

Investigating the Validity of the Minimal Disease Activity State for Patients with Rheumatoid Arthritis Treated with Abatacept

GEORGE A. WELLS, MAARTEN BOERS, TRACY LI, and PETER S. TUGWELL

ABSTRACT. *Objective.* To validate the definitions of minimal disease activity (MDA) in patients with rheumatoid arthritis (RA) and to compare abatacept to control with respect to patients attaining a state of MDA. *Methods.* Two randomized controlled trials comparing abatacept to control in patients with RA were considered: ATTAIN and AIM. Core set measures, Disease Activity Score 28-joint count (DAS28), and, for AIM, radiographic scores were available. The core set and DAS-based definitions for MDA were calculated and the number of patients in the treatment groups meeting the definitions was compared to determine sensitivity of the criteria to treatment differences and patient severity. The number of times achieving MDA was compared to the change in Health Assessment Questionnaire (HAQ), and for the AIM study compared to change in radiographic scores. *Results.* For both definitions of MDA, the change in radiographic scores showed a continual decrease in progression the more often a patient was in MDA. The change in HAQ, for both studies, showed a similar consistent improvement — the longer a patient was in MDA, then the better the HAQ score. Significantly more patients in the abatacept group met the core set and DAS-based definition of MDA than in the control group. *Conclusion.* The presence and persistence of MDA was associated with slowing of radiographic progression and improvement in the HAQ, providing support for discriminative and predictive validity of the measure. The MDA results were consistent with other efficacy analyses indicating a treatment advantage for abatacept. (J Rheumatol First Release Jan 15 2009; doi:10.3899/jrheum.080059)

Key Indexing Terms:

RHEUMATOID ARTHRITIS

VALIDATION

ABATACEPT

MINIMAL DISEASE ACTIVITY

RANDOMIZED CONTROLLED TRIAL

With recent advances in therapy for patients with rheumatoid arthritis (RA), the proportion of patients achieving a satisfactory state of disease activity is becoming a more important measure with which to compare different treatment strategies. Minimal disease activity (MDA) is a state between high disease activity and remission and, by definition, anyone in remission will also be in MDA. The need for a definition of MDA arose out of the observation that

achieving and maintaining a satisfactory state of disease activity is probably more important in the long term than the improvement from a high level of disease activity documented in trials, and that remission is not a frequent occurrence in regular clinical practice. Describing the number of patients with RA achieving and maintaining a satisfactory state of MDA for a specified period of time will add useful information for the practicing physician and aid in the interpretation of trial and longitudinal results.

The process to come to a definition of MDA consisted of 3 basic steps: consensus on a conceptual definition, development of an operational definition, and prospective validation of the definition¹. First, from the conceptual perspective, the definition of MDA was anchored to the clinical experience of the physician and personal experience of the patient. For the physician it is linked to treatment decisions and to prognosis; and for the patient it is linked to satisfaction and adaptation. At the Outcome Measures in RA Clinical Trials (OMERACT) 6 conference in 2002, the participants agreed on a conceptual definition of MDA: “that state of disease activity deemed a useful target of treatment by both the patient and the physician, given current treat-

From the Department of Epidemiology and Community Medicine and the Institute of Population Health, University of Ottawa, Ottawa, Ontario, Canada; Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam, The Netherlands; and Bristol-Myers Squibb, Princeton, New Jersey, USA.

Supported in part by an unrestricted research grant-in-aid from Bristol-Myers Squibb.

G.A. Wells, PhD, Department of Epidemiology and Community Medicine, University of Ottawa; M. Boers, MSc, MD, PhD, Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre; T. Li, PhD, Bristol-Myers Squibb; P.S. Tugwell, MD, MSc, Institute of Population Health, University of Ottawa.

