Preliminary Validation of Clinical Remission Criteria Using the OMERACT Filter for Select Categories of Juvenile Idiopathic Arthritis

CAROL A. WALLACE, ANGELO RAVELLI, BIN HUANG, and EDWARD H. GIANNINI

ABSTRACT. ***Objective***. To begin the validation process of the preliminary criteria for inactive disease (ID), clinical remission on medication (CRM), and clinical remission off medication (CR) in children with select forms of juvenile idiopathic arthritis (JIA).

Methods. We used the OMERACT filter paradigm to estimate the validity of the criteria within each of the filter's 3 components: truth, discrimination, and feasibility, in 5 categories of JIA: systemic arthritis, persistent and extended oligoarthritis, and rheumatoid factor-positive and negative polyarthritis. Data sources for determining validity estimates included a Delphi questionnaire survey sent to 246 pediatric rheumatologists in 34 countries, a consensus conference attended by 20 senior pediatric rheumatologists representing 9 countries, a retrospective chart review of 437 patients with JIA from 3 tertiary care clinics who had been followed between 4 and 22 years, and the literature.

Results. Truth component: face and content validity. These aspects of validity were largely established via the Delphi questionnaire exercise and the consensus conference. Using an 80% consensus level, participants felt that a set of non-redundant variables could effectively differentiate the clinical states of ID, CRM, and CR. Criterion validity could not be irrefutably established because no gold standard for inactive disease exists for JIA. As an alternative, published investigations of remission in JIA were used to estimate concurrent and convergent validity, as surrogates for criterion validity and as indicators of overall construct validity. Correlational analyses revealed the new criteria to have good construct validity. Discrimination component: the criteria demonstrated moderate to high levels of classification, prognosis, and responsiveness (sensitivity to change) using data from the chart review. Patients who were able to attain CR remained disease-free for substantially longer periods than did those who attained only ID or CRM. Responsiveness was evidenced by the ability of the criteria to allow movement of most patients between the disease states, consistent with what is known of the course of the disease. Feasibility component: Results of the Delphi and consensus conference produced a set of criteria that are easily, quickly, and inexpensively completed in the physician's office, and present minimal or no risk to the patient.

Conclusion. The preliminary criteria demonstrated moderate to excellent validity characteristics in some, but not all components of the OMERACT filter. Prospective validation studies are under way. (First Release Feb 15, 2006; J Rheumatol 2006;33:789–95)

Key Indexing Terms:
JUVENILE IDIOPATHIC ARTHRITIS VALIDATION REMISSION CRITERIA

From the Division of Immunology, Rheumatology, and Infectious Disease, Children's Hospital and Regional Medical Center; and Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington, USA; the IRCCS G. Gaslini, Pediatria II, Genoa, Italy; and the Center for Epidemiology and Biostatistics, and William S. Rowe Division of Rheumatology, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio, USA.

Supported by grants from Pfizer Inc., Centocor Inc., the Howe Endowment for JRA Research, and NIAMS P60AR47784.

C.A. Wallace, MD, Associate Professor, Division of Immunology, Rheumatology, and Infectious Disease, Children's Hospital and Regional Medical Center, Department of Pediatrics, University of Washington School of Medicine; A. Ravelli, MD, Dirigente Medico, IRCCS G. Gaslini, Pediatria II; B. Huang, PhD, Assistant Professor, Center for Epidemiology and Biostatistics; E.H. Giannini, MSc, DrPH, Professor, William S. Rowe Division of Rheumatology, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine.

Address reprint requests to Dr. C.A. Wallace, Division of Rheumatology, Children's Hospital and Regional Medical Center, 4800 Sandpoint Way NE, Seattle, WA 98105, USA. E-mail: cwallace@u.washington.edu

Recent clinical trials in both adult rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA) of new immune response modifiers (biologicals) and new combinations of therapeutic agents have shown that therapeutically induced disease quiescence may be achievable in a substantial number of patients[1-9]. Clinicians seem to agree that brief periods of *inactive disease* (ID) prior to disease flare represent little more than a temporary quieting of signs and symptoms of arthritis. Longer, continuous periods of ID have been classified as either *complete clinical response* or, if all antiarthritis medications have been discontinued, *remission*[10-13]. However, there is little agreement on criteria for ID or on how long the patient must remain in ID to be classified as being in a state of complete clinical response. Similarly, the length of time a patient must remain off all antiarthritis medications, while maintaining ID, in order to be classified as being in remission

is unresolved. While all these terms convey a low level or complete absence of clinically apparent disease, there are currently no definitive laboratory tests or biological markers to serve as gold standards to establish incontrovertibly that the disease is truly "biologically inactive."

