ARTHRITIS & RHEUMATISM
Vol. 54, No. 8, August 2006, pp 2665–2673
DOI 10.1002/art.21972

Classification Criteria for Psoriatic Arthritis

Development of New Criteria From a Large International Study

William Taylor,[1] Dafna Gladman,[2] Philip Helliwell,[3] Antonio Marchesoni,[4] Philip Mease,[5] Herman Mielants,[6] and the CASPAR Study Group

Objective. **To compare the accuracy of existing classification criteria for the diagnosis of psoriatic arthritis (PsA) and to construct new criteria from observed data.**

Methods. **Data were collected prospectively from consecutive clinic attendees with PsA and other inflammatory arthropathies. Subjects were classified by each of 7 criteria. Sensitivity and specificity were compared using conditional logistic regression analysis. Latent class analysis was used to calculate criteria accuracy in order to confirm the validity of clinical diagnosis as the gold standard definition of "case"-ness. Classification and Regression Trees methodology and logistic regression were used to identify items for new criteria, which were then constructed using a receiver operating characteristic curve.**

Results. **Data were collected on 588 cases and 536 controls with rheumatoid arthritis (n = 384), ankylosing spondylitis (n = 72), undifferentiated arthritis (n = 38), connective tissue disorders (n = 14), and other diseases (n = 28). The specificity of each set of criteria was high. The sensitivity of the Vasey and Espinoza method (0.97) was similar to that of the method of McGonagle et al (0.98) and greater than that of the methods of Bennett (0.44), Moll and Wright (0.91), the European Spondylarthropathy Study Group (0.74), and Gladman et al (0.91). The CASPAR (ClASsification criteria for Psoriatic ARthritis) criteria consisted of established inflammatory articular disease with at least 3 points from the following features: current psoriasis (assigned a score of 2; all other features were assigned a score of 1), a history of psoriasis (unless current psoriasis was present), a family history of psoriasis (unless current psoriasis was present or there was a history of psoriasis), dactylitis, juxtaarticular new bone formation, rheumatoid factor negativity, and nail dystrophy. These criteria were more specific (0.987 versus 0.960) but less sensitive (0.914 versus 0.972) than those of Vasey and Espinoza.**

Conclusion. **The CASPAR criteria are simple and highly specific but less sensitive than the Vasey and Espinoza criteria.**

Despite the recognition that psoriatic arthritis (PsA) is a distinct disease, clinical and basic research into this disorder is often confounded by the absence of a widely agreed-upon or validated case definition (1). Several classification criteria have been proposed and used in the literature, but it is unclear which of these best represent "true" PsA (2,3). These include criteria proposed by Moll and Wright (4), Bennett (5), Gladman et al (6), Vasey and Espinoza (7), the European Spondylarthropathy Study Group (ESSG) (8), McGonagle et al (9), and Fournie et al (10). The operational definitions

Supported by the European League Against Rheumatism, Barnsley District NHS Trust (UK), Groote Schuur Hospital (Cape Town, South Africa), Department of Medical Sciences, University Hospital (Uppsala, Sweden), the Krembil Foundation (Toronto, Ontario, Canada), St. Vincent's University Hospital Radiology Department (Dublin, Ireland), Inkosi Albert Luthuli Central Hospital (Durban, South Africa), El Ayachi Hospital (Rabat-Salé, Morocco), the National Psoriasis Foundation (US), The Foundation for Scientific Research of the Belgian Society of Rheumatology, and Arthritis New Zealand.

[1]William Taylor, MBChB: University of Otago, Wellington, New Zealand; [2]Dafna Gladman, MD: University of Toronto, Toronto, Ontario, Canada; [3]Philip Helliwell, MD: University of Leeds, Leeds, UK; [4]Antonio Marchesoni, MD: University of Milan, Milan, Italy; [5]Philip Mease, MD: Seattle Rheumatology Associates, Seattle, Washington; [6]Herman Mielants, MD: University of Ghent, Ghent, Belgium.

Dr. Gladman has received consulting fees or honoraria (less than $10,000 each) from Abbot, Amgen, Centocor, Schering, and Wyeth.

Address correspondence and reprint requests to William Taylor, MBChB, Department of Medicine, Wellington School of Medicine and Health Sciences, University of Otago, PO Box 7343, Wellington 6039, New Zealand. E-mail: will.taylor@otago.ac.nz.

Submitted for publication August 17, 2005; accepted in revised form April 13, 2006.

of these criteria are available from the authors. Also, rheumatologists in practice appear to exhibit substantial variation in how they diagnose PsA (11). Variability in case definition is problematic. Heterogeneous study populations confound clinical research so that prognostic or intervention trials are difficult to interpret when patients with different diseases are included.