Address reprint requests to Prof. G.A. Wells, Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada. E-mail: gawells@ottawaheart.ca

Accepted for publication September 9, 2008.

ment possibilities and limitations². Second, to determine an operational definition, a data-driven consensus process was required. The approach taken was the judgmental approach that gauges the opinion of patients and physicians on a useful target using patient profiles. To prepare for a preliminary operational definition of MDA for use in clinical trials, a survey of rheumatologists was conducted in which they assessed patient profiles describing real patients with RA seen in routine clinical practice. Based on their responses, several candidate definitions for MDA were designed and discussed at the OMERACT 7 conference in 2004. Feedback from participants and additional on-site analyses in a cross-sectional database allowed the formulation of 2 preliminary, equivalent definitions of MDA³: one based on the Disease Activity Score 28-joint count (DAS28) index of disease activity⁴; and one based on meeting cutpoints in 5 out of the 7 World Health Organisation (WHO)/International League of Associations for Rheumatology (ILAR) core set measures⁵.

Applying these 2 definitions, one first needs to choose whether to use the DAS28 or the core set definition, because although each selects a similar proportion in a population, these are not always the same patients. For both the core set definition and the DAS-based definition, a decision node is used to include patients with high pain or Health Assessment Questionnaire (HAQ) levels but otherwise in MDA. This initial decision node places all patients in MDA who have a tender joint count (TJC) of 0, a swollen joint count (SJC) of 0, and an erythrocyte sedimentation rate (ESR) not greater than 10 mm. If this condition is not met then for the core set definition, the core set measures are assessed and the definition places patients in MDA when they meet 5 out of 7 of the following criteria: pain (0–10)

≤ 2 ; SJC (0–28) ≤ 1 ; TJC (0–28) ≤ 1 ; HAQ (0–3) ≤ 0.5 ; physician global assessment (0–10) ≤ 1.5 ; patient global assessment (0–10) ≤ 2 ; and ESR ≤ 20 mm. For the DAS-based definition: DAS28 is calculated and the definition places patients in MDA when $\text{DAS28} \leq 2.85$. This cutpoint value of DAS28 between 2.6 for remission and 3.2 for low disease activity state was derived empirically at OMERACT 7. These 2 definitions for MDA are displayed in Figure 1. The concepts and derivation of MDA have been further described in an article on low disease activity⁶.

The third step in developing the MDA definition is to prospectively validate the definition². For this, datasets will be required to determine whether being in a state for a period of time leads to benefits in terms of functional disability and structural damage. The objective of our study was to compare abatacept to control with respect to patients attaining a state of MDA, and to validate the 2 preliminary MDA definitions by determining whether being in a state for a period of time leads to benefits in terms of functional disability as assessed by the HAQ and structural damage as assessed by radiographic progression.

MATERIALS AND METHODS

Datasets. The data from 2 double-blind, randomized controlled trials (RCT) in patients with active RA were used for this evaluation. The ATTAIN Study was a Phase III multicenter, 6-month trial evaluating the efficacy and safety of abatacept on a background of disease modifying antirheumatic drug (DMARD) therapy in patients with active RA (RA functional class I, II, or III) who were anti-tumor necrosis factor (TNF) therapy failures⁷. Eligible and consenting patients were randomized 2:1 to receive abatacept (n = 258) or placebo (n = 133) on a background of DMARD. The primary objectives were to compare abatacept to placebo regarding the clinical efficacy as assessed by the American College of Rheumatology (ACR) 20 response rate at 6 months, and the improvement in physical function as assessed by the HAQ disability index at 6 months.

