

Minimal Disease Activity for Rheumatoid Arthritis: A Preliminary Definition

GEORGE A. WELLS, MAARTEN BOERS, BEVERLEY SHEA, PETER M. BROOKS, LEE S. SIMON, C. VIBEKE STRAND, DANIEL ALETAHA, JENNIFER J. ANDERSON, CLAIRE BOMBARDIER, MAXIME DOUGADOS, PAUL EMERY, DAVID T. FELSON, JAAP FRANSEN, DAN E. FURST, JOHANNA M.W. HAZES, KENT R. JOHNSON, JOHN R. KIRWAN, ROBERT B.M. LANDEWÉ, MARISSA N.D. LASSERE, KALEB MICHAUD, MARIA SUAREZ-ALMAZOR, ALAN J. SILMAN, JOSEF S. SMOLEN, DESIREE M.F.M. van der HEIJDE, PIET L.C.M. van RIEL, FRED WOLFE, and PETER S. TUGWELL

ABSTRACT. Agreement on response criteria in rheumatoid arthritis (RA) has allowed better standardization and interpretation of clinical trial reports. With recent advances in therapy, the proportion of patients achieving a satisfactory state of minimal disease activity (MDA) is becoming a more important measure with which to compare different treatment strategies. The threshold for MDA is between high disease activity and remission and, by definition, anyone in remission will also be in MDA. True remission is still rare in RA; in addition, the American College of Rheumatology definition is difficult to apply in the context of trials. Participants at OMERACT 6 in 2002 agreed on a conceptual definition of minimal disease activity (MDA): “that state of disease activity deemed a useful target of treatment by both the patient and the physician, given current treatment possibilities and limitations.” To prepare for a preliminary operational definition of MDA for use in clinical trials, we asked rheumatologists to assess 60 patient profiles describing real RA patients seen in routine clinical practice. Based on their responses, several candidate definitions for MDA were designed and discussed at the OMERACT 7 in 2004. Feedback from participants and additional on-site analyses in a cross-sectional database allowed the formulation of 2 preliminary, equivalent definitions of MDA: one based on the Disease Activity Score 28 (DAS28) index, and one based on meeting cutpoints in 5 out of the 7 WHO/ILAR core set measures. Researchers applying these definitions first need 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. In both MDA definitions, an initial decision node places all patients in MDA who have a tender joint count of 0 and a swollen joint count of 0, and an erythrocyte sedimentation rate (ESR) no greater than 10 mm. If this condition is not met:

- The DAS28 definition places patients in MDA when $DAS28 \leq 2.85$
- The core set definition places patients in MDA when they meet 5 of 7 criteria: (1) Pain (0–10) ≤ 2 ; (2) Swollen joint count (0–28) ≤ 1 ; (3) Tender joint count (0–28) ≤ 1 ; (4) Health Assessment Questionnaire (HAQ, 0–3) ≤ 0.5 ; (5) Physician global assessment of disease activity (0–10) ≤ 1.5 ; (6) Patient global assessment of disease activity (0–10) ≤ 2 ; (7) ESR ≤ 20 . This set of 2 definitions gained approval of 73% of the attendees. These (and other) definitions will now be subject to further validation in other databases. (J Rheumatol 2005;32:2016–24)

Key Indexing Terms:

MINIMAL DISEASE ACTIVITY
OUTCOME MEASURES

CLINICAL TRIALS

RHEUMATOID ARTHRITIS
SURVEY

From the Department of Epidemiology and Community Medicine, University of Ottawa; Institution of Population Health, Ottawa, Canada; Medical University Vienna, Vienna, Austria; Boston University School of Public Health, Boston, USA; Institution of Work and Health, University of Toronto, Toronto, Canada; University of Queensland, Royal Brisbane Hospital, Brisbane, Australia; Hopital Cochin, Rene Descartes University, Paris, France; University of Leeds, Leeds, UK; School of Medicine, Boston, USA; Department of Rheumatology, University Hospital Nijmegen; David Geffen School of Medicine, UCLA, Los Angeles, USA; Erasmus University Medical Center, Rotterdam, The Netherlands; Mater Hospital, University of Newcastle, Newcastle, Australia; University of Bristol, Bristol Royal Infirmary, Bristol, UK; University Hospital Maastricht, Maastricht, The Netherlands; Department of Rheumatology, St. George Hospital, Sydney, Australia; National Data Bank for Rheumatic Diseases, Wichita, USA; Baylor College of Medicine and Michael E. De Bakey VAMC, Houston, USA; University of Manchester,

Manchester, UK; Harvard Medical School, Beth Israel Deaconess Medical Center, Boston; and Stanford University School of Medicine, Stanford, USA;

G.A. Wells, PhD, Department of Epidemiology and Community Medicine, University of Ottawa; M. Boers, MD, PhD, Department of Clinical Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam; B. Shea, PhD Candidate, Institute of Population Health, University of Ottawa; D. Aletaha, MD, Medical University Vienna, Division of Rheumatology, Vienna General Hospital; J.J. Anderson, PhD, Boston University School of Public Health; C. Bombardier, MD, University of Toronto, Institute for Work and Health; P.M. Brooks, MD, University of Queensland, Royal Brisbane Hospital; M. Dougados, MD, Hopital Cochin, Rene Descartes University, Department of Rheumatology; P. Emery, MD, Rheumatology, University of Leeds; D.T. Felson, MD, School of Medicine, Boston University; J. Fransen, MD, Department of Rheumatology, University Hospital Nijmegen; D.E. Furst,

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2005. All rights reserved.

