Development of an Assessment Tool for Dactylitis in Patients with Psoriatic Arthritis

PHILIP S. HELLIWELL, JILL FIRTH, GAMAL H. IBRAHIM, RICHARD D. MELSOM, IRFAN SHAH, and DEBORAH E. TURNER

ABSTRACT. ***Objective.*** Psoriatic arthritis (PsA) is characterized by inflammatory arthritis in the presence of psoriasis. Certain clinical features help characterize this disorder, one of which is dactylitis. Hitherto an instrument for quantifying dactylitis has not been developed.

Methods. A dactylitis score sheet was developed. The score is a function of finger circumference and tenderness, assessed and summed across all dactylitic digits. Initial results were obtained on a small sample of patients attending clinics. Inter and intraobserver agreement on the presence of dactylitis using kappa agreement statistics, and the validity and reliability of the instrument, using intraclass correlation coefficients (ICC), were assessed in a further group of 7 patients with PsA.

Results. Tender dactylitis was deemed present in 74 digits out of a total of 280 (140 digits on each occasion). Kappa agreement scores for the presence of tender dactylitis were poor to good, both within and between observers (0.25 to 0.89 between observers and 0.29 to 0.91 within observers). Agreement scores for non-tender dactylitis were poor (0.01 to 0.66 between observers and 0.01 to 0.59 within-observer agreement). The new dactylitis instrument was simple and easy to administer and was found to measure appropriate scores in patients with different severity of dactylitis. Inter and intraobserver agreement was good (interobserver ICC 0.90, 95% CI 0.74–0.98; intraobserver ICC 0.84, 95% CI 0.71–0.92). Intraobserver ICC improved but interobserver ICC deteriorated by rating simply presence or absence, rather than a 4 point grade, of tenderness.

Conclusion. A new method for quantifying dactylitis based on digital circumference and tenderness has been described. This instrument has shown good inter and intraobserver reliability. Further studies of responsiveness are now required. (J Rheumatol 2005;32:1745–50)

Key Indexing Terms:
PSORIATIC ARTHRITIS DACTYLITIS RELIABILITY

Psoriatic arthritis (PsA) is characterized by chronic arthritis in the presence of psoriasis. Certain clinical features help differentiate this disorder from rheumatoid arthritis (RA) including presence of axial involvement, distal interphalangeal involvement, dactylitis, and enthesitis. The assessment of disease activity and outcome in RA and ankylosing spondylitis (AS) is well established[1,2], but there are currently no validated measures to assess the specific clinical features of PsA[3]. The tools currently used to assess PsA are instruments that were developed and validated for use in AS and RA populations. In order to develop a clinically meaningful disease activity score relevant to PsA, specific clinical features characteristic of this disorder should be included in the assessment.

Dactylitis, one of the hallmark clinical features of PsA, occurs in 16–48% of reported cases[4-6]. Rothschild, *et al* defined dactylitis as "uniform swelling such that the soft tissues between the metacarpophalangeal and proximal interphalangeal, proximal and distal interphalangeal, and/or distal interphalangeal joint and digital tuft were diffusely swollen to the extent that the actual joint swelling could no longer be independently recognized."[7] According to some authors dactylitis is predominantly due to swelling and inflammation in the flexor tendon sheaths[8], although other groups have recorded joint synovitis as well as tenosynovitis[9]. It is possible that enthesitis may also contribute to the clinical picture[10]. However, chronic, non-tender diffuse dactylitic swelling occurring in PsA may be less of an indicator of active disease than tenderness within the swollen digit[11]. An instrument suitable for quantifying psoriatic dactylitis must therefore be able to measure the amount of swelling and distinguish tender from non-tender digits. This article describes the development and validation of such an instrument for assessing dactylitis.

From St. Lukes Hospital, Bradford NHS Foundation Trust, and the Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds, Leeds, UK.