We have recently compared the test performance characteristics of these criteria using data retrospectively obtained from case notes and radiographs of patients with PsA or rheumatoid arthritis (RA) from 2 European centers (12). That study showed that the criteria of Vasey and Espinoza, Gladman et al, and McGonagle et al had similar test performance characteristics. In order to resolve some of the problems associated with retrospective data collection from only 2 centers and with a control group of RA patients alone, we designed a larger, prospective, international study with 2 objectives: 1) to compare the performance of existing criteria more rigorously, and 2) to see whether more accurate criteria could be derived from examination of data directly. It is important to stress that this study concerned classification criteria, for use with groups of patients in the context of clinical research, rather than diagnostic criteria, which might be useful for patient diagnosis in individual clinical encounters. Members of our group, the CASPAR (ClASsification criteria for Psoriatic ARthritis) Study Group, are listed in Appendix A.

PATIENTS AND METHODS

Patient and control selection. For selection of study patients and controls, the recommendations by Felson and Anderson were carefully considered and followed as far as possible (13). Consecutive clinic attendees with physician-diagnosed PsA were enrolled in the study from 30 rheumatology clinics in 13 countries. The diagnosis was based upon opinions by rheumatologists with longstanding expertise in PsA. Controls were also consecutive clinic attendees with other forms of inflammatory arthritis matched for approximate disease duration ($<$12 months or $\geq$12 months). The only other criterion was that at least 50% of controls were to have RA to reflect the disease distribution seen in normal rheumatologic practice. The choice of appropriate controls was considered carefully. Since the objective of the study was to identify criteria for use in clinical research settings, it was believed that patients with other inflammatory arthropathies would be appropriate controls. This was because recruitment for clinical research generally comes from rheumatology clinics where different types of inflammatory arthritis are highly prevalent.

Data collection. Data were recorded on standardized forms and included demographics (ethnicity, sex, age) and information, psoriasis (whether currently evident or previously observed, nail involvement), time of onset of skin disease and arthritis, dactylitis, chest wall pain, diffuse enthesis pain, inflammatory heel pain, clinical sacroiliitis, family history of psoriasis, inflammatory spinal pain (lumbar, thoracic, and cervical), rheumatoid factor (RF) status, elevated acute-phase reactant levels, rheumatoid nodules, primary clinical diagnosis, other rheumatic diseases, number of joint surgeries, Psoriasis Area Severity Index (14), Health Assessment Questionnaire score (15), spinal mobility measures, joint counts (tender, swollen, damaged), and treatment. The precise definitions of all clinical, radiographic, and laboratory items are available from the authors. Altogether, information on more than 50 variables was collected. Clinical and basic laboratory data were recorded at each study site.

Laboratory studies. Sera and whole blood DNA extracts were submitted to the central coordinating center. Antibodies to cyclic citrullinated peptide (anti-CCP) were measured by enzyme-linked immunosorbent assay (ELISA) using a second-generation commercially available kit (QUANTA Lite IgG ELISA; A. Menarini Diagnostics, Wokingham, UK). A cutoff of 20 units/ml was used to define positivity. Radiographs of the hands, feet, pelvis, lumbar spine, and cervical spine were obtained using local facilities, unless films from the previous 12 months were available. Hard-copy films or digitized images were submitted to the central coordinating center and read in tandem by 2 observers who were blinded to the physician diagnosis. Radiographic features were identified using the definitions developed by an earlier observer reliability study (16).

Statistical analysis. The overall discriminant validity of each individual feature was assessed using the specificity and sensitivity of each item for the diagnosis of PsA, and the independent contribution of each feature was assessed using multivariate logistic regression models. Cases with missing data items were included in the regression analysis by categorizing each variable as "present," "absent," or "missing." The odds ratios (ORs) presented are with reference to the "absent" state.

We chose 2 methods of validating the diagnostic "gold standard" against which the rules were to be tested. Expert physician diagnosis is the traditional gold standard when there is no pathognomonic feature. However, it is known that criteria for diagnosing PsA vary among physicians, particularly for nonclassic forms of disease (17).

The first method involved randomly selecting 10% of case record forms for scrutiny by a Data Quality Committee. Any disagreements in diagnosis based on data collected for the study were referred back to the local investigator for clarification. The second method employed the technique of latent class analysis (LCA) (18). This statistical approach assumes that each classification method classifies patients into the same classes, although with different errors. It is also assumed that agreement between the classes identified by each set of criteria is mediated only by the underlying true status of the patient. This method enables one to specify a model so that the error rates (false-positive and false-negative results) or conditional response probabilities (sensitivity and specificity) of each set of criteria within the model can be derived without actually knowing the true classification state of the patient. A procedure suggested by Uebersax (19) was used to test for conditional independence, which is an assumption of LCA. To assess the concordance between the latent classes and the clinical

Table 1. Patient characteristics*

	PsA patients (n = 588)	Controls (n = 536)
Age, mean ± SEM years	50.3 ± 0.54	55.2 ± 0.62
Disease duration, mean ± SEM years	12.5 ± 0.40	13.3 ± 0.46
Disease duration <12 months	31 (5.3)	20 (3.7)
Male sex	306 (52)	198 (37)
Caucasian ethnicity	511 (87)	445 (83)
RF positive	27 (4.6)	307 (57)
Anti-CCP positive†	27 (7.6)	188 (55)

* Except where indicated otherwise, values are the number (%) of patients. PsA = psoriatic arthritis; RF = rheumatoid factor; anti-CCP = anti–cyclic citrullinated peptide.
† Measured in 353 cases and 343 controls.

diagnosis, each subject was reclassified according to the latent class, and agreement between the clinical diagnosis and the latent class model was assessed with a kappa statistic. The purpose of the LCA was to use an independent approach to validate the clinical diagnosis. Latent Gold, version 4 (Statistical Innovations, Belmont, MA) was used for the LCA.