MD, David Geffen School of Medicine, UCLA; J.M.W. Hazes, MD, Erasmus University Medical Center; K.R. Johnson, MD, Clinical Pharmacology, Mater Hospital, University of Newcastle; J.R. Kirwan, BSc, MD, Academic Rheumatology Unit, University of Bristol, Bristol Royal Infirmary; R.B.M. Landewé, MD, University Hospital Maastricht; M.N.D. Lassere, MB, PhD, Department of Rheumatology, St. George Hospital; K. Michaud, MSc, National Data Bank for Rheumatic Diseases; M. Suarez-Almazor, MD, PhD, Baylor College of Medicine, Michael E. De Bakey VAMC; A.J. Silman, MD, University of Manchester, ARC Epidemiology Unit; L. Simon, MD, Harvard Medical School, Beth Israel Deaconess Medical Center; J.S. Smolen, MD, Medical University Vienna, Division of Rheumatology, Vienna General Hospital; C.V. Strand, MD, Stanford University School of Medicine; D.M.F.M. van der Heijde, MD, PhD, University Hospital Maastricht; P.L.C.M. van Riel, MD, Department of Rheumatology, University Hospital Nijmegen; F. Wolfe, MD, National Data Bank for Rheumatic Diseases; P.S. Tugwell, MD, MSc, Institute of Population Health, University of Ottawa.

Address reprint requests to Dr. G. Wells, Department of Epidemiology and Community Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario K1H 8M5, Canada.

The threshold for minimal disease activity (MDA) is between high disease activity and remission and, by definition, anyone in remission will also be in MDA. In this context, we define remission conceptually as “absence of disease activity.” The need for a definition of MDA for patients with rheumatoid arthritis (RA) 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 remission is not a frequent occurrence in regular clinical practice. Describing the number of patients achieving and maintaining 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.

Any definition of MDA should be a compromise that best reflects the opinion of patients and physicians. The process to come to such a definition consists of 3 basic steps: conceptual definition, operational definition, and prospective validation.

(1) From the conceptual perspective, the definition of MDA is 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; for the patient it is linked to satisfaction and adaptation. One definition suggested for MDA is: that state deemed a “useful target” of treatment by both the physician and patient given current treatments and knowledge.

(2) To determine an operational definition, a data-driven consensus process is required, and 2 fundamental approaches can be taken: the judgmental approach that gauges the opinion of patients and physicians on a useful target using methods such as direct questioning, patient profiles, physician submitted cases, and direct observation of clinical practice; or the statistical approach that considers the range of states obtained using the judgmental approach applied in existing datasets to determine which best distinguishes a weak from a strong treatment.

(3) To prospectively validate the definition, longitudinal

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.

Important limitations and caveats of this work need to be listed: First, in this initial phase of development the use of any MDA definition should be limited to research settings (trials and observational studies). A definition for use in the clinic is an eminently worthy goal, and could follow when the preliminary definition has been shown to be valid in different settings. Second, what constitutes “a useful target of therapy” is likely a moving target, so any MDA definition should be regularly reviewed. Third, the MDA concept is linked to the concept of remission. By definition, anyone in remission will also be below the threshold of MDA. In this article the distinction between MDA and remission is not the object of study. It should be noted that current definitions of remission have their own problems and are also undergoing review. For example, in the American College of Rheumatology (ACR) definition one needs to meet 5 of the 6 criteria, so patients meeting this definition can have signs of disease activity in the remaining measure (e.g., high swollen joint counts). Also, any definition based on reduced joint counts can be shown to include patients that have significant remaining disease activity, especially in the ankle and feet joints that are not assessed. Finally, the definition of MDA is designed to be a secondary endpoint in studies. Like remission or the ACR70 response, it will not be the most efficient way to discriminate between treatment groups.

Work to define MDA has been going on for more than 4 years. The original term was Low Disease Activity State (LDAS), and the various OMERACT and ACR meetings, surveys, and presentations used the term LDAS. Over time, it became apparent that LDAS gave the impression of referring to a “low” state of activity and excluded remission. The change to MDA was, in part, to address this misconception.

The background work for MDA began with the OMERACT 6 meeting in 2002, Brisbane, Australia. The objective of the OMERACT 6 LDAS Workshop was to meet many challenges that exist in determining minimal disease activity by reviewing concepts and terminologies and deciding on a process for developing an operational definition¹⁻³. At OMERACT 6, the workshop had 4 breakout groups. One of these groups comprised patients who attended the conference. In the Patient Perspective group, patient concerns were critically reviewed and discussed with the goal of ensuring that any definition of MDA would take into consideration the patient perspective and ultimately be acceptable to patients. The final voting supported the development of a research agenda for measuring sleep and fatigue outcomes that were important to the patients so that these could be considered in the definition of MDA. The methods group was concerned with the methods and consensus process for developing an operational definition. A wide range of possible judgmental and statistical approaches were discussed,

with the goal of developing a comprehensive methodological strategy to be implemented for the development of an operational definition of MDA. The voting supported both an opinion-based approach (judgmental) and an observational-based approach (statistical). The candidate measures group reviewed the core measures used in indexes such as ACR20 (American College of Rheumatology) and DAS (Disease Activity Score), and added (e.g., fatigue) and subtracted measures as needed with the goal of deriving a comprehensive and parsimonious list of candidate measures for use in a definition of MDA. The voting supported a comprehensive list of outcomes for assessing pain, function, inflammation, health related quality of life, structure/damage, and toxicity and comorbidity for consideration in the definition of MDA. The definition formulation group focused on the levels and combinations of the measures considered (assuming measures used in the definition were given) with the goal of providing examples of definitions of MDA that have face validity. Approaches for formulating MDA supported by the voting included a weighted approach (i.e., outcome measures are weighted and aggregated, often using an equation), unweighted (i.e., a cutpoint is defined for each outcome measure and number of measures satisfying the cutpoint is counted), and tree approach (i.e., a step-by-step path through the outcome measures constituting the definition, with branching at any conditional point). Conference participants agreed on a research agenda, and a plan was formulated with different phases to be designed, implemented, and conducted over the following 2 years.