P.S. Helliwell, DM, PhD, FRCP, Senior Lecturer in Rheumatology, St. Luke's Hospital, and Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds; J. Firth, Rheumatology Nurse Practitioner; G.H. Ibrahim, ABMS, MRCP, MSc, Staff Grade Rheumatologist; R.D. Melsom, MD, FRCP, Consultant Rheumatologist; I. Shah, MB, ChB, Senior House Officer in Rheumatology, St. Lukes Hospital; D.E. Turner, PhD, Research Fellow, Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds.

Address reprint requests to Dr. P.S. Helliwell, Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds, 36 Clarendon Road, Leeds, LS2 9NZ, UK. E-mail: p.helliwell@leeds.ac.uk

Accepted for publication April 4, 2005.

MATERIALS AND METHODS

Theoretical considerations. The definition of dactylitis indicates uniform swelling of a finger or toe, and measurement of digital volume would be the ideal variable. Further, some measure of normality would be required to compare the involved digit to a reference value. As finger and toe volume varies across the hand or foot, a reference value for each digit would be required. Further, hand and foot size varies from person to person and is generally linked to other anthropometric variables. However, it is possible to use the contralateral digit for comparison, providing that digit is not also considered to be affected by dactylitis.

Measurements of digital volume can be made directly or indirectly. Indirect methods use surface measurements to calculate the volume. One such method assumes the digit to be 3 (or in the case of the thumb/great toe, 2) truncated cones. These measurements can readily be taken from the finger and thumb, although they are harder to take reliably from the feet, and standardization of the landmarks has not been agreed on. In comparison, using the direct method involves the use of Archimedes' principle and requires the subject to place the digit into a vessel containing water, the displaced volume representing the volume of the digit. From the point of view of the clinician the above measurements are impractical for use in the clinic, and such maneuvers required for volumetric assessment are difficult to carry out on the toes.

In order to provide an instrument that is simple and easy to use, yet has a direct link to pathophysiology, it was decided to use digital circumference at a single point on the digit as a surrogate for finger volume. Finger circumference was a frequent metric used in the 1980s to monitor the efficacy of antirheumatic drugs but was discontinued after poor interobserver reliability was demonstrated[12]. In order to circumvent this problem it was decided to use the contralateral digit as a comparator expressing the result as a ratio. This method allows both a definition of dactylitis (such as a 5 or 10% difference in circumference) and a metric for use as an outcome measure.

A further modification of this technique was the subjective assessment of the degree of tenderness of the dactylitic digit. A grading system such as that used by the Ritchie index was employed[13]. The Ritchie index assigns a score from 0 to 3 in response to a squeeze from the examiner's hand: 0 = no tenderness; 1 = patient reports tenderness; 2 = patient winces; 3 = patient withdraws digit. The circumference ratio for each affected digit is then multiplied by the tenderness score to provide an overall score for each digit, the total score being the arithmetic sum of the individual scores.

The method thus assesses the main features of the condition, i.e., swelling and tenderness, and provides a score that reflects these qualities and that should improve as dactylitis improves. The only problem occurs where dactylitis is symmetrical, excluding use of the contralateral digit as a comparator. To make provision for this situation normative data were required.

Normative data were sought for men and women. Anthropometric data for the fingers are available from population studies in the 1980s so that, in the cases where affected digits are symmetric, these population norms can be used[14]. Unfortunately, normal anthropometric data for the toes are not available so a normative data set was created. Fifty-two people [25 women, 27 men, mean age 38, age range 23–60, mean body mass index (BMI) 23.9, range 19.5–36.8] agreed to have their toe circumference measured. Toe dimension was measured with a plastic device to measure circumference calibrated using wooden dowel pegs of known dimensions. Circumference of each digit was measured as close to the web space as possible. The mean of right and left digits was calculated and the circumference rounded to the nearest whole number.

Dactylitis score sheet. The proposed instrument (see Appendix) is used as follows:

1. Use separate sheet for each patient encounter
2. Record date
3. Record patient details
4. Indicate on diagram which digits are affected
5. Measure circumference of affected digits either with a tape or precalibrated loop at the level of the proximal phalanx
6. Measure circumference of contralateral digit at same level
7. If contralateral digit is involved, use appropriate value from table
8. Squeeze affected digit with moderate pressure (enough to blanch examiner's nailbed) and record response: 0 = no tenderness; 1 = tender; 2 = tender and winces; 3 = tender and withdraws
9. Perform calculation for each digit and record result
10. Add score to give grand total.

