

The 2010 American College of Rheumatology/ European League Against Rheumatism Classification Criteria for Rheumatoid Arthritis

Phase 2 Methodological Report

Tuhina Neogi,¹ Daniel Aletaha,² Alan J. Silman,³ Raymond L. Naden,⁴ David T. Felson,¹ Rohit Aggarwal,⁵ Clifton O. Bingham, III,⁶ Neal S. Birnbaum,⁷ Gerd R. Burmester,⁸ Vivian P. Bykerk,⁹ Marc D. Cohen,¹⁰ Bernard Combe,¹¹ Karen H. Costenbader,¹² Maxime Dougados,¹³ Paul Emery,¹⁴ Gianfranco Ferraccioli,¹⁵ Johanna M. W. Hazes,¹⁶ Kathryn Hobbs,¹⁷ Tom W. J. Huizinga,¹⁸ Arthur Kavanaugh,¹⁹ Jonathan Kay,²⁰ Dinesh Khanna,²¹ Tore K. Kvien,²² Timothy Laing,²³ Katherine Liao,¹² Philip Mease,²⁴ Henri A. Ménard,²⁵ Larry W. Moreland,⁵ Raj Nair,²⁶ Theodore Pincus,²⁷ Sarah Ringold,²⁸ Josef S. Smolen,² Ewa Stanislawska-Biernat,²⁹ Deborah Symmons,³⁰ Paul P. Tak,³¹ Katherine S. Upchurch,²⁰ Jiří Vencovský,³² Frederick Wolfe,³³ and Gillian Hawker³⁴

Objective. The American College of Rheumatology

Supported by the American College of Rheumatology and the European League Against Rheumatism.

¹Tuhina Neogi, MD, PhD, FRCPC, David T. Felson, MD, MPH: Boston University School of Medicine, Boston, Massachusetts; ²Daniel Aletaha, MD, MSc, Josef S. Smolen, MD: Medical University of Vienna, Vienna, Austria; ³Alan J. Silman, FRCP, FmedSci, DSc (Hons): Arthritis Research UK, Chesterfield, UK; ⁴Raymond L. Naden, MB ChB, FRACP: Ministry of Health, Auckland, New Zealand; ⁵Rohit Aggarwal, MD, MS, Larry W. Moreland, MD: University of Pittsburgh, Pittsburgh, Pennsylvania; ⁶Clifton O. Bingham, III, MD: Johns Hopkins University, Baltimore, Maryland; ⁷Neal S. Birnbaum, MD: California Pacific Medical Center and University of California, San Francisco; ⁸Gerd R. Burmester, MD: Charité Hospital–University Medicine Berlin, Free University and Humboldt University, Berlin, Germany; ⁹Vivian P. Bykerk, MD, FRCPC: Mount Sinai Hospital and University of Toronto, Toronto, Ontario, Canada; ¹⁰Marc D. Cohen, MD: National Jewish Medical and Research Center, Denver, Colorado; ¹¹Bernard Combe, MD, PhD: Lapeyronie Hospital and Montpellier I University, Montpellier, France; ¹²Karen H. Costenbader, MD, MPH, Katherine Liao, MD: Brigham and Women’s Hospital and Harvard University, Boston, Massachusetts; ¹³Maxime Dougados, MD: Cochin Hospital, Assistance Publique Hôpitaux de Paris, and Paris-Descartes University, Paris, France; ¹⁴Paul Emery, MA, MD, FRCP: University of Leeds and NIHR Leeds Biomedical Research Unit, Leeds, UK; ¹⁵Gianfranco Ferraccioli, MD: School of Medicine, Catholic University of the Sacred Heart, Rome, Italy; ¹⁶Johanna M. W. Hazes, MD, PhD: Erasmus Medical Center and University of Rotterdam, Rotterdam, The Netherlands; ¹⁷Kathryn Hobbs, MD: University of Colorado School of Medicine, Denver; ¹⁸Tom W. J. Huizinga, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands; ¹⁹Arthur Kavanaugh, MD: University of California, San Diego; ²⁰Jonathan Kay, MD, Katherine S. Upchurch, MD:

and the European League Against Rheumatism have developed new classification criteria for rheumatoid arthritis (RA). The aim of Phase 2 of the development process was to achieve expert consensus on the clinical and laboratory variables that should contribute to the final criteria set.

Methods. Twenty-four expert RA clinicians (12 from Europe and 12 from North America) participated in Phase 2. A consensus-based decision analysis approach was used to identify factors (and their relative

UMassMemorial Medical Center and University of Massachusetts Medical School, Worcester; ²¹Dinesh Khanna, MD, MS: David Geffen School of Medicine at University of California, Los Angeles; ²²Tore K. Kvien, MD, PhD: Diakonhjemmet Hospital, Oslo, Norway; ²³Timothy Laing, MD: University of Michigan, Ann Arbor; ²⁴Philip Mease, MD: Swedish Medical Center and University of Washington, Seattle; ²⁵Henri A. Ménard, MD: McGill University Health Centre and McGill University, Montreal, Quebec, Canada; ²⁶Raj Nair, MD: University of North Carolina, Chapel Hill; ²⁷Theodore Pincus, MD: New York University Hospital for Joint Diseases, New York, New York; ²⁸Sarah Ringold, MD, MS: Seattle Children’s Hospital, Seattle, Washington; ²⁹Ewa Stanislawska-Biernat, MD, PhD: Institute of Rheumatology, Warsaw, Poland; ³⁰Deborah Symmons, MD, FFPH, FRCP: University of Manchester, Manchester, UK; ³¹Paul P. Tak, MD, PhD: Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands; ³²Jiří Vencovský, MD, DSc: Institute of Rheumatology, Prague, Czech Republic; ³³Frederick Wolfe, MD: National Data Bank for Rheumatic Diseases and University of Kansas, Wichita; ³⁴Gillian Hawker, MD, MSc, FRCPC: Women’s College Hospital and University of Toronto, Toronto, Ontario, Canada.

weights) that influence the probability of “developing RA,” complemented by data from the Phase 1 study.