The longer term objective, to be fulfilled at OMERACT 7, was to seek consensus on a definition of MDA that could be recommended as a secondary endpoint in randomized clinical trials and could be further validated in other datasets and longterm outcome databases.

MATERIALS AND METHODS

A 3-step process was followed to develop and gain consensus on the definition of MDA. (1) At an MDA discussion group convened at the ACR meeting in October 2003, agreement was reached on candidate measures to consider in the initial definition of MDA, and options for opinion-based questions and design issues on surveying stakeholders on possible operational definitions of MDA were considered. (2) Based on these discussions a survey was designed and conducted among stakeholders between January and April 2004, to derive a limited set of possible definitions for MDA. (3) At the OMERACT 7 LDAS Module, participants were presented with this limited set of candidate definitions to discuss and from which to choose an agreed definition.

ACR meeting. The objectives of the MDA session at the ACR meeting were: to review the fundamental concepts associated with MDA, to obtain consensus on the candidate measures to be considered in the definition of MDA, to consider options for an opinion-based survey using direct (profiles of measures) or indirect (individual measures) procedures for determining MDA, and to consider design issues on surveying stakeholders on possible operational definitions of MDA.

The meeting process consisted of a slide presentation summarizing the work of the OMERACT 6 workshop and the tasks that had been accomplished since the workshop. A series of questions on key issues associated with the next steps in the development of MDA were posed during the pres-

entation and discussed by the meeting participants. The goal was to help design the "Survey of Stakeholders," which would be conducted in order to derive a limited set of possible definitions for MDA for consideration at the LDAS Module of OMERACT 7.

Survey of stakeholders. The sampling frame and sampling methodology for the survey were discussed at the ACR meeting. It was determined to survey attendees of that meeting, previous chairs and co-chairs of the OMERACT Minimal Clinically Important Difference (MCID)/LDAS modules and workshops, key research and opinion leaders, and others that these individuals identified. Further, for the opinion-based questions, the general stem-and-leaf format of the profiles was determined. Over the next 2 months the lists of those to be surveyed were assembled and the questionnaire was designed and tested. In addition, the questionnaire was posted on the OMERACT 7 conference website and participants were invited to complete the questionnaire.

The survey questionnaire consisted of 60 profiles. The profiles used measures taken from the core set⁴ to describe patients with various states of disease activity. The examples were of real patients with rheumatoid arthritis (RA), selected from the Rheumatoid Arthritis Evaluation Survey (RAES) database⁵. This database contains the results of a cross-sectional survey of disease activity in 730 consecutive RA patients attending 40 clinics in the US and Canada. The data presented for each patient was unaltered from the database record. The profiles were selected to encompass the full range of disease activity present in the dataset, and were enriched with profiles with physician global assessments between 1 and 3 (range 0–10). Surveyed stakeholders were instructed to consider the same setting for each profile. That is, to consider that the profile corresponded to a patient with RA started on methotrexate that had been increased to the dose usually used by the stakeholder. The profile described the disease activity after at least 6 months of therapy at that dose. The core measures provided in the profile were as follows (note that for all measures, "better" is indicated by a lower score):

Pain: visual analog scale (VAS): 0 to 10

Swollen joint count: 0 to 28

Tender joint count: 0 to 28

Physical function/HAQ: 0 to 3

Physician global assessment, VAS: 0 to 10

Patient global assessment, VAS: 0 to 10

Acute phase protein/ESR, mm/h: 0 to 120

For each profile, the question to be answered was: "Is the patient described in the profile in MDA?" using the definition agreed on at OMERACT 6 and reinforced at the ACR meeting:

MDA is that state which is deemed a useful target of treatment by both physician and patient, given current treatment possibilities and limitations

A typical profile is given in Figure 1. A 2-step process was suggested for "scoring" the profiles, and the instructions provided to those surveyed were as follows: (1) For each profile, consider the "result column" and the "% of max" column for each core measure, and indicate (with an "x") whether you think this patient is in MDA. (2) When you have completed scoring in the first step, go back to the profiles you scored as being in MDA and consider the "If yes" part of the question and indicate how much any single measure could increase ("highest result tolerated"), given the others stay the same, before MDA would be lost in your opinion.