The term "remission" has been used with considerable latitude in the literature, particularly in children, and no set of criteria for remission has been systematically validated in JIA. Not surprisingly, there are scant data to allow determination of the clinical relevance of having reached a state of remission. Does remission imply a cure, or some probability of flare within 5 or 10 years, or something else? This question has not been addressed through adequately designed clinical investigations that would provide an evidence based resolution.

To address these issues an international project was begun in 2002 that has resulted in preliminary criteria for ID, clinical remission on medication (CRM; a term thought to be more descriptive than complete clinical response), and clinical remission off medication (CR), implying inactive disease off medication for a protracted period for 5 different categories of JIA. These categories include systemic arthritis, persistent and extended oligoarthritis, and rheumatoid factor-positive (RF+) and negative (RF–) polyarthritis. The chief goal of this effort was to begin the process of establishing a common vocabulary for use by clinicians, researchers, regulatory agencies, and sponsors to describe the same clinical states. Results of the effort to establish consensus based preliminary criteria have been published[14], and are shown in Table 1. (Although we consider the criteria in Table 1 "preliminary," we have used the term "criteria" for purposes of brevity throughout this report.)

We applied the Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) filter[15] in an attempt to begin the process of validating the newly developed criteria described by Wallace, *et al*. The filter is considered a user-friendly paradigm that facilitates the capture of 3 essential components of an outcome measure or set of criteria, and follows the guidelines for validity formulated by Tugwell and Bombardier for measurement methodology focused on trials[16]. To our knowledge the work described here represents the first effort to use the OMERACT filter in a pediatric rheumatic illness. Results of these validation efforts are the subject of this report.

MATERIALS AND METHODS

The OMERACT filter and terminology used in validation studies. The OMERACT filter was developed to simplify and clarify the vast array of nonstandardized terminology used in validation studies. It encompasses 3 components of clinimetrics: truth, discrimination, and feasibility.

Truth

Truth refers to whether a variable or set of criteria actually measures what it is designed to measure in an unbiased, relevant way. Truth captures 4 aspects of validity: face, content, criterion, and construct.

Face validity. Face validity provides evidence that the criteria are sensible (on their face) to practitioners who will use them in the relevant field.

Content validity. Content (or comprehensiveness) validity asks: (1) if the criteria include the most relevant, crucial clinical characteristics of the condition under study so that a valid, comprehensive assessment of subjects can be made; and (2) if each variable in the criteria contributes something distinct and important to facilitate classification of the patient.

Criterion validity. Criterion validity asks if the criteria under investigation produce the same or similar results as does a gold standard criterion. Because gold standards do not exist in many situations, authors frequently rely on measures of concurrent validity as a surrogate. Concurrent validity asks if the criteria agree "concurrently" with other criteria designed to measure the same or similar constructs [e.g., whether patients with similar clinical pictures are classified into the same (or similar) categories using both sets of criteria].

Construct validity. Construct validity, the fourth aspect of the truth domain, is a broad view of the criteria, and asks, "Overall, do the criteria adequately measure the underlying construct?" There is no single measure of construct validity. Rather, it is based on the accumulation of knowledge about the criteria and their relationship to other tests or criteria. Construct validity can be assessed by evaluating convergent and divergent validity. Convergent validity describes how well the results produced by the new criteria (e.g., remis-

Table 1. Preliminary criteria for inactive disease and clinical remission of JIA. Adapted from[14] Wallace, *et al,* J Rheumatol 2004; 31:2290-4.