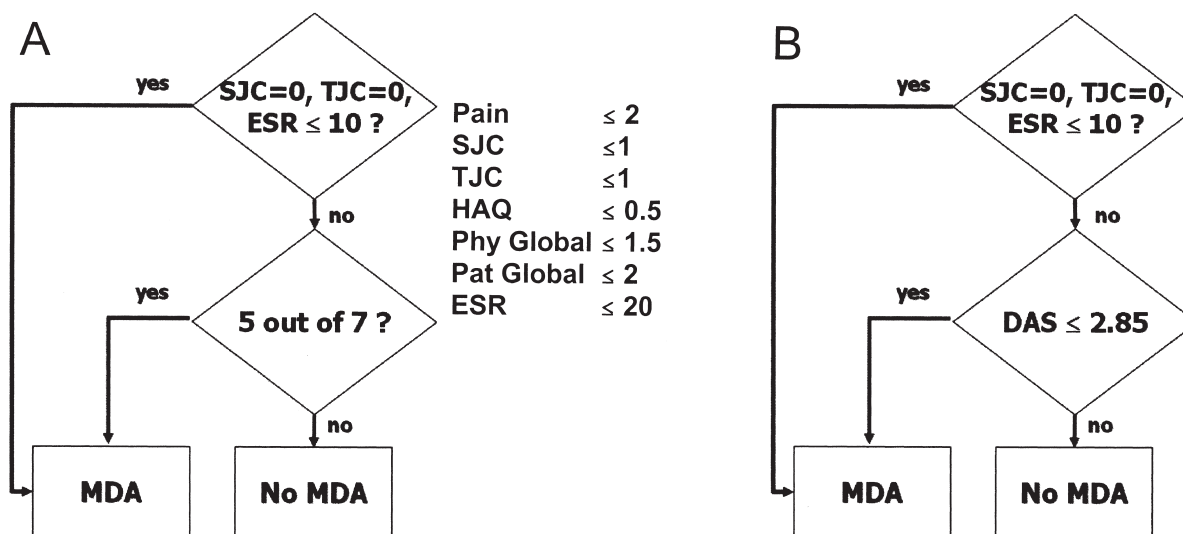


Figure 1. The core set definition (A) and the DAS-based definition (B) for minimal disease activity (MDA) for RA. DAS: Disease Activity Score; ESR: erythrocyte sedimentation rate; HAQ: Health Assessment Questionnaire; Phy Global: physician global assessment; Pat Global: patient global assessment; SJC: swollen joint count; TJC: tender joint count.

The AIM Study was a Phase III multicenter, 12-month trial evaluating the efficacy and safety of abatacept on a background of methotrexate (MTX) therapy in patients with active RA (RA functional class I, II, or III) who had an inadequate response to MTX⁸. Eligible and consenting patients were randomized 2:1 to receive abatacept (n = 433) or placebo (n = 219) on a background of MTX. The primary objectives were to compare abatacept to placebo regarding clinical signs and symptoms of RA as measured by ACR20 response following 6 months of treatment, physical function as measured by the HAQ disability index at 12 months, and radiographic progression as assessed by erosion score using the Genant-modified Sharp method at 12 months of treatment.

Outcome measures. For both RCT a number of outcomes measures were assessed at 3, 6, 9, and 12 months for the AIM trial, and at 6 and 12 months for the ATTAIN study, including the core set measures: TJC (0–68), SJC (0–68), patient assessment of disease (Pat Global; 0–100), physician assessment of disease (Phy Global; 0–100), pain assessment (Pain; 0–100), HAQ (0–3), and ESR. The DAS28 and the European League Against Rheumatism (EULAR) response criteria were calculated. In the AIM trial, radiographic progression was assessed using joint space narrowing (JSN), erosion score, and total score at 6 and 12 months.

Analysis. The percentage of patients reaching MDA was calculated based on the core set definition and the DAS-based definition of MDA. In addition to calculating this percentage at the end of study, the frequency distribution according to the month achieving MDA was assessed by treatment group for each of the studies. The relationship of radiographic progression (erosion score, JSN, total score) and functional progression (HAQ) to MDA was assessed according to the number of times in MDA (number of assessments times at 3, 6, 9, and 12 mo; 0–4). An analysis of variance (ANOVA) was used to compare the radiographic and functional progression over the number of times in MDA. In addition to the overall F-test in the ANOVA, Tukey's honest significance difference for all pairwise comparisons was considered.