The instrument is quick and easy to complete, and initial results on response to treatment (intraarticular injection) were encouraging. Patients with 1, 3, and 4 dactylitic digits had scores of 28.6, 34.4, and 85.5, respectively, the first of these falling to 4.6 one week after an injection of 20 mg methylprednisolone into the flexor tendon sheath.

Reliability study. Informed consent was obtained and appropriate ethical committee approval obtained. Ten patients with PsA were invited to participate; however, on the day of the study only 7 patients attended. Each patient was examined twice by each of 5 examiners, with an interval of no more than 2 hours between examinations. Discussion of cases between examinations was discouraged. The order in which each patient was examined was randomized and differed between first and second examinations. Each examiner performed 2 assessments with each patient. First, the examiner decided which digits had dactylitis. This was recorded on the dactylitis instrument sheet so that a dactylitis score could be completed for each patient. Second, the examiner proceeded to measure the circumference of each of the 20 digits using a standard tool for measuring circumference (Rehaboutlet.com, Florida, USA; Figure 1). The examiners consisted of 2 consultant rheumatologists (one with a long-standing interest in PsA), an experienced rheumatology nurse practitioner/metrologist, an experienced "staff grade" doctor, and a relatively inexperienced junior doctor with 4 months' training in rheumatology.

Statistics. Since only 7 patients attended for the reliability study, instead of using a replicated latin square design, agreement was computed between individual examiners using appropriate kappa statistics and between the group of examiners using multiple observer kappa statistics[15]. In addition, intraclass correlation was calculated for the dactylitis scores between examiners and within examiners across the 2 timepoints.

RESULTS

Agreement statistics. The patients consisted of 3 men and 4 women, mean age 46 years, mean duration of disease 5.3 years. In the opinion of the principal investigator, only 4 of these people had tender dactylitis on the day of the study, although tender dactylitis was deemed present in 74 digits out of a total of 280 (140 digits on each occasion). Examiners showed poor to good interobserver agreement at both time-

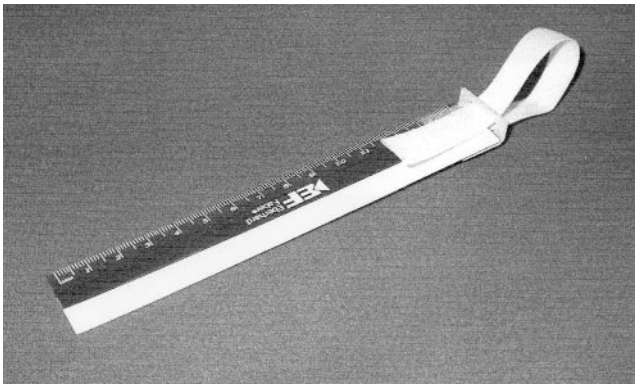

Figure 1. The circumferometer used in this study.

points for tender dactylitis, with kappa scores ranging from 0.25 to 0.89 (Table 1). Agreement statistics for non-tender dactylitis were poor, ranging from 0.01 to 0.66 (Table 2). Intraobserver kappa scores between the 2 timepoints ranged from 0.29 to 0.91 for any tender dactylitis and 0.01 to 0.59 for any non-tender dactylitis. Despite low kappa values the percentage of exact agreement between examiners for tender dactylitis was good, ranging from 71% for the great toe to 100% for the ring and little fingers. The median percentage of exact agreement for tender dactylitis was 93%.

At the first timepoint, kappa score between the group of examiners as a whole (multiobserver kappa) was good for any tender dactylitis (0.81) but poor for any non-tender dactylitis (0.14, although agreement was found 95% of the time). At the second timepoint, the scores were 0.77 and 0.02 for tender and non-tender dactylitis, respectively. Overall agreement was worse for feet, and in particular for great toes and for those patients whose BMI exceeded 30.