Dr. Aletaha has received consulting fees, speaking fees, and/or honoraria from Abbott, Bristol-Myers Squibb, UCB, Schering-Plough, Wyeth, and Roche (less than \$10,000 each). Dr. Naden has received consulting fees from the American College of Rheumatology in regard to the methodology of developing weighted scoring systems (more than \$10,000). Dr. Bingham has received consulting fees, speaking fees, and/or honoraria from UCB, Roche, Genentech, Celgene, and Merck Serono (less than \$10,000 each); he has received research and/or educational grant support from Bristol-Myers Squibb, Genentech, UCB, Centocor, Abbott, and Amgen. Dr. Birnbaum has received consulting fees, speaking fees, and/or honoraria from Amgen, Pfizer, Centocor, Abbott, and UCB (less than \$10,000 each). Dr. Burmester has received consulting fees, speaking fees, and/or honoraria from Abbott, Bristol-Myers Squibb, Pfizer, UCB, and Roche (less than \$10,000 each). Dr. Bykerk has received consulting fees, speaking fees, and/or honoraria from Amgen, Wyeth, Abbott, Schering-Plough, Roche, Bristol-Myers Squibb, and UCB (less than \$10,000 each); her spouse is employed by Genzyme and owns stock in the company. Dr. Cohen has received consulting fees, speaking fees, and/or honoraria from UCB, Genentech, Bristol-Myers Squibb, and Human Genome Sciences (less than \$10,000 each). Dr. Combe has received consulting fees, speaking fees, and/or honoraria from Abbott, Bristol-Myers Squibb, Pfizer, Roche, Schering-Plough, and Merck, Sharpe, and Dohme (less than \$10,000 each). Dr. Emery has received consulting fees, speaking fees, and/or honoraria from Pfizer, Abbott, Centocor, UCB, Roche, Bristol-Myers Squibb, and Merck, Sharpe, and Dohme (less than \$10,000 each). Dr. Ferraccioli holds a patent for T cell receptor clonotype analysis (PCT/IB 2008/053152 NP). Dr. Huizinga has received consulting fees, speaking fees, and/or honoraria from Schering-Plough, Bristol-Myers Squibb, UCB, Biotest AG, Wyeth/Pfizer, Novartis, Roche, Sanofi-Aventis, Abbott, and Axis-Shield (less than \$10,000 each). Dr. Kavanaugh has conducted clinical research for Amgen, Abbott, Bristol-Myers Squibb, UCB, Roche, Centocor, Genentech, and Sanofi-Aventis. Dr. Kay has received consulting fees from Array BioPharma, Bristol-Myers Squibb, Celgene, Centocor, Genentech, Roche, UCB, and Sanofi-Aventis (less than \$10,000 each). Dr. Khanna has received consulting fees, speaking fees, and/or honoraria from UCB and Abbott (less than \$10,000 each). Dr. Mease has received consulting fees, speaking fees, and/or honoraria from Abbott, Amgen, Biogen Idec, Bristol-Myers Squibb, Centocor, Roche, Genentech, UCB, Pfizer, Novartis, and Eli Lilly (less than \$10,000 each). Dr. Ménard has received unrestricted educational and research grants as well as consulting and speaking fees from Abbott, Amgen, Inova, Merck, Pfizer, Roche, Schering-Plough, UCB, and Wyeth (less than \$10,000 each) and investigator-initiated research grants from Bristol-Myers Squibb, EuroImmune AG, and Roche (more than \$10,000 each); he owns stock or stock options in Merck; and he has a license agreement with EuroImmune AG for an anti-Sa enzyme-linked immunosorbent assay. Dr. Moreland has received consulting fees, speaking fees, and/or honoraria from Biogen Idec, Centocor, Pfizer, Takeda, KaloBios, ChemoCentryx, UCB, Genentech, Incyte, and Eli Lilly (less than \$10,000 each). Dr. Pincus has received consulting fees, speaking fees, and/or honoraria from Amgen, Abbott, Bristol-Myers Squibb, Centocor, UCB, Wyeth, and Genentech (less than \$10,000 each) and investigator-initiated research grants from Amgen, Bristol-Myers Squibb, UCB, and Centocor. Dr. Stanislawska-Biernat has received speaking fees from Abbott and Pfizer (less than \$10,000 each). Dr. Vencovský has received speaking fees from Pfizer, UCB, Abbott, Roche, and Merck, Sharpe, and Dohme (less than \$10,000 each).

Address correspondence and reprint requests to Gillian Hawker, MD, MSc, Department of Medicine, Women's College Hospital, 76 Grenville Street, 8th Floor, Room 815, Toronto, Ontario M5S 1B2, Canada. E-mail: g.hawker@utoronto.ca.

Patient case scenarios were used to identify and reach consensus on factors important in determining the probability of RA development. Decision analytic software was used to derive the relative weights for each of the factors and their categories, using choice-based conjoint analysis.

Results. The expert panel agreed that the new classification criteria should be applied to individuals with undifferentiated inflammatory arthritis in whom at least 1 joint is deemed by an expert assessor to be swollen, indicating definite synovitis. In this clinical setting, they identified 4 additional criteria as being important: number of joints involved and site of involvement, serologic abnormality, acute-phase response, and duration of symptoms in the involved joints. These criteria were consistent with those identified in the Phase 1 data-driven approach.

Conclusion. The consensus-based, decision analysis approach used in Phase 2 complemented the Phase 1 efforts. The 4 criteria and their relative weights form the basis of the final criteria set.

The new American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria for rheumatoid arthritis (RA) were developed in 3 phases (1). Phase 1, led by the EULAR (AS and DA), used cohort data to identify the key factors to be considered in the new criteria, and their associated weights (2). The current report outlines the second phase, led by the ACR (TN, DF, and GH). A consensus-based, decision science-informed approach was used to identify factors that influence expert RA clinicians' opinions about the probability of developing persistent inflammatory or erosive arthritis (“developing RA”). The rationale for this approach was 2-fold: to ensure that expert clinicians' perspectives were captured, and to ensure that potentially important factors not captured in the Phase 1 cohort data might be identified. Results from Phases 1 and 2 were subsequently integrated to determine the final criteria set (1).