Among the subjects who could be classified by every rule, the specificity and sensitivity of each classification rule were compared using forward conditional logistic regression analysis. The classification method was regressed against the result of the method (positive or negative) separately for cases of PsA (comparing sensitivity) and controls (comparing specificity). In this analysis, the definition of case or control status was by clinical diagnosis. SPSS software, version 12 (SPSS, Chicago, IL) was used for this analysis.

Data-derived classification rules were derived using Classification and Regression Trees (CART) analysis (20). This method of recursive partitioning iteratively selects variables that split the sample into progressively purer groups. It has a theoretical advantage over techniques such as logistic regression, in that the structure of the classes in relation to the predictor variables is not assumed—that is, different combinations of the predictor variables may identify subgroups

Table 2. Discriminant value of individual clinical and laboratory features*

Feature (no. of patients)	Univariate analysis		Multivariate analysis
	Sensitivity, %	Specificity, %	OR (95% CI)†
Normal acute-phase reactant (279)	32	80	
RF negative (727)‡	95	60	0.057 (0.019–0.17)
Anti-CCP negative (483)	92	55	
Any tender enthesis (436)	53	77	
Chest wall pain (239)	25	83	
Diffuse enthesis pain (122)	14	93	
Inflammatory heel pain (335)	39	81	
Inflammatory LBP (408)	46	74	
Lumbar stiffness (481)	53	51	
Inflammatory neck pain (407)	39	66	
Neck stiffness (695)	69	31	
Inflammatory thoracic spinal pain (180)	20	88	
Thoracic stiffness (176)	17	84	
Clinical sacroiliitis (263)	31	84	
Absence of subcutaneous nodules (1,039)	100	16	
Iritis (62)	4.6	93	
Dactylitis (338)	54	95	20 (5.9–67) for current dactylitis; 5.5 (1.8–17) for history of dactylitis
Nail dystrophy (352)	58	98	
History of psoriasis (538)	94	98	97 (36–261)
Family history of psoriasis (318)	47	91	5.7 (2.4–13)
Current psoriasis on examination (532)	88	98	22 (8.4–56)
<4 MCP joints involved (628)	66	55	
Any DIP joint involved (202)	29	94	
Any toe DIP joint involved (168)	23	94	
Symmetry of joint involvement (924)	80	15	
All joints of any single ray involved (90)	11	96	

* 95% CI = 95% confidence interval; LBP = low back pain; MCP = metacarpophalangeal; DIP = distal interphalangeal (see Table 1 for other definitions).
† Forward conditional logistic regression of every item regressed against clinical diagnosis. The odds ratio (OR) refers to the ratio of the odds of having PsA with an item being present to the odds of having PsA with the item being absent. Only those items selected by the model are presented (model $\chi^2 = 1{,}329$, 11 degrees of freedom, $P < 0.001$, $R^2 = 0.69$).
‡ The variable used in the regression model was RF positivity.

Table 3. Discriminant value of individual radiographic features*

Feature (no. of patients)	Univariate analysis		Multivariate analysis OR (95% CI)†
	Sensitivity, %	Specificity, %	
Andersson lesion (1)	0	100	
Interphalangeal bony ankylosis (74)	12	97	
Bilateral sacroiliitis (111)	11	87	
DIP erosive disease (170)	62	89	
Entheseal erosion (76)	6.7	90	
Entheseal bony proliferation (143)	16	85	
Juxtaarticular new bone formation (116)	19	95	4.75 (2.88–7.84)
Marginal syndesmophytes (56)	4.5	91	
Nonmarginal syndesmophytes (87)	10	90	
Paravertebral ossification (5)	0.64	99	
Joint osteolysis (102)	13	92	
Ray involvement (37)	6.1	99	
Romanus lesion (9)‡	0.65	99	
Tuft osteolysis (22)	4.3	100	
Unilateral sacroiliitis (34)	5.4	98	
Any peripheral radiographic feature (230)	31	83	
Any axial radiographic feature (196)	24	79	
Any proliferative radiographic feature (319)	43	66	

* See Table 2 for other definitions.
† Forward conditional logistic regression of every item regressed against clinical diagnosis. The OR refers to the ratio of the odds of having psoriatic arthritis (PsA) with an item being present to the odds of having PsA with the item being absent. Only those items selected by the model are presented (model $\chi^2 = 55.4$, 2 degrees of freedom, $P < 0.001$, $R^2 = 0.048$).
‡ Clearly defined erosion of the anterior margin of the discovertebral junction at the superior or inferior portions of the vertebral body.

within each class. For instance, axial features are likely to be more important for PsA with predominant axial disease and peripheral features are likely to be less important. CART is potentially able to uncover this structure. Second, when a splitting variable is missing, CART is able to specify surrogate predictors at each node, which minimizes the impact of missing data. Testing of the rule uses the technique of v-fold cross-validation, whereby v random samples of the data are used to grow trees that are compared with a tree grown from the entire data set. Typically, a 10-fold cross-validation procedure is used. CART for Windows, version 5.0 (Salford Systems, San Diego, CA) was used for this analysis (21).