RESULTS

The patients with RA included in our analysis had an average age in their early 50s and were mostly female and Caucasian (Table 1). Disease duration varied between the 2 studies included in the analysis, with the patients in the ATTAIN study having a duration of disease approaching 12 years and those in the AIM study being under 9 years in duration.

For both the ATTAIN and AIM studies, a significantly greater number of patients in the abatacept group met the core set and DAS-based definition of MDA than the corresponding control group. For the ATTAIN study, 10.6% in the abatacept group compared to 3.1% in the control group (p = 0.0097) met the MDA core set definition. An even greater difference was found for the DAS-based definition, with 12.6% in the abatacept group compared to 1.8% in the control group (p = 0.0007) meeting the definition. Similarly, for the AIM study, 29.0% versus 9.2% (p < 0.0001) and 21.9% versus 2.4% (p < 0.0001) met the 2 definitions of MDA. The percentage of patients being in MDA by the length of time on treatment for the ATTAIN study was 6.8% versus 1.6%, and 10.6% versus 3.1% at 3 and 6 months of treatment, respectively, for the core set definition (only the 6-mo assessment available for the DAS-based definition). Patients in MDA by length of time on treatment for the AIM study (Figure 2) indicated that for both the core set and DAS-based definitions, the difference favoring abatacept occurred at the early assessments and the difference to the control became more pronounced for the later assessment times.

Radiographic scores were measured only in the AIM study. In general, the longer a patient was in MDA then the better the radiographic scores (Table 2). In particular, there was a continual decrease in the progression of the total score for patients as the number of times in MDA increased, with the change in the total score of 1.68, 1.67, 1.34, 0.80, and 0.49 for patients that were in MDA 0, 1, 2, 3, and 4 times, respectively, according to the core set definition for MDA. The erosion scores followed this same pattern, but not the JSN scores. For the DAS-based definition, fewer times (i.e., 0, 1, 2) were available, but the pattern was similar over the number of times in MDA. Although these trends were strong, they were not statistically significant.

Functional assessment as measured by the HAQ followed a similar pattern to that for radiographic progression: the

Table 1. Characteristics of patients in the AIM and ATTAIN studies.

Study Group	Age*, mean (SD)	Patient Characteristic			RA Duration*, mean (SD)
		Sex, % female	Race, % Caucasian		
ATTAIN					
Abatacept, n = 258	53 (12.4)	77	96		12 (8.5)
Control, n = 133	53 (11.3)	80	93		11 (8.9)
Both Groups, n = 391	53 (12.1)	78	95		12 (8.6)
AIM					
Abatacept, n = 433	52 (12.9)	78	88		9 (7.3)
Control, n = 219	50 (12.4)	82	88		9 (7.1)
Both Groups, n = 652	51 (12.7)	79	88		9 (7.2)
Both Studies					
Abatacept, n = 691	52 (12.8)	78	91		10 (7.9)
Control, n = 352	51 (12.0)	81	90		10 (7.9)
Both Groups, n = 1043	52 (12.5)	79	91		10 (7.9)