Inactive disease
1. No joints with active arthritis*,**
2. No fever, rash, serositis, splenomegaly, or generalized lymphadenopathy attributable to JIA
3. No active uveitis (to be defined)
4. Normal ESR or CRP (if both are tested, both must be normal)
5. Physician's global assessment of disease activity indicates no disease activity (i.e., best score attainable on the scale used

Clinical remission

Two types of clinical remission are proposed:
1. Clinical remission on medication. The criteria for inactive disease must be met for a minimum of 6 continuous months while the patient is on medication
2. Clinical remission off medication. The criteria for inactive disease must be met for a minimum of 12 continuous months while off all antiarthritis medications

* As defined by ACR: A joint with swelling not due to bony enlargement or, if no swelling is present, limitation of motion accompanied either by pain on motion and/or tenderness. ** Isolated finding of pain on motion, tenderness, or limitation of motion on joint examination may be present only if explained by either prior damage attributable to arthritis that is now considered inactive, or nonrheumatological reasons such as trauma.

sion) converge (correlate) with other systems (or criteria) that were designed to measure the same construct, or diverge from (do not correlate with) results that use criteria designed to test some other distinct concept. Together, convergent and divergent validity provides correlational measures for the overarching concept of construct validity.

Data sources for determining the truth domain. To estimate the face and content validity (as well as feasibility, discussed below), a Consensus Working Group was formed that conducted a series of Delphi questionnaire surveys[17] sent to 246 experienced pediatric rheumatologists in 34 countries. Results of the Delphi survey provided background information for a consensus conference attended by 20 senior pediatric rheumatologists from 9 countries that was held in 2003. Nominal group technique[17] was used to arrive at the preliminary criteria (Table 1). We also used the literature to estimate face and content validity from other studies of remission in JIA (Table 2).

To assess criterion validity (using the surrogate concurrent validity) and overall construct validity (as measured by convergent and divergent validity) we used criteria from 2 sources in the literature concerning remission in juvenile arthritis[11,12] (Table 2), and a newly collected data set based on a retrospective chart review of patients with JIA from 3 clinics: Seattle Children's Hospital & Regional Medical Center, Seattle, WA, USA; IRCCS G. Gaslini, Genova; and IRCCS Policlinico S. Matteo, Pavia, Italy.

Subjects were identified for inclusion from medical charts if, first, a diagnosis of JIA by the International League of Associations for Rheumatology classification[18] of persistent or extended oligoarthritis, RF+ or RF– polyarthritis, or systemic arthritis had been made. (The classification system of JIA was not yet in existence when the patients included in the chart review were originally diagnosed. However, information from the medical charts allowed classification of each patient into the new nomenclature in nearly all cases.) Secondly, each subject had to have been followed for a minimum of 4 years. Data were extracted using standardized case report forms and were reviewed for quality assurance by 2 of the authors (CAW, AR). A total of 437 patients were identified and reviewed. Results of the clinical findings of the chart review are reported in greater detail elsewhere[19].

Discrimination

Discrimination is the second component of the OMERACT filter and asks if the measures or criteria are able to distinguish between states of interest. The term captures the concepts of classification, prognosis, responsiveness (sensitivity to change during a relevant time interval), and reliability (reproducibility).

Classification. Classification of a patient as being in a state of inactive versus active disease is of paramount importance for the validity of these criteria, since the other states (CRM, CR) depend on correct classification of the patient's disease activity state. Prognosis or predictive validity asks if the states defined by the criteria (e.g., ID, CRM, CR) actually have clinical relevance. That is, does the likelihood of future events (e.g., disease flare) vary among the disease states that patients fall into by use of the criteria?

Responsiveness. Responsiveness refers to the ability of subjects to change disease states (as defined by the criteria) over time. Practitioners are well aware that children with rheumatic disease may experience extended periods of ID, only to be followed by rapid return of active disease. Thus, criteria for describing rheumatic disease states must allow patients to shift from one disease state to another. Reliability is an attribute of criteria that measures how concordant or discordant different assessors are when classifying patients with the same clinical picture into the same disease state category.

Data sources for estimating aspects of validity associated with discrimination. To assess classification, prognosis, and responsiveness, we utilized the data from the chart review described above. To assess reliability, we searched the literature using PubMed for estimates of reliability of each variable included in the criteria.

Feasibility

Feasibility is the third component of the OMERACT filter. Feasibility is multidimensional and refers to how practical it is for the measurement of the outcome to be carried out quickly, easily, using a simple scoring method, and at a minimal cost in a physician's office. Feasibility also incorporates the risk associated with the procedures to be used in the determination of the variable's metric. Variables that necessitate the exposure of patients to procedures of more than minimal risk (either immediate or delayed) are not likely to be considered appropriate for inclusion using the OMERACT filter.