The most discriminating items identified by logistic regression and/or CART were combined, and a receiver operating characteristic (ROC) curve was constructed to identify the number of items needed to be present to achieve optimal sensitivity and specificity. SPSS software, version 12 was also used for this analysis. Each center obtained ethics approval from its local Ethics Committee or Institutional Review Board in compliance with the Declaration of Helsinki.

RESULTS

Data were collected prospectively from 588 consecutive clinic attendees with PsA and from 536 control patients (next clinic attendee with inflammatory arthritis). Control patients had RA (n = 384), ankylosing spondylitis (n = 72), undifferentiated arthritis (n = 38), connective tissue disorders (n = 14), and other diseases (n = 28). The disease durations were similar in each disease group, but patients with PsA were younger and more likely to be male (Table 1).

The 38 patients with undifferentiated arthritis had a mean ± SD age of 42 ± 13.6 years, a mean ± SD disease duration of 9 ± 8.3 years, and a mean ± SD swollen joint count of 4.2 ± 6.2. Two of these 38 patients (5.3%) had psoriasis, 2 of 25 (8%) were anti-CCP positive, 2 of 37 (5.4%) were RF positive, 5 of 32 (15.6%) had radiographic sacroiliitis, none had peripheral radiographic features of PsA, 4 of 38 (10.5%) had dactylitis or a history of dactylitis, 4 of 38 (10.5%) had a history of iritis, and 23 of 38 (60.5%) had tenderness at an enthesis.

There were 20 patients with PsA who did not have current psoriasis or a history of psoriasis. In this subgroup, the average duration of disease was 14 years, 10 patients had a family history of psoriasis, dactylitis was observed in 8, nail dystrophy was observed in 4, none were anti-CCP positive, and 19 were negative for RF.

The discriminant value of laboratory and clinical features (Table 2) and radiographic features (Table 3) was expressed in terms of sensitivity and specificity. Sensitivity refers to the proportion of patients with PsA

Table 4. Test performance of each classification rule*

Rule (ref.)	No. of controls	No. of PsA patients	Sensitivity, %	Specificity, %	Proportion not classifiable, %
Gladman et al (6)					
Positive	17	533			
Negative	514	50	91	97	0.89
Not classifiable	5	5			
McGonagle et al (9)					
Positive	46	536			
Negative	474	10	98	91	5.2
Not classifiable	16	42			
Fournie et al (10)					
Positive	15	420			
Negative	303	25	94	95	32
Not classifiable	218	143			
ESSG (8)					
Positive	48	432			
Negative	483	149	74	91	1.1
Not classifiable	5	7			
Moll and Wright (4)					
Positive	8	497			
Negative	519	50	91	98	4.4
Not classifiable	9	41			
Bennett (5 of 8 version) (5)†					
Positive	1	231			
Negative	533	297	44	100	5.5
Not classifiable	2	60			
Vasey and Espinoza (7)					
Positive	21	552			
Negative	506	16	97	96	2.6
Not classifiable	9	20			

* PsA = psoriatic arthritis; ESSG = European Spondylarthropathy Study Group.
† The 5 of 8 version refers to a modification of original criteria described in a previous report (12).

who had the feature, and specificity refers to the proportion of controls who did not have the feature. Only 1 radiographic feature was found to be significantly and independently associated with PsA in the multivariate analysis, probably because most radiographic features occurred too infrequently to be discriminating. A third model was derived by combining the significant variables from both models (clinical/laboratory features and radiographic features were initially modeled separately). All of the following variables remained independently associated with PsA: current psoriasis (OR 22.5, 95% confidence interval [95% CI] 8.5–59.1), family history of psoriasis (OR 5.6, 95% CI 2.3–13.3), history of psoriasis (OR 102.6, 95% CI 37.8–278.5), dactylitis currently present (OR 17.9, 95% CI 5.2–62.1) or a history of dactylitis (OR 6.0, 95% CI 1.9–19.0), negative RF test result (OR 27.8, 95% CI 6.5–125), and juxtaarticular new bone formation seen on radiographs (OR 4.6, 95% CI 1.3–16.9).

The performance of each set of criteria is shown in Table 4. Since some patients could not be classified by every method, numbers of patients classified as having or not having PsA as well as numbers of patients who could not be classified are shown. The method of Fournie et al (10) was the most difficult to use, primarily because HLA data were not available at the time of this analysis (only 68% of subjects could be classified by the method of Fournie et al). Excluding the rule of Fournie et al, 949 patients (84%) could be classified by all of the other 6 methods.