Two aspects were considered to derive a definition for MDA: determination of a "cutpoint," which consisted of a maximum value for each of the core measures; and consideration of the count of core measure results that must not exceed the cutpoint in order for the patient to be in MDA. Cutpoints were derived as follows. If $\geq 80\%$ of respondents classified the profile as MDA, then the profile was considered to correspond to a patient in MDA. All other profiles were not in MDA by definition. In the set of MDA profiles summary statistics were calculated for each core set measure. Seven potential cutpoints for the core set were derived from these statistics, based on (for each measure): the mean, the rounded mean, the upper 95% confidence limit, the rounded upper 95% confidence limit, the maxi-

Measure (range)	Result	Percent of Max	Highest Result Tolerated (all else equal)	LDAS?*	
				Yes	No
Pain (0-10)	2	20			
Swollen joints (0-28)	0	0			
Tender joints (0-28)	0	0			
HAQ (0-3)	0.3	10			
Physician global (0-10)	1	10			
Patient global (0-10)	3	30			
ESR (0-120)	4	3			

* LDAS has now been replaced by MDA

Figure 1. Example of a patient profile used in the survey on minimal disease activity (MDA). LDAS: low disease activity state.

mum, the mean of highest tolerated value for each core measure, and the rounded mean of highest tolerated value. For example the “upper 95% confidence interval cutpoint” would be the 7 numbers corresponding to the upper 95% confidence limit for each measure. Considering the count of core measures that must meet (i.e., have a result no higher than) their individual cutpoint in order for the patient to be in MDA, 7 variations are obtained: 7/7, 6/7, 5/7, 4/7, 3/7, 2/7, and 1/7, where n/7 indicates that n or more of the core measures have a result at or below the cutpoint. This procedure generated 49 possible candidate definitions for MDA when the 7 possible statistics that could be used in defining a cutpoint were combined with the 7 variations in the count of the core measures that could be used to meet the cutpoint.

For each definition sensitivity and specificity were calculated with the $\geq 80\%$ respondent criterion as standard for MDA. It is noted that as the value for the cutpoint in the definition decreases (i.e., becomes more strict, less disease activity), and likewise as the count of core measures (n) to be satisfied in the definition increases, sensitivity will decrease and specificity will increase.

OMERACT 7 Module. In the opening module plenary, the goal of the module and concepts associated with MDA were reviewed, results of the survey of stakeholders were presented and the charge to the breakout groups was made. Following the plenary, the conference participants divided into 10 breakout groups (each group consisted of 10 to 20 participants with a chair and rapporteur). During the breakout session, each breakout group reviewed the definitions of MDA with 2 tasks in mind: (1) Consider and discuss the operational definitions of MDA determined from the results of the survey of stakeholders and the comfort level with each definition as an initial definition for MDA; and (2) consider a set of 10 profiles with respect to each of the candidate definitions. Each breakout group generated a report from their session and the rapporteur for each group reported back in the second module plenary. In reporting back, the tasks were to describe the process that was followed, provide a summary of the discussions, list the key concerns and issues raised, and provide a ranking of the candidate definitions. The reports of the breakout groups generated specific issues that needed to be addressed by the MDA working group. This was greatly facilitated by additional analyses on site using the RAES database. Responses to these issues were prepared and 2 operational definitions of MDA were presented and voted on at the conference plenary.

RESULTS

ACR meeting. Specific decisions regarding the definition of MDA were made at the ACR meeting. In particular, although the list from OMERACT 6 was more comprehensive, it was determined that for the initial definition of MDA

only the core measures would be included. If “other” measures were included, then this would force a redefinition of disease activity, a process that could take several years. Also, some measures (e.g., health-related quality of life) were thought to be different dimensions of “burden of disease” that were relevant to treatment but only loosely bound to the concept of disease activity. In summary, more data and consensus building were needed for other measures to be included. Also, it was decided that until patient-specific outcomes (such as sleep and fatigue) could be properly measured, the candidate variables should be limited to the core measures. This agenda is currently being executed by the study group “Patient Perspective in Outcome Assessment”^{6,7}.

The different approaches for deriving an operational definition were discussed. It was believed that the opinion-based approach involving a survey method and/or Delphi process would be more timely and feasible than an observation-based approach involving the analysis of existing data and inferring MDA from a proxy variable, such as a clinician’s decision to reduce/not increase drug treatment. Further, it was determined that a direct procedure (i.e., having respondents assess descriptions of patients using profiles that provide results of all the core set measures) was better than an indirect procedure (i.e., polling for desired levels for each core set measure separately).

The sampling frame and sampling methodology for the survey were discussed. Although different sampling methodologies for surveying groups were considered (including simple random sampling, stratified random sampling), it was believed that a non-random sampling targeting at key opinion leaders and OMERACT participants would be the initial approach. Names of key decision makers and groups that should be surveyed were suggested, and others were forwarded to the module organizers by the meeting participants. Concerns about the length and format of the survey questionnaire were expressed, but it was noted that a wide range of profiles would be needed.

Survey of stakeholders. The 60 profiles were completed by 38 respondents. There was considerable consensus among the respondents on the profiles that were felt to represent patients in MDA. There was absolute agreement on 10 profiles: these were considered to be in MDA by all the respondents. Lowering the threshold of agreement yielded more profiles considered to be in MDA: there was $\geq 90\%$, $\geq 80\%$, and $\geq 70\%$ agreement that 15, 17, and 22 profiles, respectively, were representative of patients in MDA. For those profiles in MDA, summary statistics were calculated (Table 1) and cutpoints determined (i.e., values such that if the core measure for the profile did not exceed the corresponding cutpoint then the patient was considered to be in MDA) for the core measures. The 80% consensus was selected as a reasonable compromise between consensus and a sufficient number of observations of cases in MDA, so the corresponding 17 profiles were taken as being in MDA and the other 43 profiles were considered not to be in MDA. Note that cutpoint levels were not substantially different when the $\geq 90\%$ and $\geq 70\%$ levels of agreement were used (Table 1).