* in years

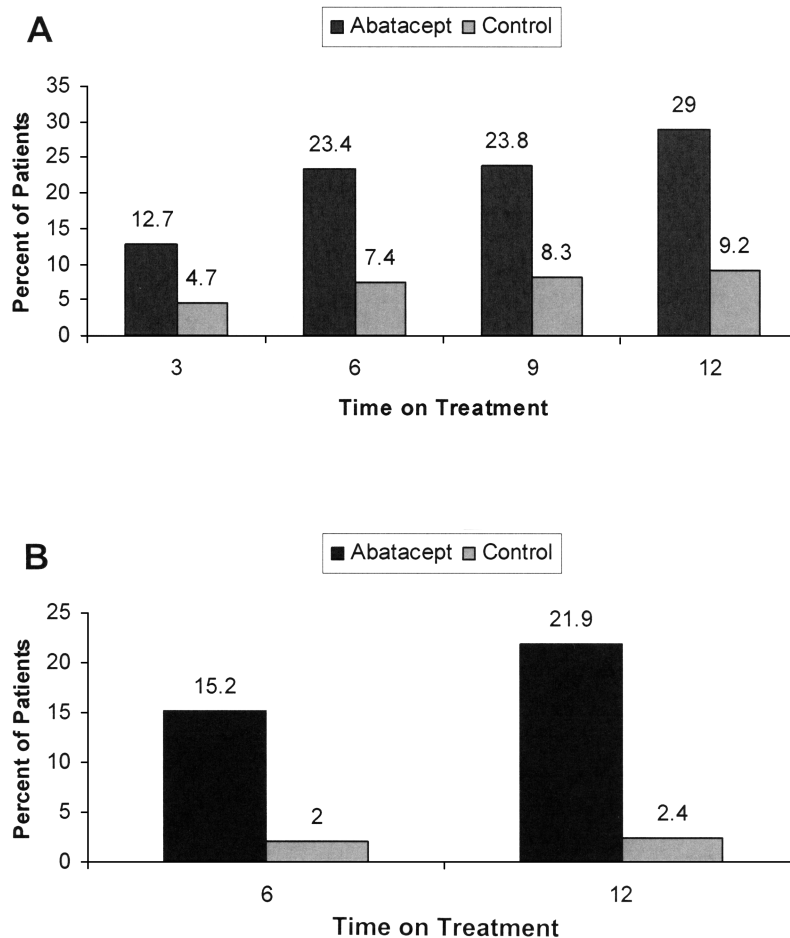


Figure 2. Patients in the AIM study in minimal disease activity (MDA) by time on treatment and treatment group: abatacept versus control. A. Core set definitions of MDA; B. DAS-based definition of MDA.

longer a patient was in MDA, the better the HAQ score (Table 2). For the AIM study the change in the HAQ score was -0.39 , -0.74 , -0.83 , -1.18 , and -1.17 ($p < 0.0001$) for patients in MDA 0, 1, 2, 3, and 4 times, respectively, according to the core set definition for MDA. A similar pattern was observed for the DAS-based definition (for number of times in MDA 0, 1, 2) and for both definitions in the ATTAIn study. These patterns were significant and, in addition, pairwise significance occurred when patients who were categorized the higher number of times in MDA were compared to those patients never in MDA.

Only one patient was reclassified as in MDA if the initial node was not included. The patient, a Caucasian woman in her mid-50s, was enrolled in the AIM study and met the core set and DAS-based definitions for MDA at only the 12-month assessment. Her core set profile was pain 42, SJC 0, TJC 0, HAQ 1.25, Phy Global 22, Pat Global 31, and ESR 6. She was classified as being in MDA since both her SJC and TJC were 0 and $ESR \leq 10$. When the initial node was not included, she did not meet the core set definition of MDA since only 3 of the 7 core set measures met the thresh-

old (Figure 1). For the DAS-based definition, she was classified as being in MDA with or without the initial node since DAS28 was 1.56 and met the threshold of 2.85 for MDA (Figure 1). Although the exclusion of the initial node had only a minor effect on the classification of patients in the AIM and ATTAIn studies, it did properly include a patient with chronic pain syndrome and low disease activity as being in MDA.

DISCUSSION

The objective of the OMERACT 7 module for MDA was to seek consensus on a definition of MDA that could be recommended as a secondary endpoint in RCT. Two definitions for MDA were derived: the core set definition and the DAS-based definition³. As part of the research agenda, the third step in developing the MDA definition was to prospectively validate the definition in high-quality datasets. The AIM and the ATTAIn were both well defined and properly conducted double-blind RCT in patients with active RA, yielding high-quality data in which to assess MDA.

The results for MDA mirrored the clinical results of the

Table 2. Radiographic and functional progression by length of time being in MDA: mean (standard deviation) (sample size).