Data sources for determining feasibility. Aspects of feasibility were addressed in the Delphi exercise and discussed further during the consensus conference described above. We relied on the results of these exercises to create a set of criteria that (1) can be conducted quickly and easily in the physician's office, (2) are inexpensive, and (3) present minimal or no risk to the patient.

Weighting variables in criteria

Bellamy[20] has suggested that weighting of variables in composite indices rep-

Table 2. Selected other criteria for remission as evidence of face and content/comprehensive validity of the preliminary proposed by Wallace, *et al*[14].

FDA criteria, 1999[10]
Defines 2 states, complete clinical response (CCR) and remission
Remission: identical to CCR, but achieved while off all antirheumatic drugs
CCR: While on drug, 6 consecutive months of morning stiffness < 15 min duration, no active synovitis, no extraarticular features (including fever, serositis, adenopathy, hepatosplenomegaly, rash, uveitis), and normal laboratory variables (including ESR, platelets, WBC), and where applicable, no ongoing structural damage while continuing therapy. Residual damage from prior disease, including extraarticular manifestations, is acceptable in meeting criteria for CCR
Oen, *et al* criteria, 2002[12]
Remission: Absence of active arthritis while off all medications for at least 2 years
Fantini, *et al* criteria, 2003[11]
Inactive disease: No active joints, but still undergoing treatment
Remission: No signs of disease activity (active joints and/or positive laboratory tests) in the absence of antirheumatic therapy, including local corticosteroid injection and NSAID, for at least 6 mo (12 mo for oligoarthritis if a joint had been previously injected with long-acting corticosteroids)
Flato, *et al* criteria, 2003[13]
Remission: Five or more of the following criteria must be met for at least 2 years: morning stiffness ≤ 15 min, no fatigue, no joint pain, tenderness, swelling, no swelling of tendon sheaths, ESR < 20. Off all antirheumatic medications for the past 2 yrs

ESR: erythrocyte sedimentation rate; WBC: white blood cells; NSAID: nonsteroidal antiinflammatory drugs.

resents a rather complex undertaking and that this area of index development requires considerable further investigation. We had no evidence at this preliminary stage on which to base a weight factor on any variable in the criteria, and therefore made no attempt to place differing measures of importance on those included.

RESULTS

Measures of truth: Face and content validity. These aspects of truth were largely established via the Delphi questionnaire exercise and the consensus conference described above. Using a consensus level of 80%, participants felt that, collectively, the variables shown in Table 1 were those considered crucial for legitimate classification of a patient as being in a state of inactive disease.

Additional evidence of face and content validity can be found in the efforts that were part of establishing construct validity (below), in which other sets of criteria for remission were used[11] (see Table 2 for remission criteria used by other investigators). The Wallace criteria contain many of the same clinical features and physician and laboratory assessments as do criteria employed by other investigators. Thus, it appears that the Wallace criteria represent a broad cross-section of possible disease manifestations and evaluations that have been considered important to assess by earlier investigators as well as by consensus conference participants, and appear to have high face and content validity.

Criterion and construct validity. In the absence of a gold standard, we attempted to establish a measure of criterion validity by use of the surrogate concurrent validity. Further, concurrent validity is considered a correlational measure of convergent validity. Thus, by performing correlational analyses, we were able to determine if our criteria "converged" or "diverged" from other similar criteria designed to define remission (and thereby estimate overall construct validity). We searched the literature for reports that focused on describing remission in JIA, and that contained criteria for classification that might be used to classify patients from our retrospective chart review. The studies by Oen, *et al* and by Fantini, *et al*[11,12] (Table 2) appeared to be the most appropriate for this exercise.

In the cross-sectional study by Oen, *et al*[12], the criterion for remission is simply "no active arthritis while off medication for a minimum of 2 years," without specification of what is meant by "no active arthritis," and the status of uveitis is not mentioned. If one assumes that no active arthritis means inactive disease by our criteria, then a comparison can be made. If the time off all medication while maintaining ID is increased to 2 years in our patient cohort, as suggested by Oen, then 28.4% (95% CI 24.15%–32.60%) of our patients would have achieved CR. This compares to 39% (given N = 392, 95% CI 34.17%–43.83%) in the Oen study (chi-square comparison between the 2 rates of CR, p = 0.0007). However, if both the Wallace and Oen criteria are applied to our data, there is very good agreement between the 2 (kappa coefficient of 0.66, 95% CI 0.60–0.73.) In fact, because the Oen criteria are more stringent, all subjects who were classified as failures (to attain CR) according to the Wallace criteria also were considered failures by the Oen criteria. Among those who were classified as failures according to the Oen criteria, 22% would be classified as successes according to the Wallace criteria.