Of the summary statistics of the 17 MDA profiles, the values for the core measures for the cutpoints corresponding to the mean and rounded mean of the highest tolerated value for each core measure were very high, lacked face validity, and were not considered further. The remaining 5 cutpoint

levels are given in Table 2. Application of the 7 variations for each cutpoint based on the count of measures at or below their maximum yielded 35 definitions to test for sensitivity and specificity against the $\geq 80\%$ criterion. Receiver operating characteristic curves showed generally high accuracy for all definitions (Figure 2).

A closer inspection of the critical upper left quadrant (Figure 3) revealed interesting trends. The cutpoints based on the maximum values with 7/7 criteria (termed Definition C, see below) had the best combination of sensitivity (100%) and specificity (97%). The next best combination was for the cutpoints based on the mean values with 3/7 criteria, which had a sensitivity and specificity of 100% and 86%, respectively. However, to be of practical use the rounded values were used, and the corresponding cutpoints for the rounded mean with 4/7 criteria (Definition A, see below) had a sensitivity and specificity of 88% and 79%, respectively. The cutpoints based on the upper 95% limit with 5/7 criteria had a sensitivity of 83% and specificity of 95%. Again, the rounded version of these values would be of more use in practice and the rounded upper 95% limit with 5/7 criteria (Definition B, see below) had the best combination of sensitivity (88%) and specificity (82%) for this choice. Finally, for the DAS28, the best choices for the cutpoints were 2.50 with a sensitivity and specificity of 88% and 92%, respectively, and 2.85 with a sensitivity and speci-

Table 1. Average value of each core measure for profiles selected as minimal disease activity by different levels of agreement among the survey respondents.

	Pain, 0–10	Swollen Joints, 0–28	Tender Joints, 0–28	HAQ, 0–3	Physician Global, 0–10	Patient Global, 0–10	ESR, 0–120	DAS28, 0–9.2
$\geq 90\%$ Respondent agreement (15 profiles)								
Mean (SD)	1.2 (0.9)	0.0 (0.0)	0.3 (0.8)	0.2 (0.2)	0.7 (0.7)	1.1 (1.3)	11.7 (11.4)	1.6 (0.9)
95% CI	0.7, 1.7	—	–0.1, 0.8	0.1, 0.3	0.3, 1.0	0.5, 1.7	6.0, 17.5	1.2, 2.1
Min, max	0.0, 3.0	0.0, 0.0	0.0, 3.0	0.0, 0.8	0.0, 2.0	0.0, 5.0	1.0, 36.0	0.0, 3.2
$\geq 80\%$ Respondent agreement (17 profiles)								
Mean (SD)	1.4 (1.0)	0.1 (0.5)	0.3 (0.8)	0.2 (0.4)	0.8 (0.8)	1.21 (1.3)	11.7 (11.0)	1.7 (0.8)
95% CI	0.9, 1.9	–0.1, 0.4	–0.1, 0.7	0.1, 0.4	0.4, 1.2	0.6, 1.8	6.4, 16.9	1.3, 2.1
Min, max	0.0, 3.0	0.0, 2.0	0.0, 3.0	0.0, 1.3	0.0, 2.0	0.0, 5.0	1.0, 36.0	0.0, 3.2
$\geq 70\%$ Respondent agreement (22 profiles)								
Mean (SD)	1.6 (1.3)	0.1 (0.4)	0.4 (0.8)	0.4 (0.5)	0.9 (1.0)	1.9 (2.1)	13.3 (10.3)	2.0 (0.9)
95% CI	1.1, 2.2	–0.1, 0.3	0.03, 0.7	0.2, 0.6	0.5, 1.3	1.0, 2.7	9.0, 17.6	1.6, 2.3
Min, max	0.0, 4.5	0.0, 2.0	0.0, 3.0	0.0, 2.1	0.0, 3.0	0.0, 9.0	1.0, 36.0	0.0, 3.5

HAQ: Health Assessment Questionnaire, ESR: erythrocyte sedimentation rate, DAS: Disease Activity Score.

Table 2. Cutpoints of the core measures for the 5 candidate definitions for minimal disease activity.

	Pain, 0–10	Swollen Joints, 0–28	Tender Joints, 0–28	HAQ, 0–3	Physician Global, 0–10	Patient Global, 0–10	ESR, 0–120
Mean	1.38	0.12	0.29	0.24	0.82	1.21	11.65
Rounded mean (Definition A)	1.5	1	1	0.25	1	1.5	15
Upper 95% limit	1.85	0.35	0.66	0.41	1.21	1.81	16.88
Rounded upper 95% limit (Definition B)	2	1	1	0.5	1.5	2	20
Maximum (Definition C)	3	2	3	1.25	2	5	36

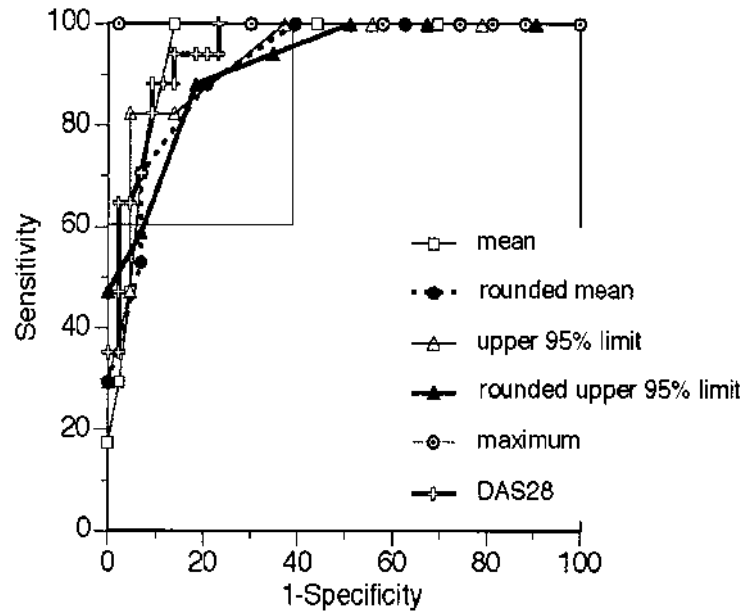


Figure 2. Receiver operating characteristic curves for the 5 candidate definitions for minimal disease activity. DAS: Disease Activity Score.

