR) improvement at study end appear comparable for different drugs in patients with baseline high-disease activity, but stability of the disease activity may not be comparable^{8,9,10}. With reaching remission as a goal and the knowledge that periodic worsening is associated with radiographic damage, the frequency, number, and severity of flares might be an interesting additional variable in the near future to compare the efficacy of bDMARD¹¹.

There are, however, some issues to be clarified when considering heterogeneity of flare criteria used in clinical research. There is indeed a considerable variation noticeable in flare criteria that have been used in clinical studies or have been proposed in literature^{1,12}. These criteria vary from an increase in number of swollen joints to physician's decision to change treatment (which would be an interesting circularity of course when used in clinical practice), a worsening of components of the ACR response criteria or worsening based on DAS28, to patient-reported flares^{1,12,13}. Indeed, recently it was shown by Yoshida, *et al* that almost all biologic discontinuation studies have used a different criterion to decide on treatment resumption or to determine disease worsening¹². This wide variety of different criteria is undesirable because data on flares may be difficult or impossible to compare.

How to resolve this issue of heterogeneity? Looking at the variety of criteria, there seem to be 2 main approaches: the patient-reported flare-based criteria, and the joint-score and laboratory-test based criteria. Both approaches have pros and cons that mainly concern the content of the domains measured by the criterion, and second, the need or not for face-to-face contact with the patient.

When, for example, disease activity measurement (e.g., DAS28 measurement) is incorporated in routine clinical care, a flare criterion based on this measure is easily calcu-

See Flares in RA, page 227

lated to guide patient and physician. This however requires that patients have low threshold access to healthcare once they experience a flare, which means that travel distance and admission times should be acceptable for the outpatient clinic visit. Also, questions on self-management will not be included in joint-score-based criteria, although this domain has proved to be very interesting in the OMERACT 10 Delphi procedure¹⁴.

Patient-reported flares, on the other hand, could easily be used at home to guide a patient in contacting their physician to discuss over the phone what treatment changes are necessary, or maybe even to execute a predetermined plan with their physician to change treatment. However, because no input from the physician nor objective disease activity indices (e.g., acute-phase reactants) are incorporated, there is risk of underreporting or overreporting flares in patients; moreover, because of response shift, patients' judgment has been shown to be impaired regarding the level of disease activity, as well as longterm changes in disease activity (although for short-term changes such as disease flare this problem should be smaller). Therefore the ideal criterion is probably a combination of both patient-reported joint score and laboratory testing-related items as demonstrated by the validation of the different flare domains used in the OMERACT preliminary flare criteria¹³. Whether this is feasible in daily clinical practice remains a question and can heavily depend on local contexts of healthcare.

A complicating factor in validating flare criteria and resolving this heterogeneity issue is the lack of gold standard for flare. Looking at the validation of patient-reported flare criteria and the joint score and laboratory testing-based flare criteria, researchers used either patients' report on worsening, or a worsening in joint score and vice versa, thus mutually anchoring their flare criterion. For example, where Bykerk, *et al* (published in this journal) demonstrate a relation between patients reporting a flare and the DAS28, we in turn reported a relation between DAS28-based flare criteria and patients reporting disease worsening¹⁵. This well-known, back-and-forth stepping stone technique remains the solution for validation studies when no gold standard is available. However, a more external standard is necessary to resolve the question of which of the approaches is favorable in which situation. Interesting alternatives for a gold standard could be using radiographic outcome as an anchor, although it reflects a late consequence of flare rather than the concept of flare itself. Another anchor could be "functioning," which also demonstrated to be strongly correlated to flare; however, function is also a patient-reported outcome (e.g., modified Health Assessment Questionnaire). Other more novel techniques, including positron emission tomography and biomarkers could be used, but those have the disadvantage that they represent a more technical pathophysiological representation of flare that is further from the patient experience of flare. So, the

ideal gold standard for flare to use in validation studies has yet to be found.

A final issue with regard to validation and use of flare criteria is that the concept of flare might be a moving target. As treat-to-target strategies have demonstrated that aiming for low disease activity and remission has become an accessible goal, a (threshold) shift could occur in what patients and physicians see as a flare. Interestingly, this is exemplified by the OMERACT working definition of flare, as it includes the phrase "any worsening of (or return of) disease activity that would, if persistent, lead to (re)initiation, increase or/and change of therapy." Recent decades have certainly taught us as clinical rheumatologists that disease activity — once considered acceptable — should now be viewed as uncontrolled disease and be treated as such. This effect of "moving goalposts" has also been inferred from the data from Bykerk's study. Although Bykerk, *et al* demonstrated that flares were reported more by patients in moderate to severe disease activity than by patients in remission, flares still seem to occur in patients with RA in remission as shown by Hewlett, *et al*¹⁶. Because both studies asked the patient "whether they were experiencing a flare or not," these flares might be fundamentally different because of the different baseline level of disease activity. On a critical note, instead of debating the best flare criterion, we should perhaps first focus more on optimally treating to target in our patients¹⁷. Although the benefits of treat-to-target have been demonstrated, many patients still do not receive this level of care, as witnessed by the relatively high mean DAS28 in several large RA registries and cohort studies, including the BRASS registry, as Bykerk, *et al* also mention in their discussion⁷.