METHODS

Phase 2 included the following steps: 1) assembly of an expert panel, 2) development and rank ordering of patient case scenarios, 3) a 2-day in-person consensus meeting, and 4) assessment of face and construct validity.

The expert panel. With input from the ACR and EULAR leadership, equal numbers of North American and European expert clinical rheumatologists were selected. The

Submitted for publication January 27, 2010; accepted in revised form May 20, 2010.

expert panel included community and academic rheumatologists, and was diverse in terms of geography and numbers of men and women.

Development of patient case scenarios. Expert panel members used a standardized template (Supplementary Figure 1, available in the online version of this article at <http://www3.interscience.wiley.com/journal/76509746/home>) to submit 3–5 real-life case scenarios representing patients with early (within 1 year of symptom onset) undifferentiated inflammatory arthritis. These scenarios included all patient information that the experts considered relevant to rule in (positive factors) or out (negative factors) an eventual diagnosis of RA.

Each scenario captured the following patient elements: age and sex, duration of joint pain, duration of joint swelling, average duration of morning stiffness, and distribution of affected joints (swollen and tender joints, indicated on joint homunculi). The expert also provided information on the subsequent disease course, whether or not treatment with methotrexate (MTX) had been initiated at that assessment time point, and the expert's opinion, using a 5-point Likert scale from 1 (very low probability) to 5 (very high probability), of the probability that the patient would, if untreated, "develop RA."

Each completed case scenario was assigned a unique name. Two members of the steering committee (TN and GH) selected a subset of 30 case scenarios that best represented the spectrum of probability of RA development. Most of the cases were in the middle 3 probability categories. These 30 scenarios were then simplified and standardized. The submitting expert's identity, opinion regarding the probability of RA, and information on the subsequent disease course were removed.

Rank ordering of case scenarios by the expert panel. Following review of the Phase 1 study results (2), expert panel members rank ordered the 30 cases, from 1 (highest probability of developing RA), to 30 (lowest probability). Additionally, for each case, the expert panel members indicated whether they would initiate treatment with MTX (yes or no, assuming that there were no contraindications and the patient was agreeable). The mean and distribution of rankings for each case scenario and for each expert panel member were plotted.

In-person consensus meeting. The expert panel met for 2 days in May 2009. The meeting was facilitated by RLN, an internist from Auckland, New Zealand who has expertise in consensus conferencing and use of decision analytic software (3).

Identification of domains, categories, and weights. Members of the expert panel were presented their case rankings, and in-depth group discussion of potential reasons for lack of agreement identified key positive and negative factors that were important in evaluating patients' probability of developing RA. Evidence from both Phase 1 study data and published data, as available, was used to support the discussions and decisions. After the panelists identified a list of key factors, the ones deemed most important and essential were agreed upon. These essential factors or criteria form the basis of the final criteria set. They then defined specific categories within each criterion that signified different levels of probability. For example, for pattern and extent of joint involvement, a hierarchy of a combination of number and type of involved joints defined the various categories within that criterion, signifying increasing levels of probability of developing RA.

Since the resulting criteria and their respective categories produce multiple possible combinations of clinical features, decision analytic software (1000Minds [www.1000minds.com]) was used to facilitate the quantification of the relative importance or "weight" for each criterion and category. The decision analytic software program used choice-based conjoint analysis (sometimes referred to as "discrete choice experiments" or "multi-criteria decision analysis") to evaluate, through discrete pairwise choices, the weights attached to the categories within each criterion. This approach has been used successfully in other projects (4–8), for example to enumerate factors affecting urgency of need for referral to rheumatologists for acute rheumatic conditions. The pairwise ranking employed by this methodology is a natural human activity that people experience in their daily lives. Deciding between just 2 alternatives is cognitively less burdensome, and therefore arguably more valid and reliable, than alternative methods for eliciting preferences to derive the weights. This method is more efficient than others because any pairwise decisions in which one option clearly has higher probability "RA development" (e.g., "high-positive serology and >10 joints involved" has a higher probability than "low-positive serology and 1–3 small joints involved") are not presented for decision-making. Efficiency is also gained by not requiring further discussion when there is consensus. The program can also be administered over the Internet, allowing for the process to be conducted without an in-person meeting when necessary. A major advantage is that individual categories can be modified, such as when new information becomes available, and the weightings recalculated without disrupting the validity of the method or the previous consensus decisions made.

An experienced facilitator guided the use of the software program. The expert panel participants were presented a series of paired scenarios, each comprising information relevant to 2 different criteria. For each pair, they independently and anonymously (using touch keypads) chose the scenario they thought had the higher probability of "developing RA." Figure 1 shows an example in which case 1 has a lower level of joint involvement and a higher level of abnormal serology, while case 2 has the opposite. The distribution of opinions (case 1 more likely to develop RA, case 2 more likely, equally likely) was presented to the group and the reasons for disagreement, if any, were discussed. The group then re-voted. Consensus was considered achieved when all participants either indicated agreement or indicated that they could accept the majority decision.

Based on these discrete choices, the decision analytic software program uses mathematical methods to determine the relative importance, and thereby weight, of each category within each criterion. The process is iterative, such that each successive result further refines the weights derived through prior choice outcomes. The final weights determine the scores assigned to each category, and the sum of the weights produces a total score for each case, from low to high probability. The weights are scaled such that those associated with the highest categories in each criterion sum to 100. Thus, possible scores range from 0 to 100, with a higher score indicating a higher probability of developing RA.