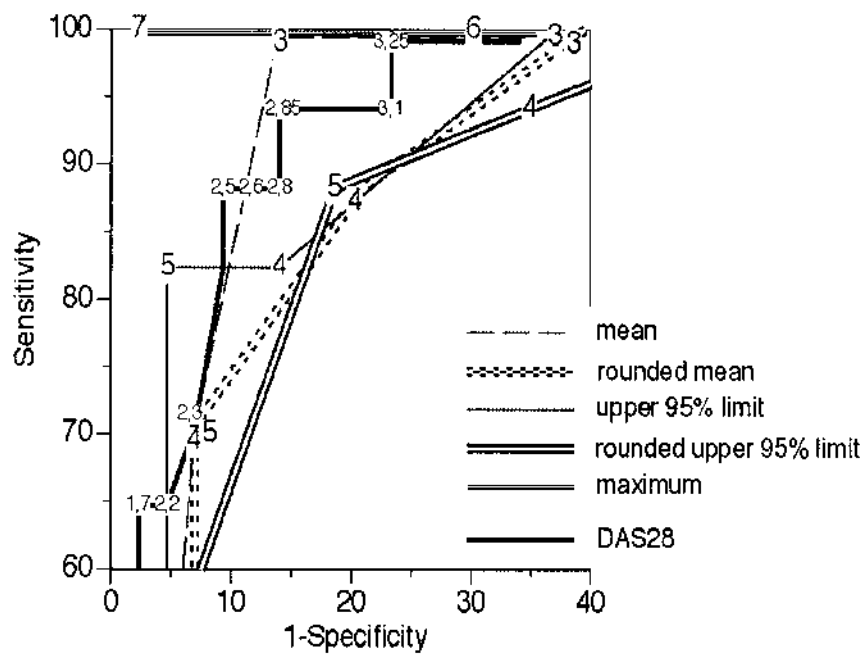


Figure 3. Critical region of the receiver operating characteristic curves for the 5 candidate definitions for minimal disease activity. DAS: Disease Activity Score.

ficiency of 94% and 86%, respectively. It should be noted that these results should be regarded with some caution in view of the low number of MDA profiles and the low number of respondents.

The candidate definitions for MDA based on the core set that were presented to the participants for their consideration were: the rounded mean (Definition A), rounded upper

95% limit (Definition B), and the maximum (Definition C) (Table 2). In addition, participants were asked to determine the best level of DAS28, equivalent with the preferred core set MDA definition.

OMERACT 7 Module. The results from the survey were presented at the plenary of the OMERACT 7 Module. During breakout sessions key aspects of concepts and results issues

were discussed, and rapporteurs reported several key issues including: the usefulness of new definitions given that remission has already been defined, the usefulness of measures in the proposed definitions, misclassification, change versus state, and potential for abuse. Feedback from the breakout sessions was recorded and reviewed in order to clarify the purpose, identify misunderstandings, and address concerns raised. With access to the RAES database⁵ and the Vienna profile survey⁸, onsite consideration of these issues was possible prior to the vote at final conference plenary.

1. MDA versus remission. To be useful the MDA definition must be distinct from current remission definitions. Specific (“strict”) candidate definition [such as the 7 out of 7 definition (Definition C) and the lower cutpoint for the DAS definitions] lowers sensitivity and classifies only patients close to remission. This was confirmed based on RAES database analysis and resolved by applying more sensitive definitions, such as Definition B, with 5 of 7 of the core measures meeting the cutpoint and the higher DAS cutoff of 2.85. Analysis of the RAES database indicated that about 20% of patients met either definition (but not always both). It is worth noting that even the 2.85 cutoff for the DAS28 is close to the current definition for DAS remission of 2.6. However, we feel this is more a problem of the remission definition, especially when obtained with the 28-joint count. Recently, some of us showed that patients with a DAS28 of 2.6 or less can have substantial residual disease, especially in the feet⁹. Probably all definitions of remission based on reduced joint counts need to be reconsidered in this light.

2. Usefulness of measures in the definition. Several points were raised regarding the candidate measures in the definition. Fatigue was recognized as an important measure to include once an instrument measuring it passes the OMERACT filter¹⁰. There was some support among participants for using C-reactive protein (CRP), by substituting CRP for ESR after applying a validated nomogram. Restricted joint counts were recognized as an increasing problem for all definitions (including the ACR remission definition) as disease activity goes down. This issue was identified for the research agenda. That global questions (e.g., patient global, physician global, and pain) need to focus on disease activity was recognized and was identified as an important point to consider when the core set is reviewed.

3. Misclassification. Although misclassification can arise with any definition (it is more acceptable in a randomized trial setting since misclassification applies equally in all treatment arms), doubts were raised about face validity: Fixed disability in established disease will keep HAQ elevated, and MDA will be difficult or impossible to achieve according to a definition that includes a low HAQ level. However, in an analysis using the full RAES database, the proportions of patients meeting MDA by various candidate definitions were similar in early, established, and late disease, suggesting this may not be a substantive problem.