	Number of times in MDA					p
	0	1	2	3	4	
AIM Study						
Core set definition						
Radiographic progression						
Erosion score	0.86 (2.45) (n = 413)	0.75 (1.51) (n = 80)	0.74 (1.31) (n = 45)	0.40 (1.44) (n = 37)	0.20 (0.94) (n = 40)	0.3221
JSN score	0.51 (0.50) (n = 459)	0.46 (0.50) (n = 84)	0.57 (0.50) (n = 49)	0.56 (0.50) (n = 41)	0.73 (0.45) (n = 40)	0.0748
Total score	1.68 (4.22) (n = 413)	1.67 (2.97) (n = 80)	1.34 (2.41) (n = 45)	0.80 (2.20) (n = 37)	0.49 (1.90) (n = 40)	0.2571
Functional progression						
HAQ	-0.39 (0.58) (n = 443)	-0.79 (0.64) (n = 82)	-0.83 (0.63) (n = 48)	-1.18 (0.66) (n = 39)	-1.17 (0.61) (n = 37)	< 0.0001 0 vs 1,2,3,4 1 vs 3,4
DAS-based definition						
Radiographic progression						
Erosion score	0.85 (2.32) (n = 482)	0.39 (1.41) (n = 72)	0.36 (1.00) (n = 40)			0.1136
JSN score	0.51 (0.50) (n = 519)	0.64 (0.48) (n = 78)	0.50 (0.51) (n = 46)			0.0874
Total score	1.71 (4.07) (n = 482)	0.79 (2.44) (n = 72)	0.78 (1.99) (n = 40)			0.0695
Functional progression						
HAQ	-0.47 (0.63) (n = 503)	-0.95 (0.55) (n = 75)	-1.05 (0.64) (n = 42)			< 0.0001 0 vs 1,2
ATTAIN Study						
Core set definition						
Functional progression						
HAQ	-0.26 (0.48) (n = 344)	-0.96 (0.58) (n = 24)	-1.23 (0.67) (n = 13)	—	—	< 0.0001 0 vs 1,2
DAS-based definition						
Functional progression						
HAQ	-0.29 (0.50) (n = 287)	-0.92 (0.72) (n = 28)		—	—	< 0.0001

MDA: minimal disease activity; JSN: joint space narrowing; DAS: disease activity score; HAQ: health assessment questionnaire.

abatacept studies, which demonstrated positive and sustained response for the abatacept-treated group in clinical endpoints, physical function, quality of life (physical and mental), and various patient-reported outcomes of fatigue, sleep quality and activity limitation. This confluence of results provides a level of face validity for MDA.

The ability of MDA to predict something it should theoretically be able to predict (predictive validity) and to be different from something that it should theoretically be dissimilar to (discriminative validity) was assessed by considering the length of time patients were in MDA with respect to 2 important outcomes, radiographic progression and functional progression. The time in MDA was determined according to the number of assessment times over the course of the study that the patient was determined to be in MDA (number of assessment times at 3, 6, 9, and 12 mo; 0–4). It was found that the more often patients were in MDA then the better were their radiographic progression and functional

progression. The results were consistent for the 2 patient populations considered in the studies: subjects with an inadequate response to MTX and subjects with an inadequate response to TNF therapies.

During the development of the definitions for MDA, a concern was raised that patients with chronic pain syndrome but low disease activity would be misclassified as having high disease activity due to high scores in pain, TJC, and Pat Global. To address this issue, an initial decision is made to classify patients as being in MDA if they have high pain or HAQ levels, but are otherwise in MDA. As part of the prospective validation, the importance of this decision node was to be evaluated. The sensitivity analysis indicated that this decision node in the definitions did not lead to substantive differences in the results. However, this result was not expected since the patient groups for both the AIM and ATTAIN studies had high disease activity and the number of patients with chronic pain syndrome but low disease activi-

ty would be exceptional.