The LCA showed similar test performance characteristics for each set of criteria. Further, the latent class model agreed closely with the clinical diagnosis (Cohen's $\kappa = 0.96$, 95% CI 0.95–0.98), confirming the validity of the clinical diagnosis and confirming that this could be used as a satisfactory gold standard.

Examination of 124 randomly selected case record forms by the Data Quality Committee revealed a single instance of an incorrect diagnosis (0.8%). Subsequent analyses therefore used the clinical diagnosis as the definition of "case"-ness.

The differences in specificity and sensitivity be-

Table 5. Test performance of the classification rules, using cases that could be classified by every rule except for that of Fournie et al (n = 949)*

Rule (ref.)	No. of controls	No. of PsA patients	*P,* for a difference in sensitivity compared with Vasey and Espinoza rule†	*P,* for a difference in specificity compared with Vasey and Espinoza rule‡
Gladman et al (6)				
Positive	8	397	0.001	0.49
Negative	500	44		
McGonagle et al (9)				
Positive	38	431	0.176	<0.001
Negative	470	10		
ESSG (8)				
Positive	33	326	<0.001	0.001
Negative	475	115		
Moll and Wright (4)				
Positive	4	399	0.001	0.081
Negative	504	42		
Bennett (5 of 8 version) (5)§				
Positive	0¶	224	<0.001	0.031
Negative	508	217		
Vasey and Espinoza (7)				
Positive	11	424	Reference	Reference
Negative	497	17		

* Statistical comparisons of sensitivity and specificity used logistic regression analysis with the Vasey and Espinoza criteria as the reference variable. PsA = psoriatic arthritis; ESSG = European Spondylarthropathy Study Group.
† Model log likelihood −1,267, 5 degrees of freedom, $P < 0.001$, and the effect of criteria type was significant ($P < 0.001$).
‡ Model log likelihood −431, 5 degrees of freedom, $P < 0.001$, and the effect of criteria type was significant ($P < 0.001$).
§ The 5 of 8 version refers to a modification of original criteria described in a previous report (12).
¶ 0.5 was added to zero cells to estimate model parameters.

tween the criteria were confirmed using conditional logistic regression (Table 5) applied to the data using clinical diagnosis to define case–control status (Table 5). Only those subjects who could be classified by every method (except for the method of Fournie et al) were included in this analysis. The Vasey and Espinoza method was chosen to be the reference, since it appeared to perform best in the earlier analysis. Only the *P* value for each OR is presented for reasons of clarity. The method of McGonagle et al had sensitivity similar to that of the method of Vasey and Espinoza, but the other methods, especially those of Bennett and the ESSG, were significantly less sensitive. The methods of Gladman et al and of Moll and Wright had specificity similar to that of the method of Vasey and Espinoza, and the method of Bennett had better specificity, but the other methods (of the ESSG and of McGonagle et al) had significantly less specificity.

A CART model was derived using data priors (specifying prevalence as it was actually observed). The presence of psoriasis differentiated between the 2 groups to such an extent that only a history of psoriasis and current psoriasis were selected as independent predictors when all variables were tested together. The specificity of this model was 96.8% and its sensitivity was 96.1%. Surrogate variables (as if the primary splitting variable were missing) chosen with the CART algorithm were nail disease, dactylitis, RF negativity, and a family history of psoriasis.

A further model was constructed using the variables identified in the logistic regression analysis as being independently predictive of PsA in addition to variables identified in the CART analysis: psoriatic nail dystrophy (from CART analysis), a history of psoriasis (if no current evidence of psoriasis), current evidence of psoriasis, dactylitis, a history of dactylitis (if no current dactylitis evident), juxtaarticular new bone formation (plain radiography of hands and feet), a family history of psoriasis (if no current evidence or history of psoriasis), and a negative RF test result. The number of these features being present for each subject from the entire data set were summed to form a score, from which an ROC curve was derived. We did not use weighting suggested by the ORs from the logistic regression analysis, but we did weight current psoriasis more highly than other possible features of psoriasis, since we believed that the reliability and accuracy of this feature were likely to be greater than the reliability and accuracy of features based on the patients' recall or the medical record.

Table 6. The CASPAR criteria*

To meet the CASPAR (ClASsification criteria for Psoriatic ARthritis) criteria, a patient must have inflammatory articular disease (joint, spine, or entheseal) with ≥3 points from the following 5 categories:

1. Evidence of current psoriasis, a personal history of psoriasis, or a family history of psoriasis.
   Current psoriasis is defined as psoriatic skin or scalp disease present today as judged by a rheumatologist or dermatologist.† A personal history of psoriasis is defined as a history of psoriasis that may be obtained from a patient, family physician, dermatologist, rheumatologist, or other qualified health care provider.
   A family history of psoriasis is defined as a history of psoriasis in a first- or second-degree relative according to patient report.
2. Typical psoriatic nail dystrophy including onycholysis, pitting, and hyperkeratosis observed on current physical examination.
3. A negative test result for the presence of rheumatoid factor by any method except latex but preferably by enzyme-linked immunosorbent assay or nephelometry, according to the local laboratory reference range.
4. Either current dactylitis, defined as swelling of an entire digit, or a history of dactylitis recorded by a rheumatologist.
5. Radiographic evidence of juxtaarticular new bone formation, appearing as ill-defined ossification near joint margins (but excluding osteophyte formation) on plain radiographs of the hand or foot.