Another concern was that patients with a chronic pain syndrome but with low RA disease activity would be misclassified as high disease activity due to high scores in pain, tender joint counts, and patient global assessment. In theory, a DAS definition of MDA should be less vulnerable to this problem because pain is not a component, and patient global assessment carries only a small weight in the index. In the RAES database about 20% of patients were identified with a probable concomitant pain syndrome. Of these, 8% more were in MDA according to the DAS definition than according to core set definition, suggesting some misclassification occurs in the core set definition. To address this problem the tree approach was suggested: a decision node was placed before the definition to better classify patients with MDA who have high pain scores deemed unrelated to disease activity. Strict terms at this node would be required to avoid introducing new misclassification problems. The node suggested was: If swollen joint count = 0, tender joint count = 0, and ESR ≤ 10, then the patient is considered to be in MDA, regardless of the results of other core set measures. Most patients meeting this node would in fact be in remission.

The data were reanalyzed to include this initial decision with the following results: in the profile database from the stakeholder survey, there was no difference in the classification of patients; in the RAES database, most patients with high pain levels were now “correctly” classified; and in the Vienna survey data correspondence was found to be fair. The preliminary feedback at OMERACT 7 was that the initial decision could be partially redundant, and perhaps only swollen joint count and “normal” acute phase reactant are needed. This was considered to be an important item for a later research agenda.

Other concerns with misclassification included patients with only high joint counts and patients having comorbid conditions with chronic pain. However, patients with only high joint counts are a rare occurrence, since joint counts will drive up the physician global and patient global assessment and pain scores. The precise prevalence is a research agenda item.

Comorbid conditions with chronic pain syndromes are an issue, but inflammatory and other conditions causing high ESR are not seen as a large problem. Again, this is an item for a research agenda.

4. Change versus state. Interest was expressed in having a measure of both state and change. It was noted that the EULAR response requires both a minimum change and a certain state to be reached, whereas the ACR20 requires only a change. If agreement on a preliminary definition of MDA was achieved, the next step could be to explore combinations with response criteria such as the ACR70.

5. Potential for abuse. In particular, 2 potential areas for misuse of the MDA definition were noted: (1) There was concern that the definitions would be used for individual

patient management and reimbursement. This concern cannot be fully resolved, but it must be emphasized that the proposed definitions are not intended for or useful to guide decisions for individual patient care. The development of other definitions for guiding patient care may be useful and would be a worthwhile item to consider on a research agenda. Other possible abuse involves using the definition to make comparisons of results between trials. Such comparisons are of dubious validity as the study populations may differ substantially. However, such cross-trial comparisons are commonplace and occur with other response criteria as well (e.g., ACR20).

Definition of MDA. The 3 candidate definitions of MDA were considered by all the breakout groups. Although Definition C (with the cutpoints based on the maximum values and requiring all 7 criteria to be satisfied) had the best combination of sensitivity and specificity, OMERACT 7 participants indicated that this definition did not have great face validity and it was not scored highly in the breakout sessions. Also, as noted earlier, because of the specific nature of this definition, only patients close to remission would be classified as being in MDA. Of the 3 definitions, Definition B (with the cutpoints based on the upper 95% limits and requiring 5 or more of the 7 criteria to be satisfied) garnered the greatest support.

There are 2 sets of outcome measures currently used as the primary endpoint in RA clinical trials: the WHO/ILAR core set (corresponding ACR response criteria)⁴ and the DAS28 (corresponding EULAR response criteria)¹¹. To follow current practice in trial methodology, 2 equivalent preliminary definitions of MDA for use as a secondary outcome measure in clinical trials in RA were proposed. Researchers applying these definitions first need to choose between the DAS28 and core set definition because each definition selects a similar proportion in a population but not always the same patients.

1. Core set definition using Definition B (upper 95% limit; 5 out of 7) preceded by a decision node for including patients with high pain or HAQ levels but otherwise in MDA (Figure 4): For this definition, a patient with no tender or swollen joints and an $ESR \leq 10$ would be considered to be in MDA, otherwise the full set of core measures is considered. If 5 of the following 7 criteria are met, then the patient is considered to be in MDA: pain ≤ 2 , swollen joint count ≤ 1 , tender joint count ≤ 1 , HAQ ≤ 0.5 , physician global ≤ 1.5 , patient global ≤ 2 , and $ESR \leq 20$. Otherwise, the patient is not in MDA.

2. DAS-based definition using DAS of 2.85 preceded by the same decision node (Figure 5): For this definition, a patient with no tender or swollen joints and an $ESR \leq 10$ would be considered as MDA, otherwise the DAS28 would be considered. Only if $DAS \leq 2.85$ would the patient be classified as MDA.