Dactylitis scores. Dactylitis scores showed large interpatient and moderate interobserver variability (Figure 2). However, interobserver reliability statistics using the mean score from both timepoints were good (ICC 0.90, 95% CI 0.74–0.98). Overall intraobserver scores were also good (ICC 0.84, 95% CI 0.71–0.92), but there was wide variability between examiners (Table 3).

Alternative scoring to reduce variability in dactylitis scores. To assess examiners' opinion of what constituted dactylitis, descriptive statistics of measured finger circumference ratios were generated for each dactylitic comparison and, where both digits were dactylitic, for the greater of the 2.

Table 1. Interobserver agreement for presence of tender dactylitis. Agreement statistics are based on recorded tender dactylitis for 140 digits. The first value is the interobserver kappa for the first examination, the second value the interobserver kappa for the second examination.

	Examiner				
	A	B	C	D	E
A		0.55/0.47	0.51/0.55	0.59/0.72	0.89/0.70
B			0.49/0.61	0.25/0.39	0.44/0.61
C				0.37/0.44	0.40/0.60
D					0.66/0.60

Table 2. Interobserver agreement for presence of non-tender dactylitis. Agreement statistics are based on recorded non-tender dactylitis for 140 digits. The first value is the interobserver kappa for the first examination, the second value the interobserver kappa for the second examination.

	Examiner				
	A	B	C	D	E
A		0.04/0.05	0.02/0.66	0.66/0.01	0.02/0.33
B			0.08/0.11	0.01/0.05	0.00/0.19
C				0.01/0.01	0.43/0.56
D					0.01/0.33

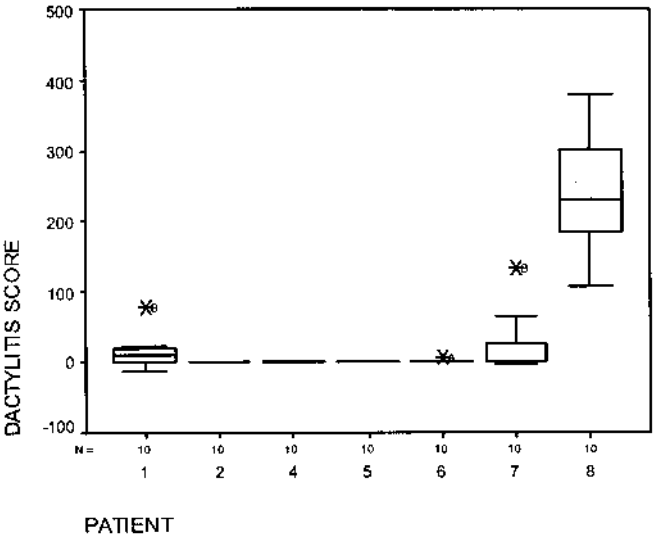

Figure 2. Box plot of scores for all examiners for all 7 patients. Boxes: interquartile range; horizontal lines: median; outer bars: range. *: Cases with values more than 3 box lengths from the upper or lower edge of the box. A: examiner A; B: examiner B.

Table 3. Intraclass correlation coefficients (ICC) for each examiner (intraobserver) between timepoints for original and revised instrument.

Examiner	Original ICC (95% CI)	Alternative ICC (95% CI)
A	0.99 (0.99–0.99)	0.99 (0.98–0.99)
B	0.71 (0.02–0.94)	0.87 (0.47–0.98)
C	0.86 (0.45–0.97)	0.84 (0.39–0.97)
D	0.74 (0.14–0.95)	0.99 (0.98–0.99)
E	0.92 (0.65–0.99)	0.97 (0.85–0.99)

CI: Confidence interval.