Assessment of the face validity of the weights. The next step assessed whether the case rankings, achieved using the "probability of developing RA" scores, reflected the experts'

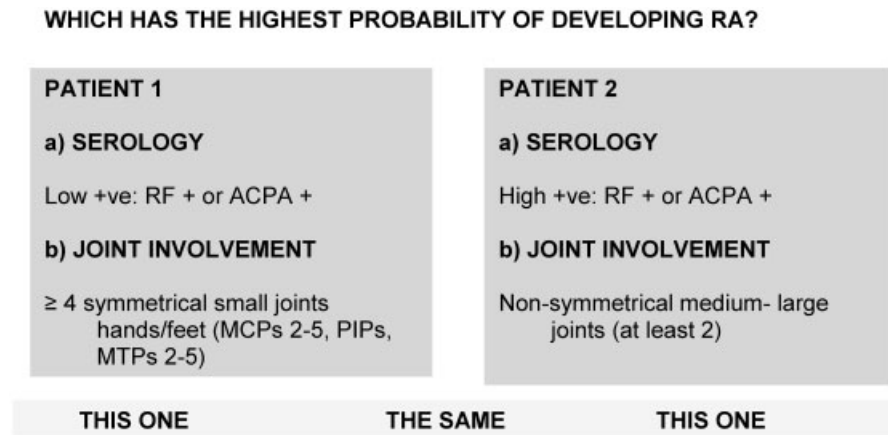


Figure 1. Example of a discrete choice experiment. RA = rheumatoid arthritis; RF = rheumatoid factor; ACPA = anti-citrullinated protein antibody; MCPs = metacarpophalangeal joints; PIPs = proximal interphalangeal joints; MTPs = metatarsophalangeal joints.

clinical judgment. Ten of the 30 case scenarios, representing very low to very high probability of RA, were selected and scored. The expert panel reviewed the rankings and discussed cases that seemed out of place, based on clinical impression. Where necessary, modifications were made to the criteria categories to address concerns raised. The decision analytic program was then used to reassign weights to the revised categories, followed by scoring and re-review of the 10 case rankings to confirm that concerns had been addressed.

Assessment of interrater reliability of scoring. At a second in-person meeting, Phase 1 and Phase 2 participants reviewed the results of the consensus meeting, further revised and defined each of the criteria and criteria categories, and outlined steps for validation. The 56 case scenarios not included in the initial set of 30 were simplified, anonymized, and assigned a unique name. The steering committee members and 2 volunteers who had not participated in the in-person consensus meeting independently categorized each case within each criterion to determine the consistency of categorization, and thus ultimate scoring. By teleconference, the group discussed potential explanations for disagreements. Modifications were made to the criteria or category definitions until full consensus was achieved. These definitions were compiled in a glossary.

Assessment of face and construct validity. Cases were rank ordered from highest probability of developing RA (score closest to 100) to lowest probability of developing RA (score closest to 0). Phase 1 and 2 participants reviewed the rank ordering to identify cases that were substantively out of place. Additionally, for each case, the panelists indicated if they would 1) treat with MTX or another disease-modifying antirheumatic drug (DMARD) due to risk for developing RA (yes/no) and 2) assuming eligibility, enroll the patient into a clinical trial of an investigational biologic therapy with inherent risks (yes/no).

RESULTS

Twenty-four expert rheumatologists (12 European, 12 North American) participated in Phase 2. Among the North American rheumatologists, 25% were female and 75% were in academic (as opposed to community) practice; among the European rheumatologists, 25% were female and 50% were in academic practice. The expert panelists submitted 86 patient case scenarios in total. From these, 30 were chosen, simplified, standardized, anonymized, and rank ordered by the expert panel based on panel members' opinion of the probability of developing RA. The distribution of rankings, shown in Figure 2, indicated a relative lack of consensus for most cases; some cases, e.g., case 14, received the full range of probability rankings, from highest to lowest.

Twenty-two of the 24 expert panel members attended the in-person consensus meeting. Through review of the expert panel rankings and the reasons for disagreement, the following were identified as important factors in determining the probability of developing RA: expert determination of evidence of joint swelling, indicating synovitis; morning stiffness; joint distribution (site, number, symmetry); temporal evolution of joint involvement; family history; age; sex; joint tenderness versus swelling; features of another inflammatory arthritis; physician global assessment; serology (anti-citrullinated protein antibody [ACPA], and rheumatoid factor [RF]); acute-phase response; duration of symptoms; duration of synovitis; and others. Each factor was

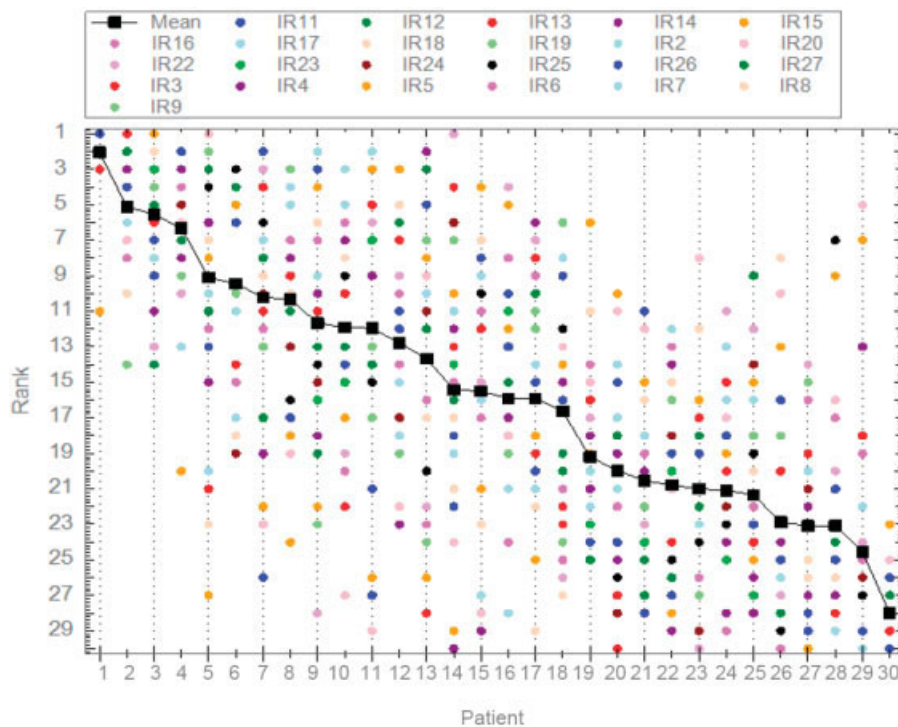


Figure 2. Initial rankings (IR) by the expert rheumatologist panel (n = 24). Expert panel members are indicated by colored dots.