Fantini's criteria[11] differ as well. To achieve CR, patients with either systemic or polyarthritis had to remain disease-free for 6 months off medication, and for 12 months for oligoarthritis if intraarticular steroids had been used. The status of uveitis is not mentioned. Nevertheless, the Fantini criteria come closest to ours, as they include "no signs of active disease and/or positive laboratory tests." If one assumes that by "no signs of active disease" systemic manifestations of disease are included, the criteria are similar. Fantini found 42% (given N = 683, 95% CI 38.3%–45.7%) of his patients achieved remission by their final visit, but 58% never achieved remission during a median followup of 6.2 years. If our criteria for clinical remission off medication are lowered to 6 months off medication for the non-oligo groups, and remain at 12 months for the oligo group (identical to Fantini's approach), then 48% (95% CI 43.3%–52.7%) of our patient cohort achieved CR (chi-square between the 2 rates of CR is NS, p = 0.06). If both the Wallace and Fantini criteria are applied to our data, the result suggests excellent agreement (kappa coefficient, 0.94; 95% CI 0.90–0.97). Among all those who failed to reach CR according to the Wallace criteria, only 5.8% were considered to be in CR according to Fantini criteria. All patients who achieved CR according to the Wallace criteria are also considered to be in CR by the Fantini criteria.

From these reviews and analyses we conclude that preliminary estimates of overall construct validity of the Wallace criteria, as evidenced by moderate to excellent concurrent and convergent validity estimates, is likely to be quite high.

Measures of discrimination: Classification, prognosis, responsiveness, and reliability. Given the results of the consensus conference and the similarity of the Wallace criteria to other systems of classification, these criteria likely perform quite well in distinguishing a child with very active versus no clinically apparent disease. However, misclassification of some patients near the boundary of active and inactive disease cannot be ruled out due to concerns about reliability of assessment of some physician-determined variables (below). This study is not able to estimate the size of this potential classification bias.

We considered the issue of prognosis, or clinical relevance, of the criteria to be critically important for determining their discriminatory ability under the OMERACT filter paradigm. For example, if children who attained CR were as likely to relapse as quickly as those who attained only ID, then the criteria for CR could be criticized for not being stringent enough, with little prognostic validity.

The data source for this analysis was the chart review of 437 patients described above. Table 3 provides evidence of the prognostic value of the various disease states. Those patients

Table 3. Survival of the disease states described by the preliminary criteria.

Disease State Attained	No. of Episodes	No. of Episodes/ Percentage of Episodes Relapsing to Active Disease (95% CI)	Time to Relapse (mo), mean ± SD median (IQR)
Inactive disease only[a]	292	249, 95.77 (93.32, 98.22)	5.04 ± 2.98 5 (3–6)
CRM only[b]	358	288, 91.43 (88.34, 94.52)	17.11 ± 12.18 14 (9–21)
CR only[c]	127	86, 83.50 (76.33, 90.66)	33.45 ± 20.32 29 (17–45)
CR after CRM	101	62, 83.78 (75.39, 92.18)	48.84 ± 24.96 41 (32–56)
Statistical comparisons over 4 groups	—	jonckheere-Terpstra test $p < 0.0001$	Kruskal-Wallis test $p < 0.0001$

[a] Episode of inactive disease did not progress to clinical remission on or off medication. [b] CRM: clinical remission on medication. [c] CR: clinical remission off medication, without first achieving CRM. IQR: interquartile range.

who attained ID but not clinical remission on medication flared earlier than those achieving CRM. Similarly, those patients that achieved CR tended to flare later than those who only achieved CRM. In consideration of the data in Table 3, with significant differences being found in prognosis among the various disease states, there appears to be prognostic justification for classifying patients into the categories described by the criteria.

Responsiveness (sensitivity to change) of the criteria was assessed using data from the retrospective chart review. The longitudinal nature of patient followup, ranging from a minimum of 4 years to a maximum of 22 years (median 6.5), permitted assessment of the frequency with which patients moved in and out of the various disease states described by the criteria. Table 4 demonstrates the responsiveness of the criteria. These data indicate that the criteria permit movement both forward and backward through the disease states, consistent with the known natural history of the disease. In many cases individual subjects experienced multiple episodes of ID, CRM, and CR, again suggesting a high level of responsiveness. As expected, there were substantial differences in the rates of relapse and duration of time spent in each disease state among the categories of JIA. These differences have been described in greater detail by Wallace, *et al*[19].