* The CASPAR criteria have specificity of 98.7% and sensitivity of 91.4%.
† Current psoriasis is assigned a score of 2; all other features are assigned a score of 1.

The area under the ROC curve was 0.989 (95% CI 0.984–0.995). The CASPAR criteria (Table 6) were identified as the presence of 3 or more of these features in a person with inflammatory arthritis (sensitivity 0.914, specificity 0.987). In those subjects classified by the Vasey and Espinoza and CASPAR methods (n = 1,095), the sensitivities were 0.972 versus 0.914, respectively ($P < 0.001$), and the specificities were 0.960 versus 0.987, respectively ($P < 0.001$).

DISCUSSION

In the present study we used patient-derived data to compare existing classification criteria and to derive new criteria for PsA. The direct comparison of existing classification criteria methods (except for the method of Fournie et al) showed that the method of Vasey and Espinoza had the best combination of sensitivity and specificity. The LCA showed similar test performance characteristics for each set of criteria. Further, the latent class model agreed closely with the clinical diagnosis, confirming the validity of the clinical diagnosis as a satisfactory gold standard. A further set of criteria suggested by the observed data (the CASPAR criteria) was found to have better specificity but less sensitivity than the Vasey and Espinoza method.

There have been few reported studies that compared different criteria for the diagnosis of PsA (22). We previously found in a study of patients with PsA and RA that the currently proposed classification criteria had nearly equivalent test performance, apart from the Bennett criteria and ESSG criteria, which were significantly less sensitive. As a rough index of feasibility, the proportions of subjects who could not be classified because of missing data items were similar for each rule except for the rule of Fournie et al. Using CART methods, alternative combinations of variables could be used to diagnose PsA with some degree of accuracy, even without inclusion of RF or psoriasis (12).

Since the Vasey and Espinoza method performed best of the existing criteria, we briefly describe the requirements of this method. First, the presence of psoriasis or nail dystrophy is mandatory. Second, 1 feature of either spinal disease or peripheral disease is required. The following features are specified: pain and soft tissue swelling with or without limitation of movement of the distal interphalangeal joint for >4 weeks; pain and soft tissue swelling with or without limitation of motion of the peripheral joints involved in an asymmetric peripheral pattern for >4 weeks (this includes a sausage digit); symmetric peripheral arthritis for >4 weeks in the absence of RF or subcutaneous nodules; pencil-in-cup deformity, whittling of terminal phalanges, fluffy periostitis and bony ankylosis (radiographic changes); spinal pain and stiffness with the restriction of motion present for >4 weeks; grade 2 symmetric sacroiliitis according to the New York criteria (23); grade 3 or 4 unilateral sacroiliitis.

The present study has a number of strengths. First, the cases were unselected consecutive clinic attendees with a clinical diagnosis of PsA who were sampled from 30 clinics in 13 countries. This would tend to reduce any diagnostic bias that might be present in 1 or 2 centers. The controls were also unselected consecutive clinic patients who were roughly matched for disease duration and who had any other type of inflammatory musculoskeletal disease. Approximately 2% of controls had psoriasis, which is within the range of expected population prevalence and suggests that there was little selection bias toward not recruiting controls who had psoriasis.

Second, we investigated for the possibility of an inaccurate gold standard (clinical diagnosis) by comparing test characteristic performances of the different classification criteria sets obtained with a clinical gold

standard versus those obtained with a statistically defined gold standard using LCA. There was broad agreement using these 2 approaches, suggesting that the clinical gold standard was adequate. A direct comparison of clinically defined diagnosis with the cluster-defined diagnosis using LCA revealed near-perfect agreement. We also examined clinical diagnoses using an independent data monitoring committee and found that the case–control status could be incorrect in only 0.8% of the subjects.

Third, we examined a large range of potential items that were not suggested by contributors to the data collection, circumventing issues of circularity. Fourth, we obtained data from a large number of cases and controls that allowed sufficient power to detect small differences in test performance characteristics.

We used 2 different statistical approaches to identify features for new criteria and found similar important variables from both CART and logistic regression. We combined the important variables in a simple way that did not rely upon the weighting suggested by the logistic regression analysis, since we believed that the weighting would probably be sample specific. Such simple criteria (the CASPAR criteria) (Table 6) were found to be highly specific but less sensitive than the Vasey and Espinoza criteria. One attraction of these criteria is that they permit the diagnosis of PsA despite RF positivity or the absence of psoriasis, as long as other typical features of PsA are present.