discussed, and the evidence supporting its usefulness reviewed. Further, all discussions about potential factors for inclusion in the new classification criteria took into account the ability of these criteria to be used throughout the world, regardless of income level or health care system. From these discussions, the expert panel identified the factors that should be incorporated in the new classification criteria.

Identifying the population to which the new criteria should be applied. To ensure that the classification criteria are applied to persons with undifferentiated inflammatory arthritis, the expert panel identified 2 mandatory criteria. First, there should be evidence, as determined by an expert assessor, of swelling, indicating synovitis, in at least 1 synovial joint, excluding joints typically involved in osteoarthritis (distal interphalangeal [DIP] joints, first metatarsophalangeal [MTP] joint, first carpometacarpal [CMC] joint). Second, signs and symptoms must not be better explained by another diagnosis. Thus, depending on the patient presentation and context (e.g., sociodemographics and geographic prevalence of specific conditions), if another definable disease better explains the presence of synovitis, the new RA criteria should not be applied.

Domains of importance for the new criteria set.

Assuming these 2 mandatory criteria were met, the following additional criteria were identified as being essential for determining the probability of developing RA: pattern and extent of joint involvement, duration of signs and symptoms of synovitis, serologic findings (ACPA or RF), and acute-phase response (erythrocyte sedimentation rate [ESR] or C-reactive protein [CRP] level).

Pattern and extent of joint involvement. Given the mandatory requirement for expert-determined synovitis (swelling) in at least 1 joint, the expert panel agreed that this criterion should refer to the number and distribution of “involved joints,” defined as tender *or* swollen joints at the time of the physician assessment. Again, due to concerns about overlap with osteoarthritis, the DIP, first MTP, and first CMC joints should not be included. Six categories associated with increasing probability of developing RA were determined within this criterion: 1) monoarthritis of a medium-large joint (shoulder, elbow, hip, knee, or ankle); 2) at least 2 *asymmetrically involved* medium-large joints; 3) at least 2 *symmetrically involved* medium-large joints; 4) 1–3 involved small joints of the hands and feet (metacarpophalangeal, proximal interphalangeal, second through fifth MTP) or wrists; 5) ≥ 4

Table 1. Initial criteria, their respective categories, and initial weights assigned to each category*

| Criterion and category | Score |
|--|-------|
| Joint involvement (pattern and distribution) | |
| 1 medium-large joint | 0.0 |
| Asymmetric medium-large joints (at least 2 joints) | 12.3 |
| Symmetric medium-large joints (at least 2 joints) | 20.0 |
| 1–3 small joints (hands/feet or wrists) | 26.2 |
| ≥4 asymmetric small joints (hands/feet or wrists) | 35.4 |
| ≥4 symmetric small joints (hands/feet or wrists) | 36.9 |
| Serology (ACPA or RF) | |
| RF negative and ACPA negative | 0.0 |
| Low-positive (RF positive and/or ACPA positive) | 27.0 |
| High-positive (RF positive and/or ACPA positive) | 43.1 |
| Acute-phase reactants (CRP or ESR) | |
| Normal | 0.0 |
| Abnormal | 7.7 |
| Duration of synovitis | |
| <4 weeks | 0.0 |
| 4–8 weeks | 10.8 |
| >8 weeks | 12.3 |

* Scores from each criterion are summed to derive the total score, which represents the probability of developing rheumatoid arthritis (RA). ACPA = anti-cyclic citrullinated protein; RF = rheumatoid factor; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate.

asymmetrically involved small joints of the hands and feet or wrists; 6) ≥4 *symmetrically involved* small joints of the hands and feet or wrists (Table 1). While a patient may fulfill more than one category, the highest category of fulfillment takes precedence for scoring.

Serology. The role of RF versus ACPA in determining probability of developing RA was reviewed. First, recent literature reviews (9,10) indicate only small systematic differences between ACPA and RF in predicting RA outcomes. Second, Phase 1 analyses did not reveal any major differences between ACPA and RF (2). Finally, the ultimate classification criteria must be applicable for use internationally, including regions where ACPA testing is not readily available. Thus, expert panel members recommended that ACPA and RF comprise a single criterion, “serology.” Since evidence indicates that prognosis varies by level of both RF and ACPA (11), the group agreed that the serology criterion should comprise 3 categories, with increasing probability of RA: abnormal result on neither test, low-level positive result on at least 1 test, or high-level positive result on at least 1 test (Table 1). A task force (KHC, TWJH, HAM, JSS, and PPT) was created to inform the definition of high and low levels for RF and ACPA.

Acute-phase response. In Phase 1, ESR and CRP had similar predictive validity for various RA outcomes. Given insufficient evidence to support the use of multiple cut points, panel members agreed that this criterion

should be represented by only 2 categories: normal (abnormal result on neither test) or abnormal (elevation of either CRP level or ESR or both), based on local laboratory standards (Table 1).

Duration. Persistence of synovitis is associated with prognosis. While most patients are currently assessed for possible RA beyond 8 weeks of symptoms, the intent of the new criteria is to encourage earlier referral, diagnosis, and treatment. Thus, the criteria should be applicable in early disease. Given the mandatory requirement for expert-determined synovitis (swelling) in at least 1 joint, the group agreed that “duration” of synovitis should be assessed based on patient self-reported signs or symptoms of synovitis (e.g., pain, swelling, tenderness) of joints that are clinically “involved,” as defined above, at the time of the physician assessment. Three criterion categories were identified, with increasing probability of RA: duration <4 weeks, 4–8 weeks, and >8 weeks (Table 1).