Table 4. Movement of patients among the disease states of inactive disease (ID), clinical remission on medication (CRM), and clinical remission off medication (CR).

	Episode					
	1	2*	3	4	5	> 5
(No. of patients/no. of episodes)						
Disease state						
ID	391	234	127	67	29	30
CRM	189	67	27	11	1	1
CR	160	34	0	0	0	0

* 234 patients had at least 2 episodes of ID; 67 patients had at least 2 episodes of CRM, and 34 patients had 2 episodes of CR.

Reliability (reproducibility). When composite criteria are used to arrive at a single determination (e.g., ID versus active disease), each variable in the criteria must demonstrate reliability, since a mistake in the assessment of any one variable may lead to misclassification of the patient. Reliability of some (ESR, CRP, fever) but not all variables in the Wallace criteria is known. We realize that a standard definition of inactive uveitis does not exist at this time, and that a consensus process is under way to develop valid criteria. Therefore, estimates of reliability of this variable are not yet possible.

We searched the literature for legitimate estimates of the reliability among practitioners for the active joint count, the determination of systemic manifestations, and global assessments, and found 2 such reports[21,22]. These articles both examined interobserver reliability of articular examination. One investigation described the results of 2 experienced examiners who agreed on a standardized format of examination and separately spent 30 minutes on each joint examination of 20 patients with JIA. This study reported substantial agreement in both swelling and limited range in the examinations by these 2 individuals. The second study found only moderate agreement for the 2 variables (joint swelling and limitation of range) between 4 examiners (one experienced and 3 junior) who spent 14 minutes per examination without prior discussion of standardized joint examination. Our experience in conducting training sessions for clinical investigators who are preparing to participate in industry-sponsored drug studies is not as favorable as these 2 reports, and has shown considerable differences among pediatric rheumatologists in assessing the number of active joints and in global

assessments (unpublished, proprietary data). At this time, therefore, estimation of overall reliability is not complete and will have to be accomplished during prospective efforts.

Measures of feasibility: ease of use, low cost, low risk. Each of the procedures necessary for assessment of the variables in the criteria for determining ID is part of the routine rheumatological assessment of a child with JIA. None requires specialized equipment, and the cost of evaluating each variable is low. Venipuncture for the determination of ESR and CRP is the only procedure associated with minimal risk; the remainder of procedures are of negligible risk. Imaging methods such as magnetic resonance imaging (MRI) that could potentially identify ongoing joint destruction and/or cartilage damage represent more than minimal risk (the requirement for sedation in young children) and considerable expense. Further, the necessary instrumentation may not be widely available in most clinics. For these reasons, the present criteria were considered to fulfill the feasibility component of the OMERACT filter.

DISCUSSION

The OMERACT filter for assessment of the validity of outcome measures in rheumatology has been used previously to assess quality of life measures[23], scoring methods for radiological change in RA trials[24], comparison of scoring systems for plain films and MRI in ankylosing spondylitis and adult RA[25-27], and measurements of adult RA disability and damage using MRI[28,29].

We chose to use it in this initial attempt to validate the criteria for inactive disease, clinical remission on medication, and clinical remission off medication for JIA for reasons of clarity, simplicity, and because of its demonstrated usefulness and acceptance in other rheumatic diseases. The filter allowed us to observe the stronger and weaker aspects of validity of the criteria within the filter's 3 components despite the fact that our data sources were limited to consensus formation, a retrospective chart review, and the literature. For example, all components of the feasibility component of the current criteria appear to be met very well. Similarly, the face and content aspects of validity of the truth component seem quite high. We are confident that these criteria will be considered by practitioners as able to measure what they are designed to measure, using a comprehensive array of non-redundant variables.