Attempting to define MDA is not a new venture. The DAS28 has identified cutpoints for remission (< 2.6) and low disease activity (≤ 3.2) and the ACR response criteria attempt to provide levels of response by using the specific ACR20, ACR50, and ACR70 criteria. The DAS-based definition for MDA is closely aligned with DAS28 criteria. In the empirical work and consensus-building process completed for the OMERACT 7 conference, the optimal cutpoint for MDA was found to be 2.85 for the DAS28, compared to the current cutpoint of 3.2 for low disease activity^{3,6}. This new lower cutpoint may reflect a greater refinement in deriving the cutpoint or greater expectations for treatment to deliver better states of disease activity, as the new cutpoint is closer to the DAS28 remission of 2.6. The core set definition is aligned with the ACR response criteria by involving the same measures and a construction of a similar nature requiring 5 of 7 measures to satisfy specified thresholds. Since the ACR criteria for specific states has not been formally developed there is no direct comparator of our work here, and the core set definition for MDA is unique when working with the core set measures. The simultaneous development of MDA corresponding to our most important response criteria provides a common ground on which both measures can be used to provide insights into minimal disease activity for users of either response criteria. Further, the initial decision that must be made regarding chronic pain syndrome offers an important step in avoiding misclassification of these patients. The choice of the tree structure for the MDA allows for other decisions of this type to be incorporated as needed to improve classification of patients as being in MDA.

After deriving the MDA core set and DAS-based definitions, the next step in the planned development was to prospectively validate the definitions. The objective at this time was not to choose one of the definitions. Even if this were the objective, there is very little difference in the results on which to base such a decision. The 2 definitions were found to have discriminative and predictive validity, and had similar results when assessed in terms of functional disability and structural damage. The initial decision node was only needed once, for the core set definition, to properly classify a patient who was in MDA but had high pain and HAQ levels. Just as we have 2 sets of responder outcome measures (i.e., the WHO/ILAR core set corresponding to

ACR response criteria and the DAS28 corresponding to EULAR response criteria), both MDA definitions should continue to be evaluated until appropriate accrual of empirical evidence is obtained to make a decision. One important consideration in using the core set definition is that it provides users of the ACR response criteria with a measure of state, which is not available, unlike the EULAR response criteria. Although MDA cannot be recommended as the primary outcome in a study at this time, with the results of this analysis and the importance of MDA state, it is recommended that MDA be considered a major outcome in a clinical trial for patients with RA.

The presence and persistence of MDA was associated with slowing of radiographic progression and improvement in the function, indicating discriminative and predictive validity. This is the first study in patients treated with a biologic therapy that has prospectively validated the definition of MDA as prescribed by OMERACT 7.

REFERENCES

1. Boers M, Anderson J, Felson D. Deriving an operational definition of low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1112-4.
2. Wells G, Anderson J, Boers M, et al. MCID/low disease activity state workshop: summary, recommendations, and research agenda. *J Rheumatol* 2003;30:1115-8.
3. Wells GA, Boers M, Shea B, et al. Minimal disease activity for rheumatoid arthritis: a preliminary definition. *J Rheumatol* 2005;32:2016-24.
4. Van Gestel A, Prevoo M, van 't Hof MA, van Ruswijk M, van de Putte L, van Riel P. 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. Felson D, Anderson J, Boers M, et al. American College of Rheumatology, preliminary definition of improvement in rheumatoid arthritis. *Arthritis Rheum* 1995;38:727-35.
6. Wells G, Boers M, Tugwell P. Low disease activity state in rheumatoid arthritis: concepts and derivation of minimal disease activity. *Clin Exp Rheumatol* 2006;24:S052-9.
7. Genovese MC, Becker JC, Schiff M, et al. Abatacept for rheumatoid arthritis refractory to tumor necrosis factor a inhibition. *N Engl J Med* 2005;353:1114-23.
8. Kremer JM, Genant HK, Moreland LW, et al. Effects of abatacept in patients with methotrexate-resistant active rheumatoid arthritis: a randomized trial. *Ann Intern Med* 2006;144:865-76.