Refinement phase: the relative importance of each domain and category. The resultant initial weights for each of the 14 criteria categories are shown in Table 1. Using these weights, 10 of the 30 cases were scored and the rank order presented to the expert panel. The expert panel identified 2 cases that received an inappropriately low relative ranking; both cases had negative serologic results in the setting of multiple small joints involved. To address this, an additional “joint distribution” category was created: >10 joints involved, including, but not limited to, small joints of the hands and feet or the wrists. The revised criteria categories were re-weighted. The 10 cases were re-scored and re-ranked. The resultant revised rank ordering was deemed appropriate. The revised categories and their associated weights are shown in Table 2.

Post-consensus meeting modifications. The draft criteria set was reviewed with Phase 1 and 2 participants at a second in-person meeting. The revisions described below were recommended.

Consideration of erosions. Due to the desire to classify and treat individuals with RA early in their disease course in order to prevent damage, the presence of erosions should *not* be included within the classification criteria. However, the criteria do need to be applicable across the spectrum of the potential disease course. Further discussion of this issue was necessary, and was addressed in the third phase of the project (1).

Pattern and extent of joint involvement. Based on their similar weightings, the categories of small joint involvement that referred to symmetric versus asymmet-

Table 2. Revised categories and weights after inclusion of additional joint involvement category*

| Criterion and category | Score |
|--|-------|
| Joint involvement (pattern and distribution) | |
| 1 medium-large joint | 0.0 |
| Asymmetric medium-large joints (at least 2 joints) | 10.1 |
| Symmetric medium-large joints (at least 2 joints) | 16.0 |
| 1–3 small joints (hands/feet or wrists) | 21.0 |
| ≥4 asymmetric small joints (hands/feet or wrists) | 27.7 |
| 4–10 symmetric small joints (hands/feet or wrists) | 29.4 |
| >10 joints including hands/feet or wrists | 50.5 |
| Serology (ACPA or RF) | |
| RF negative and ACPA negative | 0.0 |
| Low-positive (RF positive and/or ACPA positive) | 21.8 |
| High-positive (RF positive and/or ACPA positive) | 33.6 |
| Acute-phase reactants (CRP or ESR) | |
| Normal | 0.0 |
| Abnormal | 5.9 |
| Duration of synovitis | |
| <4 weeks | 0.0 |
| 4–8 weeks | 8.4 |
| >8 weeks | 10.1 |

* Scores from each criterion are summed to derive the total score, which represents the probability of developing rheumatoid arthritis (RA). ACPA = anti-cyclic citrullinated protein; RF = rheumatoid factor; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate.

ric involvement were combined in a single category: “4–10 small joints” (Table 2).

Serology. Informed by a systematic literature review of the diagnostic properties of ACPA and RF assays (9) and input from assay makers and researchers in the field (10,12), the serology task force recommended that the local laboratory and assay upper limit of normal (ULN) be used to categorize serologic results as follows: normal = less than or equal to the ULN; low-level positive = higher than the ULN but ≤3 times the ULN; and high-level positive = >3 times the ULN. If the result was not available, it should be regarded as normal or negative. Further, patients should be scored only if there are results available for at least 1 serologic test. This recommendation may be revised once standardized units become available (for ACPA) or are universally employed (for RF).

Acute-phase response. Participants agreed that if a test were unavailable, the result should be regarded as normal. As with serology, patients should be scored only if results are available for at least 1 acute-phase reactant.

Duration. Expert panel participants noted that in performing the paired discrete choice exercise, they had used the midpoint of the 4–8-week category rather than the extreme ends of the range in making their choices. Thus, this criterion was simplified to duration <6 weeks versus ≥6 weeks.

The decision analytic software program that was used can accommodate alterations in the categories, as was necessary for combining categories in the joint involvement criterion and the duration criterion. Subsequent analysis of the cases included in the in-person consensus meeting confirmed that decisions implied from the combined categories were consistent with the decisions made at the meeting. The revised criteria set and associated weights are shown in Table 3.

Assessment of interrater reliability of case categorization. The remaining unused 56 patient case scenarios were reviewed for presence of the 2 mandatory criteria outlined above. Two cases were excluded (for both, another inflammatory arthritis condition was more likely). The remaining 54 cases were independently categorized within each of the 4 domains by 7 individuals (4 steering committee members and 2 volunteers unfamiliar with the project to date). For cases in which the ULN for the laboratory had not been provided, that information was obtained whenever possible. Where either a test result or the ULN value was not provided, the raters were asked to consider the value to be normal.

Categorization discordance arose largely from the use of the originally developed standardized template, which failed to include pertinent information on the subsequently identified criteria and categories. Refinements to the criteria and category definitions were

Table 3. Final revised categories and weights at end of Phase 2*

| Criterion and category | Score |
|--|-------|
| Joint involvement (pattern and distribution) | |
| 1 medium-large joint | 0 |
| Asymmetric medium-large joints (at least 2 joints) | 10.2 |
| Symmetric medium-large joints (at least 2 joints) | 16.1 |
| 1–3 small joints (hands/feet or wrists) | 21.2 |
| 4–10 small joints (hands/feet or wrists) | 28.8 |
| 4–10 symmetric small joints (hands/feet or wrists) | 29.4 |
| >10 joints including hands/feet or wrists | 50.8 |
| Serology (ACPA or RF) | |
| RF negative and ACPA negative | 0 |
| Low-positive (RF positive and/or ACPA positive) | 22.0 |
| High-positive (RF positive and/or ACPA positive) | 33.9 |
| Acute-phase reactants (CRP or ESR) | |
| Normal | 0.0 |
| Abnormal | 5.9 |
| Duration of synovitis | |
| <6 weeks | 0 |
| ≥6 weeks | 9.3 |

* Scores from each criterion are summed to derive the total score, which represents the probability of developing rheumatoid arthritis (RA). ACPA = anti-cyclic citrullinated protein; RF = rheumatoid factor; CRP = C-reactive protein; ESR = erythrocyte sedimentation rate.