Criterion validity was unable to be definitively determined because there is no gold standard for the determination of inactive disease or remission in JIA. Nevertheless, concurrent/convergent validity estimates and correlational analyses using data from published studies suggest rather high levels of overall construct validity. Two other sets of remission criteria are present in the literature. The large study by Flato and coworkers[13] is difficult to use as a comparator as this study included morning stiffness, fatigue, and whether there was swelling of the tendon sheaths. These variables were not collected in our chart review. We considered using the US Food and Drug Administration (FDA) criteria[10] for remission in JIA; however, the guidance document that describes these criteria contains no data, and the criteria have never been validated. Further, due to the need for repeated radiographs in the FDA criteria, we felt the FDA criteria were inappropriate for use here. Other measures of concurrent and convergent validity are possible, such as correlation with a pain scale or functional ability tool. Unfortunately, standardized scales did not exist in our database, and this useful exercise in establishing overall construct validity will have to be carried out during prospective validation.

As with the truth component, the evidence for validity of some, but not all aspects of the discrimination component are quite strong. Clearly, the criteria can distinguish between those patients with very active disease versus very minimal disease. Classification validity is problematic, however, in patients near the interface of the 2 disease states due to the potential non-reliability of some measures, such as whether the patient has any active joints. Analysis of prognosis revealed that those patients who achieved a state of CR demonstrated longer periods of continued ID than did those who attained only ID or CRM. Thus, the disease categories that follow the onset of ID appear to have some prognostic validity.

A high degree of responsiveness of the criteria was observed as evidenced by movement of patients through the disease states within a reasonable time period. Still, the data must be interpreted with a degree of caution. Responsiveness could be a function of low reliability of assessment of the variables within the criteria (in addition to the other usual drawbacks of data from a retrospective chart review). Prospective efforts are being designed to establish legitimate estimates of the reliability of the variables included in the criteria.

Recent literature reports indicate that radiological progression in adult RA can occur among patients classified as in remission by the ACR preliminary criteria[30]. We therefore strongly emphasize that the criteria presented here for validation refer to clinical remission rather than to biological remission. Criteria that include radiographs or MRI studies, for the sole purpose of determining clinical remission, would fail the OMERACT filter due to insensitivity (radiographs in children), unacceptable risk, and high cost. Nevertheless, investigators and clinicians are keenly interested in knowing the absolute truth about disease activity. Ongoing studies of RNA expression and proteomics may eventually provide an accurate determination of disease inactivity thereby avoiding imaging studies.

The OMERACT filter concept is iterative, and criteria are forever considered "preliminary"[15]. The classification of these criteria as preliminary is certainly deserved here. And yet, this effort represents a start at standardizing the language and the assessments used to describe children with JIA who have low level or no active disease. The 3 pediatric rheumatology research networks [Pediatric Rheumatology Collabor-

ative Study Group (PRCSG), the Childhood Arthritis and Rheumatology Research Alliance (CARRA), and the Pediatric Rheumatology InterNational Trials Organization (PRINTO)] are currently collaborating to prospectively validate the criteria for clinical remission of JIA studied here.

ACKNOWLEDGMENT

The authors thank the additional members of the Consensus Working Group who are not named authors on this publication: Alberto Martini, MD; Nicola Ruperto, MD, MPH; and Daniel Lovell, MD, MPH. We also thank Marcia Bandeira, MD, and Kathy Salmonson, RN, for their assistance in review of medical charts.