The median difference in digital circumference was found to be 11.7%, but with a wide range of 1.5% to 60%. Defining dactylitis using a cutoff of 10% difference therefore seemed reasonable based on existing data. Variability in scores could also result from inconsistent rating of finger tenderness. To examine the effect of this variable, scores were recalculated using a multiplier of either 0 or 1 (instead of 0–3), recording simply the presence or absence of tenderness in the digit. The alternative scores (using a 10% cutoff and a tenderness rating of 0 or 1) are shown in Figure 3. The ICC statistics were marginally worse: interobserver ICC 0.76 (95% CI 0.48–0.94), but the intraobserver statistics were improved: overall ICC 0.91 (95% CI 0.83–0.95). Individual scores are given in Table 3. However, it is evident from the figure that one examiner contributed to all of the outlying points — excluding this examiner (examiner B) from analysis improved interobserver ICC statistics: ICC 0.97 (95% CI 0.90–0.99).

DISCUSSION

Dactylitis is a hallmark and characteristic feature of PsA. It

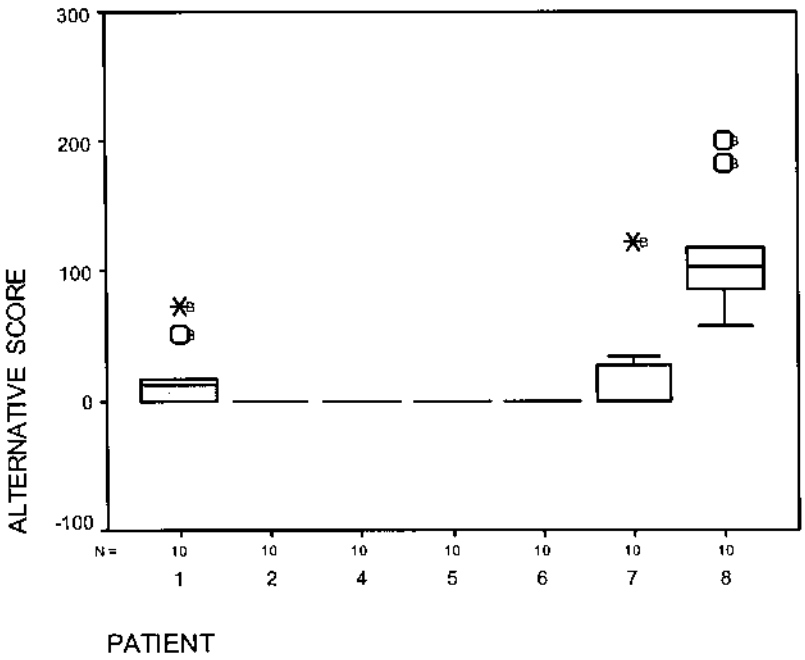

Figure 3. Box plot of scores using alternative scoring system (10% cutoff and tenderness rating of 0 or 1). Boxes: interquartile range; horizontal lines: median. □: Cases with values between 1.5 and 3 box lengths from the upper or lower edge of the box. *: Cases with values more than 3 box lengths from the upper or lower edge of the box. B: examiner B.

is characterized by uniform swelling of the digit. The underlying pathological process involves tenosynovitis and possibly enthesitis as well as synovitis. Acute and subacute forms of dactylitis are recognized; the acute form is tender and painful while the subacute form is non-tender and may have been present for many weeks. The advent of new, effective treatments for PsA has encouraged development of appropriate assessment tools for characteristic features of PsA — in this way disease activity scores can be generated and treatment effects can be quantified. This article describes the development of a new instrument for quantifying dactylitis.

For the dactylitis instrument to be applied consistently there must be agreement on which digits are deemed to be dactylitic. Agreement statistics for the presence of tender dactylitis were variable, ranging from poor to good, both between and within examiners. Although tender dactylitis was deemed to be present in only 4 of the 7 patients, the examiners were unaware of this and the inclusion of non-dactylitic patients was justified since it is as important to recognize absence as it is presence of the feature. Indeed tender dactylitis was deemed present in 74 digits out of a total of 280 (140 digits on each occasion) — roughly one in 4. The agreement statistics for the group as a whole (multi-observer kappa) were good for tender dactylitis (0.81 and 0.77 on the first and second occasions, respectively), but poor for non-tender dactylitis (0.14 and 0.02 on first and second occasions). These results compare favorably with those obtained by a Canadian group, who found kappa statistics of moderate agreement (0.57) between observers[16].