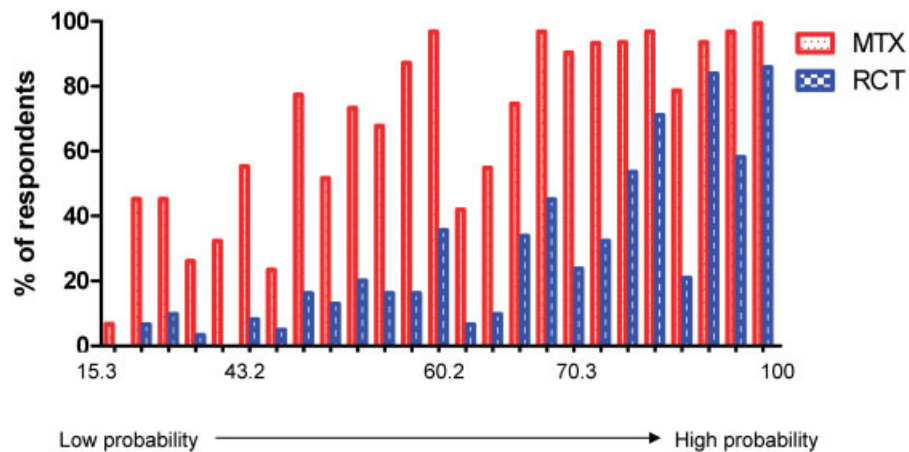


Figure 3. Proportion of respondents who would prescribe methotrexate (MTX) or another disease-modifying antirheumatic drug and proportion who would enroll the patient in a randomized controlled trial (RCT) of a biologic therapy, for clinical scenarios arranged from lowest to highest probability of “developing rheumatoid arthritis” based on the total score derived from the Phase 2 criteria set.

made, after which consensus categorization (100% agreement) was achieved for all cases.

Assessment of face and construct validity. Using the consensus categorizations and associated weights shown in Table 3, the 54 cases were scored. The resulting scores ranged from a low of 15.3 to a high of 100.0 (Figure 3). The resulting rank order, based on these scores, was reviewed by 31 Phase 1 and Phase 2 participants. They identified no substantive concerns, indicating face validity of the scoring system. The proportion who indicated that they would institute MTX or another DMARD due to a concern about risk for “developing RA,” and that they would enroll the patient into a clinical trial of a new biologic agent with inherent risks, are shown in Figure 3. As expected, the proportion of cases for which the rheumatologists would initiate MTX or another DMARD was greater than that for which they would recommend entry into a clinical trial of a biologic agent. The slight decrease in initiation of MTX or enrollment in a clinical trial at a score of ~64–65 occurred for 2 cases that had high-positive serology, but relatively few joints involved. Overall, with this exception, both proportions increased with increasing probability of RA, supporting the construct validity of the scoring system.

DISCUSSION

Using an evidence-based consensus methodology, an expert panel of rheumatologists identified 6

criteria that are important in determining the probability that a patient with undifferentiated inflammatory arthritis will develop persistent and/or erosive inflammatory arthritis that we currently consider to be RA. Two criteria were deemed essential: evidence of expert-assessed clinical joint swelling, indicating synovitis, in at least 1 joint and the absence of another condition that would better explain the patient’s presentation. The remaining 4 criteria (pattern and extent of joint involvement, serology [ACPA and/or RF], acute-phase response [ESR and/or CRP], and duration of synovitis) each contributed differently to the probability of developing RA. The relative weighting of these 4 criteria and their subcategories was determined using a new methodology with a consensus-based decision analytic software program (3). Applying the derived scoring system to a set of case scenarios produced a rank ordering close to the order determined by the clinical judgment of the group, and was consistent with the data-derived outcome of Phase 1, providing a degree of face and construct validity.

In many countries, imaging techniques, e.g., ultrasound and magnetic resonance imaging, are being used to evaluate synovitis. However, the predictive validity of synovitis detected only by imaging, and by non-experts, in the absence of clinically obvious joint swelling remains unclear. Thus, the expert panel recommended that “definite” synovitis in at least 1 joint be determined based on evidence of joint swelling on

clinical assessment by an expert assessor. Whenever possible, this should be a rheumatologist or other physician with expertise in autoimmune rheumatic diseases. Given the required presence of swelling in at least 1 joint, and the inherent imprecision of clinical determination of joint swelling, the expert panel recommended that joint involvement (i.e., number and pattern) should be assessed based on joint swelling *or* tenderness on clinical examination. Imaging modalities could be used to confirm these clinical findings.

Fulfillment of the second mandatory criterion, that clinical presentation was not better explained by another diagnosis, also requires clinical expertise. Patients in whom signs and symptoms may be explained by more than one inflammatory arthritis condition should not be assessed using these criteria until further evaluation has taken place. It was not the group's intent to imply that specific investigations should be performed to rule in or rule out alternative inflammatory arthritides or other diagnoses. Rather, the intent was to ensure that, in the physician's opinion, no other condition better explain the situation.

The one patient factor that was surprisingly not included was duration of morning stiffness. Many rheumatologists hold strong opinions regarding the value of this patient-reported symptom in making a diagnosis and determining a management approach. However, in the Phase 1 study, duration of morning stiffness >1 hour versus <1 hour did not discriminate between patients who did and those who did not receive MTX within a year of diagnosis. Furthermore, while this symptom can reflect the burden of inflammation, on an individual patient level it does not discriminate among the inflammatory arthritides, or even between inflammatory and noninflammatory disease. Thus, morning stiffness was not included.