The main limitation of the present study is the well-established nature of the disease (disease duration ~12 years) in the study subjects. It is not possible to apply the results of the present study to recent-onset disease. Problems with classification criteria in recent-onset rheumatic disease are not limited to PsA, but this is clearly an area that requires further research. Similarly, the present study was performed in subjects with known inflammatory articular disease. It is not possible to apply the results of this study to the general population or to other clinical populations (for example, dermatology clinic populations). However, since the diagnosis requires the presence of an inflammatory articular feature, it is likely that these criteria would perform well in the general population. Further studies are needed to test the performance of the criteria in these populations.

It is important to note that the first criterion of the proposed criteria is not precisely defined. Since a diagnosis of inflammatory arthritis was required for inclusion in this study (for both cases and controls), the data cannot be used to operationalize a precise definition of inflammatory articular disease. Furthermore, we did not wish to prejudge features of inflammatory articular disease by framing entry criteria in terms of particular manifestations of inflammatory disease; entry criteria were framed in terms of diagnosis rather than by prespecifying features necessary to make a diagnosis of inflammatory arthritis. This does mean that it will now be necessary to examine potential features of inflammatory articular disease in samples with and without such diseases to properly define this criterion.

One major purpose of classification criteria is to enroll appropriate patients into clinical intervention studies. For this purpose, criteria with high specificity are more desirable than criteria with high sensitivity. The CASPAR criteria have been derived from observed patient data, are more specific in a rheumatology clinic setting, and are easier to use than existing classification criteria. We therefore propose that the CASPAR criteria be adopted for future clinical studies of PsA.

REFERENCES

1. Taylor WJ, Helliwell PS. Case definition of psoriatic arthritis [letter]. Lancet 2000;356:2095.
2. Helliwell PS, Taylor WJ. Classification and diagnostic criteria for psoriatic arthritis. Ann Rheum Dis 2005;64 Suppl 2:ii3–8.
3. Gladman DD. Psoriatic arthritis. Baillieres Clin Rheumatol 1995; 9:319–29.
4. Moll JM, Wright V. Psoriatic arthritis. Semin Arthritis Rheum 1973;3:55–78.
5. Bennett RM. Psoriatic arthritis. In: McCarty DJ, editor. Arthritis and allied conditions. 9th ed. Philadelphia: Lea & Febiger; 1979. p. 645.
6. Gladman DD, Shuckett R, Russell ML, Thorne JC, Schachter RK. Psoriatic arthritis: an analysis of 220 patients. Q J Med 1987;238: 127–41.
7. Vasey F, Espinoza LR. Psoriatic arthropathy. In: Calin A, editor. Spondyloarthropathies. Orlando (FL): Grune & Stratton; 1984. p. 151–85.
8. Dougados M, van der Linden S, Juhlin R, Huitfeldt B, Amor B, Calin A, et al, and the European Spondylarthropathy Study Group. The European Spondylarthropathy Study Group preliminary criteria for the classification of spondylarthropathy. Arthritis Rheum 1991;34:1218–27.
9. McGonagle D, Conaghan PG, Emery P. Psoriatic arthritis: a unified concept twenty years on. Arthritis Rheum 1999;42:1080–6.
10. Fournie B, Crognier L, Arnaud C, Zabraniecki L, Lascaux-Lefebvre V, Marc V, et al. Proposed classification criteria of psoriatic arthritis: a preliminary study in 260 patients. Rev Rhum Engl Ed 1999;66:446–56.
11. Gorter S. Psoriatic arthritis: performance of rheumatologists in daily practice. Ann Rheum Dis 2002;61:219–24.
12. Taylor WJ, Marchesoni A, Arreghini M, Sokoll K, Helliwell PS. A comparison of the performance characteristics of classification criteria for the diagnosis of psoriatic arthritis. Semin Arthritis Rheum 2004;34:575–84.
13. Felson DT, Anderson JJ. Methodological and statistical approaches to criteria development in rheumatic diseases. Baillieres Clin Rheumatol 1995;9:253–66.

14. Fredriksson T, Pettersson U. Severe psoriasis—oral therapy with a new retinoid. Dermatologica 1978;157:238–44.
15. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
16. Taylor WJ, Porter GG, Helliwell PS. Operational definitions and observer reliability of the plain radiographic features of psoriatic arthritis. J Rheumatol 2003;30:2645–58.
17. Symmons D, Lunt M, Watkins G, Helliwell P, Jones S, McHugh N, et al. Developing classification criteria for peripheral joint psoriatic arthritis. Step I. Establishing whether the rheumatologist's opinion on the diagnosis can be used as the "gold standard". J Rheumatol 2006;33:552–7.
18. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a "gold standard." Am J Psychiatry 1994;151:650–7.
19. Uebersax J. A practical guide to local dependence in latent class models. 2000. URL: http://ourworld.compuserve.com/homepages/jsuebersax/condep.htm.
20. Breiman L, Friedman JH, Olshen RA, Stone JS. Classification and regression trees. Belmont (CA): Wadsworth; 1984.
21. Steinberg D, Colla P. CART—Classification and Regression Trees. San Diego (CA): Salford Systems; 1997.
22. Marques CD, Navaux F, Constantin A, Mazieres B, Cantagrel A. Comparative study of 4 diagnosis criteria sets for psoriatic arthritis in early inflammatory rheumatisms [abstract]. Ann Rheum Dis 2000;59 Suppl 1:202.
23. Gofton JP. Report from the Subcommittee on Diagnostic Criteria for Ankylosing Spondylitis. In: Bennett PH, Wood HN, editors. Population studies of the rheumatic diseases; 1966. New York: Excerpta Medica; 1966. p. 314–6, 456–7.