REFERENCES

1. Boers M, Verhoeven AC, Markusse HM, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet 1997;350:309-18.
2. Lovell DJ, Giannini EH, Reiff A, et al. Long-term efficacy and safety of etanercept in children with polyarticular-course juvenile rheumatoid arthritis: interim results from an ongoing multicenter, open-label, extended-treatment trial. Arthritis Rheum 2003;48:218-26.
3. Furst DE, Breedveld FC, Kalden JR, et al. Updated consensus statement on biological agents for the treatment of rheumatoid arthritis and other immune mediated inflammatory diseases. Ann Rheum Dis 2003;62 Suppl 2:ii2-9.
4. Klareskog L, van der Heijde D, de Jager JP, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. Lancet 2004;363:675-81.
5. Breedveld FC, Emery P, Keystone E, et al. Infliximab in active early rheumatoid arthritis. Ann Rheum Dis 2004;63:149-55.
6. Breedveld FC, Kalden JR. Appropriate and effective management of rheumatoid arthritis. Ann Rheum Dis 2004;63:627-33.
7. Kalden JR, Smolen JS, Emery P, et al. Leflunomide in combination therapy. J Rheumatol 2004;31 Suppl 71:25-30.
8. Landewe RB, Boers M, Verhoeven AC, et al. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis Rheum 2002;46:347-56.
9. Quinn MA, Conaghan PG, O'Connor PJ, et al. Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal: results from a twelve-month randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2005;52:27-35.
10. US Food and Drug Administration. Guidance for industry: Clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis. Washington, DC: US Government Printing Office; 1999.
11. Fantini F, Gerloni V, Gattinara M, Cimaz R, Arnoldi C, Lupi E. Remission in juvenile chronic arthritis: a cohort study of 683 consecutive cases with a mean 10 year followup. J Rheumatol 2003;30:579-84.
12. Oen K, Malleson PN, Cabral DA, Rosenberg AM, Petty RE, Cheang M. Disease course and outcome of juvenile rheumatoid arthritis in a multicenter cohort. J Rheumatol 2002;29:1989-99.
13. Flato B, Lien G, Smerdel A, et al. Prognostic factors in juvenile rheumatoid arthritis: a case-control study revealing early predictors and outcome after 14.9 years. J Rheumatol 2003;30:386-93.
14. Wallace CA, Ruperto N, Giannini EH, for the CARRA, PRINTO and PRCSG. Preliminary criteria for clinical remission for select categories of juvenile idiopathic arthritis. J Rheumatol 2004;31:2290-4.
15. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198-9.
16. Tugwell P, Bombardier C. A methodologic framework for developing and selecting endpoints in clinical trials. J Rheumatol 1982;9:758-62.
17. Delbecq AL, Van de Ven AH, Gustafson DH. Group techniques for program planning. A guide to nominal group and delphi processes. Middleton, WI: Green Briar Press; 1986.
18. Petty RE, Southwood TR, Manners P, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31:390-2.
19. Wallace CA, Huang B, Bandeira M, Ravelli A, Giannini EH. Patterns of clinical remission in select categories of juvenile idiopathic arthritis. Arthritis Rheum 2005;52:3554-62.
20. Bellamy N. Clinimetric concepts in outcome assessment: the OMERACT filter. J Rheumatol 1999;26:948-50.
21. Guzman J, Burgos-Vargas R, Duarte-Salazar C, Gomez-Mora P. Reliability of the articular examination in children with juvenile rheumatoid arthritis: interobserver agreement and sources of disagreement. J Rheumatol 1995;22:2331-6.
22. Senocak O, Unsal E, Akalin E, Ergor G. Interobserver reliability of articular examination in juvenile idiopathic arthritis. Turk J Pediatr 2003;45:29-32.
23. Ortiz Z, Shea B, Garcia Dieguez M, et al. The responsiveness of generic quality of life instruments in rheumatic diseases. A systematic review of randomized controlled trials. J Rheumatol 1999;26:210-6.
24. Molenaar ET, Boers M, van der Heijde DM, et al. Imaging in rheumatoid arthritis: results of group discussions. J Rheumatol 1999;26:749-51.
25. van der Heijde D, Calin A, Dougados M, Khan MA, van der Linden S, Bellamy N. Selection of instruments in the core set for DC-ART, SMARD, physical therapy, and clinical record keeping in ankylosing spondylitis. Progress report of the ASAS Working Group. Assessments in Ankylosing Spondylitis. J Rheumatol 1999;26:951-4.
26. Braun J, Baraliakos X, Golder W, et al. Analysing chronic spinal changes in ankylosing spondylitis: a systematic comparison of conventional x rays with magnetic resonance imaging using established and new scoring systems. Ann Rheum Dis 2004;63:1046-55.
27. Boers M, van der Heijde DM. Prevention or retardation of joint damage in rheumatoid arthritis: issues of definition, evaluation and interpretation of plain radiographs. Drugs 2002;62:1717-24.
28. Lassere MN, Bird P. Measurements of rheumatoid arthritis disease activity and damage using magnetic resonance imaging. Truth and discrimination: does MRI make the grade? J Rheumatol 2001;28:1151-7.
29. McQueen F, Lassere M, Edmonds J, et al. OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR Imaging Module. J Rheumatol 2003;30:1387-92.
30. Molenaar ET, Voskuyl AE, Dinant HJ, Bezemer PD, Boers M, Dijkmans BA. Progression of radiologic damage in patients with rheumatoid arthritis in clinical remission. Arthritis Rheum 2004;50:36-42.