During the in-person consensus meeting, expert panel members agreed that symmetry of joint involvement was important in determining the probability of RA. However, the weights derived for symmetric versus asymmetric involvement, for medium-large joints as well as for small joints, were remarkably similar. This suggests that our decision-making is not based on symmetry once other factors, e.g., the number and type of joints involved and the serologic results, are taken into consideration. Again, this was consistent with findings in the Phase 1 analyses.

Phase 2 used a new methodology to derive consensus among expert clinicians, which is more transparent and flexible than usual Delphi consensus approaches. This method is also cognitively and timewise

less burdensome than other methods, with a high degree of validity and reliability (3). As with all consensus methodologies, the result is dependent on the expertise and information of the expert panel.

In summary, Phase 2 utilized a novel consensus methodology in which decision analysis was integrated to derive a preliminary set of criteria and criteria categories, with associated weights representing their relative importance. The process was informed by Phase 1 data and results and the published literature, wherever possible. Together with Phase 1, this work informed the final phase of criteria development, outlined in the companion report (1), in which the final criteria set, including the cut point to be used to define definite RA, were determined and preliminary validation performed.

ACKNOWLEDGMENTS

The authors would like to thank the following individuals for their support of this project: Samra Mian for her stewardship of the collection and refinement of the patient case scenarios, Amy Miller and Regina Parker for their outstanding organizational support and for keeping us on track, Alison Barber for her assistance during the consensus meeting, and Drs. Steven Vlad and Gunnar Tomasson for their participation in the assessment of the reliability of case categorization.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Hawker had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Neogi, Aletaha, Silman, Naden, Felson, Aggarwal, Birnbaum, Bykerk, Combe, Costenbader, Dougados, Emery, Hazes, Huizinga, Kay, Khanna, Kvien, Moreland, Nair, Smolen, Stanislawska-Biernat, Vencovský, Wolfe, Hawker.

Acquisition of data. Neogi, Aletaha, Silman, Naden, Aggarwal, Bingham, Birnbaum, Burmester, Bykerk, Combe, Costenbader, Dougados, Emery, Hazes, Huizinga, Kavanaugh, Kay, Kvien, Laing, Liang, Ménard, Ringold, Smolen, Stanislawska-Biernat, Tak, Upchurch, Vencovský, Hawker.

Analysis and interpretation of data. Neogi, Aletaha, Silman, Naden, Aggarwal, Bingham, Birnbaum, Burmester, Bykerk, Cohen, Combe, Dougados, Emery, Ferraccioli, Hazes, Hobbs, Huizinga, Kay, Laing, Mease, Ménard, Moreland, Pincus, Smolen, Stanislawska-Biernat, Symmons, Tak, Upchurch, Vencovský, Hawker.

REFERENCES

1. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, et al. 2010 Rheumatoid Arthritis Classification Criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* 2010;62:2569–81.
2. Funovits J, Aletaha D, Bykerk V, Combe B, Dougados M, Emery P, et al. The 2010 American College of Rheumatology/European

- League Against Rheumatism classification criteria for rheumatoid arthritis: methodological report Phase 1. *Ann Rheum Dis* 2010; 69:1589–95.
3. Hansen P, Ombler F. A new method for scoring multi-attribute value models using pairwise rankings of alternatives. *J Multi-Crit Decis Anal* 2009;15:87–107.
 4. De Coster C, Noseworthy T. Improving wait times in the referral-consultation process: WCWL priority referral scores. *Proceeding of the Taming of the Queue VI Conference, Canadian Policy Research Networks*. 2009 Mar 27–28; Ottawa, Ontario, Canada.
 5. De Coster C, Fitzgerald A, Noseworthy T. Developing priority-setting referral tools for medical sub-specialties. *Proceedings of the Annual Canadian Association for Health Services and Policy Research Conference*. 2008 May 26–28; Gatineau, Quebec, Canada.
 6. Fitzgerald A, De Coster C, Naden R, Noseworthy T, Western Canada Wait List Project Rheumatology Clinical Panel. Priority-setting for referrals from primary care physicians to rheumatologists [abstract]. *Arthritis Rheum* 2008;58 Suppl:S884.
 7. Naden R, Paterson R, Hansen P, Barber A, Ombler F, Stewart R, et al. Assigning clinical priority: a systematic methodology. *Proceedings of the 6th International Conference on Priorities in Health Care, International Health Economics Association*. 2006 Sep 20–22; Toronto, Ontario, Canada.
 8. Barber A, Doolan Noble F, Stewart R, Wilkins G, Naden R, North D. Prioritisation for coronary artery bypass surgery: can the process be improved? *Proceedings of the 5th International Conference on Priorities in Health Care, International Health Economics Association*. 2004 Nov 3–5; Wellington, New Zealand.
 9. Aggarwal R, Liao K, Nair R, Ringold S, Costenbader KH. Anti-citrullinated peptide antibody assays and their role in the diagnosis of rheumatoid arthritis [review]. *Arthritis Rheum* 2009; 61:1472–83.
 10. Van der Linden MP, van der Woude D, Ioan-Facsinay A, Levarht EW, Stoeken-Rijsbergen G, Huizinga TW, et al. Value of anti-modified citrullinated vimentin and third-generation anti-cyclic citrullinated peptide compared with second-generation anti-cyclic citrullinated peptide and rheumatoid factor in predicting disease outcome in undifferentiated arthritis and rheumatoid arthritis. *Arthritis Rheum* 2009;60:2232–41.
 11. Mjaavatten MD, van der Hiejde DM, Uhlig T, Haugen AJ, Nygaard H, Sidenvall G, et al. Levels of anti-CCP and IgM RF predict persistent arthritis in patients with very early undifferentiated arthritis (UA) [abstract]. *Arthritis Rheum* 2009;60 Suppl:S123.
 12. Van Gaalen FA, Visser H, Huizinga TW. A comparison of the diagnostic accuracy and prognostic value of the first and second anti-cyclic citrullinated peptides (CCP1 and CCP2) autoantibody tests for rheumatoid arthritis. *Ann Rheum Dis* 2005;64:1510–2.