The following question was posed to the OMERACT 7 participants at the conference plenary: "Do you agree that

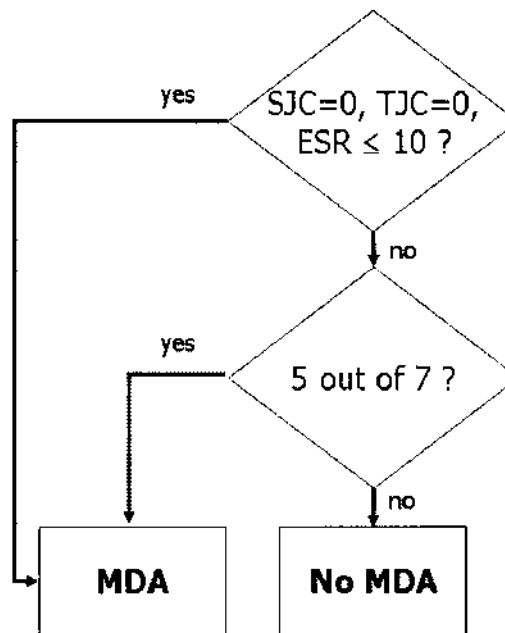


Figure 4. The core-set definition for minimal disease activity for RA. Pain ≤ 2 ; swollen joint count (SJC) ≤ 1 ; tender joint count (TJC) ≤ 1 ; Health Assessment Questionnaire ≤ 0.5 ; physician global assessment 1.5; patient global assessment ≤ 2 ; $ESR \leq 20$ mm/h.

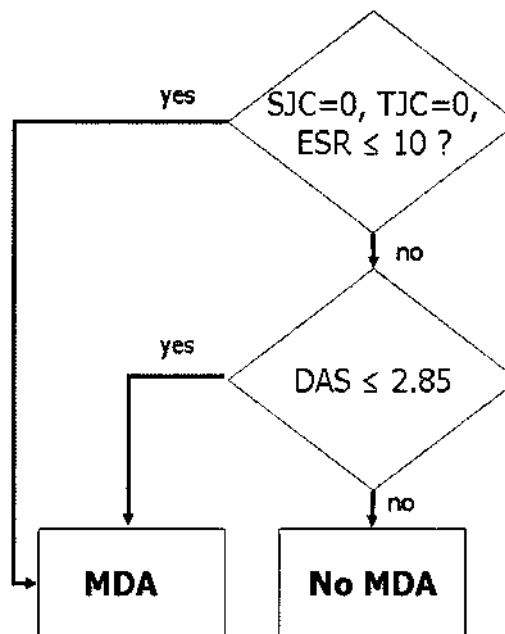


Figure 5. The DAS-based definition for minimal disease activity (MDA) for RA. SJC: swollen joint count; TJC: tender joint count; ESR: erythrocyte sedimentation rate; DAS: Disease Activity Score.

both the core set definition and the DAS-based definition have sufficiently passed the OMERACT filter to be recommended as preliminary definitions of MDA for use in randomized clinical trials, to be further validated in other datasets and longterm outcome databases?"

There was 73% agreement endorsing the core set definition and DAS-based definition as preliminary definitions of MDA.

DISCUSSION

Consensus was sought on a definition of MDA that can be recommended for use in randomized clinical trials and further validated in other datasets and longterm outcome databases. The proposed definitions were not intended or useful to guide decisions in individual patient care.

A research agenda was identified to evaluate preliminary definitions and related issues. The discussion at OMERACT 7 provided a framework for next steps, with several issues tabled as “research agenda items”, including prospective validation against longterm outcome (using existing databases) and against the opinion of real patients classified as being in MDA (to address the issue of clinical context); misclassification due to extraneously elevated acute phase proteins, other core set measures, and high disease activity in ankles or feet; validation of a CRP nomogram for substituting CRP for ESR in the definitions. Initiatives included: to explore the usefulness of a combination of minimal disease activity and response; to include a measure of fatigue; to develop a definition for patient care; to test other definitions; and to test definitions for redundancy.

The preliminary definitions for MDA satisfy the OMERACT filter¹⁰ for truth, discrimination, and feasibility. *Truth* is the opinion of the physician being met by the proposed definition in the setting of randomized clinical trials; all important issues have been considered, no “fatal” issues remain, and a large research agenda has been identified for other issues. *Discrimination* in classification criteria is subsumed under truth. *Feasibility* of using the definitions in a trial setting is achievable.

REFERENCES

1. Wells G, Boers M, Shea B, et al. MCID/Low Disease Activity State Workshop: Low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1110-1.
2. Boers M, Anderson JJ, Felson DT. Deriving an operational definition of low disease activity state in rheumatoid arthritis. *J Rheumatol* 2003;30:1112-4.
3. 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.
4. Felson DT, Anderson J, Boers M, et al. American College of Rheumatology preliminary criteria for improvement in rheumatoid arthritis patients. *Arthritis Rheum* 1995;38:727-35.
5. Wolfe F, Michaud K, Pincus T, Furst D, Keystone E. The DAS score is not suitable as the criterion for TNF initiation in the clinic: evidence of disagreement between physician and patient assessments and DAS scores [abstract]. *Arthritis Rheum* 2004;50 Suppl:S459.
6. Kirwan J, Heiberg T, Hewlett S, et al. Outcomes from the Patient Perspective Workshop at OMERACT 6. *J Rheumatol* 2003;30:868-72.
7. Kirwan JR, Hewlett S, Heiberg T, et al. Incorporating the patient perspective into outcome assessment in rheumatoid arthritis — progress at OMERACT 7. *J Rheumatol* 2005;32: (in press).
8. Aletaha D, Ward MM, Machold KPM, Nell VPK, Stamm T, Smolen JS. Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states. Submitted for publication.
9. Landewe R, van der Heijde D, Voskuyl AE, Boers M. Condensed 28-joint counts jeopardise the construct validity of DAS28-remission: a comparison with the original DAS [abstract]. *Arthritis Rheum* 2004;50 Suppl:S458.
10. Boers M, Brooks P, Strand V, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-9.
11. 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.