It should be appreciated that Bykerk, *et al* and the OMERACT RA flare group are studying flare thoroughly; and we share the desire to come up with valid flare criteria that can easily be used both in research as well as in daily clinical practice, because that will improve care for our patients and research alike.

AATKE van der MAAS, MD, PhD;

ALFONS A. den BROEDER, MD, PhD;

Department of Rheumatology,
Sint Maartenskliniek,
Nijmegen, The Netherlands

Address correspondence to Dr. van der Maas, Sint Maartenskliniek, Department of Rheumatology, Hengstdal 3, 6522 JV, Nijmegen, The Netherlands. E-mail: a.vandermaas@maartenskliniek.nl

REFERENCES

1. Bingham CO III, Pohl C, Woodworth TG, Hewlett SE, May JE, Rahman MU, et al. Developing a standardized definition for disease "flare" in rheumatoid arthritis (OMERACT 9 Special Interest Group). *J Rheumatol* 2009;36:2335-41.
2. 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.
3. 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.
 4. van Gestel AM, Prevoo ML, Van't Hof MA, van Rijswijk MH, van De Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum* 1996;39:34-40.
 5. van Gestel AM, Anderson JJ, van Riel PL, Boers M, Haagsma CJ, Rich B, et al. ACR and EULAR improvement criteria have comparable validity in rheumatoid arthritis trials. American College of Rheumatology European League of Associations for Rheumatology. *J Rheumatol* 1999;26:705-11.
 6. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis* 2011;70:404-13.
 7. Bykerk V, Shadick N, Frits M, Bingham C III, Jeffrey I, Iannaccone C, et al. Flares in rheumatoid arthritis: frequency and management. A report from the BRASS registry. *J Rheumatol* 2014;41:227-34.
 8. Pavelka K, Kavanaugh AF, Rubbert-Roth A, Ferraccioli G. Optimizing outcomes in rheumatoid arthritis patients with inadequate responses to disease-modifying anti-rheumatic drugs. *Rheumatology* 2012;51 Suppl 5:v12-21.
 9. Schiff M, Weinblatt ME, Valente R, van der Heijde D, Citera G, Elegbe A, et al. Head-to-head comparison of subcutaneous abatacept versus adalimumab for rheumatoid arthritis: two-year efficacy and safety findings from AMPLE trial. *Ann Rheum Dis* 2013 Aug 20 (E-pub ahead of print).
 10. Chen YF, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, et al. A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness. *Health Technol Assess* 2006;10:1-229.
 11. Welsing PM, van Gestel AM, Swinkels HL, Kiemeny LA, van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis Rheum* 2001;44:2009-17.
 12. Yoshida K, Sung YK, Kavanaugh A, Bae SC, Weinblatt ME, Kishimoto M, et al. Biologic discontinuation studies: a systematic review of methods; authors' response to van der Maas et al. *Ann Rheum Dis* 2013 Oct 23 (E-pub ahead of print).
 13. Lie E, Woodworth TG, Christensen R, Kvien TK, Bykerk V, Furst DE, et al. Validation of OMERACT preliminary rheumatoid arthritis flare domains in the NOR-DMARD study. *Ann Rheum Dis* 2013 Jul 12 (E-pub ahead of print).
 14. Bingham CO III, Alten R, Bartlett SJ, Bykerk VP, Brooks PM, Choy E, et al. Identifying preliminary domains to detect and measure rheumatoid arthritis flares: report of the OMERACT 10 RA Flare Workshop. *J Rheumatol* 2011;38:1751-8.
 15. van der Maas A, Lie E, Christensen R, Choy E, de Man YA, van Riel P, et al. Construct and criterion validity of several proposed DAS28-based rheumatoid arthritis flare criteria. *Ann Rheum Dis* 2013;72:1800-5.
 16. Hewlett S, Sanderson T, May J, Alten R, Bingham CO 3rd, Cross M, et al. 'I'm hurting, I want to kill myself': rheumatoid arthritis flare is more than a high joint count—an international patient perspective on flare where medical help is sought. *Rheumatology* 2012;51:69-76.
 17. Schoels M, Knevel R, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas DT, et al. Evidence for treating rheumatoid arthritis to target: results of a systematic literature search. *Ann Rheum Dis* 2010;69:638-43.

J Rheumatol 2014;41:189–91; doi:10.3899/jrheum.131217