APPENDIX A: THE CASPAR STUDY GROUP

Members of the Writing and Steering Committee of the CASPAR Study Group are as follows: Dafna Gladman, MD (University of Toronto, Toronto, Ontario, Canada); Philip Helliwell, MD, Principal Investigator (University of Leeds, Leeds, UK); Antonio Marchesoni, MD (University of Milan, Milan, Italy); Philip Mease, MD (Seattle Rheumatology Associates, Seattle, Washington); Herman Mielants, MD (University of Ghent, Ghent, Belgium); William Taylor, MBChB (University of Otago, Wellington, New Zealand).

Study Coordinators are as follows: Philip Helliwell, MD, Robin Waxman, MPH (University of Leeds, Leeds, UK). Data Quality Committee members are as follows: Mike Green, MD, Philip Helliwell, MD, Dennis McGonagle, MD (Leeds, UK); Deborah Symmons, MD (Manchester, UK).

Investigators are as follows: Marissa Lassere, MBBS, PhD, Joy Rappo (Sydney, New South Wales, Australia); Herman Mielants, MD, Marthe Van de Berghe, MD, Hans Georg Zmierczak, MD (Ghent, Belgium); Kurt de Vlam, MD (Leuven, Belgium); Anthony Russell, MD (Edmonton, Alberta, Canada); Dafna Gladman, MD, Cathy Schertag, MD (Toronto, Ontario, Canada); Emmanuelle Dernis, MD, Maxime Dougados, MD, Laure Gossec, MD, Djamila Zerkak, MD (Paris, France); Bernard Fournie, MD (Toulouse, France); Oliver FitzGerald, MD, Marie O'Rourke, MD, Doug Veale, MD (Dublin, Ireland); Antonio Marchesoni, MD (Milan, Italy); Ennio Lubrano, MD (Naples, Italy); Ignatio Olivieri, MD (Potenza, Italy); Pierluigi Macchioni, MD, Carlo Salvarani, MD (Reggio Emilia, Italy); Noufissa Lazrak Bentalha, MD, Najia Hajjaj-Hassouni, MD (Sale, Morocco); Paul Healy, MBChB, William Taylor, MBChB (Wellington, New Zealand); Asgar Ali Kalla, MD, Jenny Potts, MD (Cape Town, South Africa); Girish Modi, MD, Neeta Patel, MD (Durban, South Africa); Juan Carlos Torre Alonso, MD (Oviedo, Spain); Gunilla Holmstrom, MD, Ulla Lindqvist, MD, Björn Svensson, MD, Elke Theander, MD (Malmo and Uppsala, Sweden); Gerd Marie Alenius, MD, Krister Ek, MD, Solbritt Rantapaa-Dahlqvist, MD (Umea, Sweden); Ade Adebajo, MD, Hisham Sharlala, MD (Barnsley, UK); Jenny Lewis, MD, Neil McHugh, MD, Pat Owen (Bath, UK); Philip Helliwell, MD, Gamal Ibrahim, MD (Bradford, UK); Julie Holdsworth, MD, Dennis McGonagle, MD (Halifax, UK); Nichol Barkham, MD, Victoria Bejarano, MD, Paul Emery, MD, Julie Henry, MD, Karen Henshaw, MD (Leeds, UK); Lesley Kay, MD (Newcastle-upon-Tyne, UK); Amanda Isdale, MD (Northallerton, UK); Liliana Candia, MD, Luis Rolan Espinoza, MD, Javier Marquez, MD (New Orleans, Louisiana); Robert Durham, MD, Christopher Ritchlin, MD (Rochester, New York); Lisa Gunter, MD, Philip Mease, MD, Lynn Wang (Seattle, Washington).

Contributors are as follows: Dr. William Taylor conceived the study. Drs. Bernard Fournie, Philip Helliwell, Antonio Marchesoni, Herman Mielants, and William Taylor designed the study. Analysis of data was performed by Drs. Philip Helliwell and William Taylor with technical assistance from John Horwood, MSc (Christchurch School of Medicine, Christchurch, New Zealand). Drs. Philip Helliwell and Guy Porter (Airedale General Hospital, Keighley, UK) performed the radiographic assessment. The anti-CCP tests were performed by Dr. Neil McHugh and Pat Owen (Royal National Hospital for Rheumatic Diseases, Bath, UK).