The examiners taking part in this study were, with the exception of the principal investigator, unfamiliar with this instrument and had not experienced finger circumference measurements prior to the study day. Further, no training was given. The examiners were simply told the purpose of the study and were given instructions on what measurements to take and the sequence of patients, who were labeled by study number only. The study results could have been improved if training had been administered. However, the results represent the minimum that can be achieved with this instrument without prior training. In the clinical trial situation, where the same person is likely to be making the measurements, more reliable results would be expected.

Our instrument is quick and easy to complete. When only one or 2 dactylitic digits are present, in our experience, it takes only a minute or 2 to complete the scoring sheet. Even when multiple digits are affected, the instrument is not lengthy — in this study examiners were able to measure 140 digits in about 50 minutes, or about 3 digits per minute.

There are a number of limitations with this study. First, defaulting participants precluded employment of the original study design. Because of the design of the study, where participants and examiners gather on one occasion for the purpose of establishing reliability, further recruitment of participants at some other time was not possible. Nevertheless, we believe that useful and valid information has been obtained from the study design that resulted from the nonattendance of participants.

A further weakness of the study is the use of reference

scores to enable the calculation of a ratio when matching digits are involved. The reference values for the hand are taken from a large population sample, but are necessarily simplified for the purpose of this instrument. Ideally reference values should be appropriate for age and sex — these data are available, but the reference tables would be too lengthy and impractical for use in the clinic.

Another important factor is weight, or more appropriately, BMI, but BMI data are not available within appropriate age and sex groups. The data for feet generated for this study are subject to the same caveats, only more so, as our sample size was relatively limited. Because of how the instrument is scored, this limitation is probably not important since the same reference values will be used for all subjects at each timepoint, thus reducing variability. However, the use of uncorrected (for BMI) reference values will in some cases be inappropriate and may confound the proposed definition of dactylitis (10% difference in circumference). Further use of this instrument will reveal the extent of this problem.

The variability of dactylitis scores was due in part to variability in assessment of tenderness and in part to variability in measurement of finger circumferences. In order to reduce the former source of variability the scores were recalculated using presence or absence of tenderness rather than a grading of tenderness. The result was to increase the intraobserver reliability but to reduce the interobserver reliability. However, eliminating the scores of one observer with consistently outlying scores resulted in an improvement of both inter and intraobserver reliability. In the context of the clinical trial this improvement in reliability would be offset by a reduction in sensitivity to change and consequently the effect size of the outcome measure.

We describe a new method for quantifying dactylitis based on digital circumference and tenderness. This instrument has shown good inter and intraobserver reliability. Further studies are now required, including studies of responsiveness, comparing this instrument to other methods of quantifying dactylitis such as a simple count of involved digits[17]. Further validation should also include comparison with recognized methods of imaging such as ultrasound and magnetic resonance. Such a study could be extended to examine the pathophysiological basis of non-tender dactylitis.

APPENDIX

DACTYLITIS SCORE SHEET

Addressograph Label

Date:

Please indicate dactylitic joints

Finger or toe	Circumference involved digit (A)	Circumference contralateral Digit (or Tables) (B)	Tenderness score (C)	Final score: [{(A/B) - 1} x 100] x C
TOTAL				

Standard reference. Table - hands

Digit	Men	Women
Thumb	70	58
Index	63	54
Middle	63	54
Ring	59	50
Little	52	44

Table - feet

Digit:	Men	Women
Great toe	82	72
Second	52	46
Middle	50	44
Fourth	50	44
Little	52	45

Tenderness score: response to squeeze
0 no tenderness
1 tender
2 tender and wince
3 tender and withdraw

ACKNOWLEDGMENT

We are grateful to Dr. Elizabeth Hensor for statistical advice.

REFERENCES

1. 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.
2. Pincus T, Sokka T. Quantitative measures for assessing rheumatoid arthritis in clinical trials and clinical care. Best Pract Res Clin Rheumatol 2003;17:753-81.
3. Gladman DD, Helliwell PS, Mease PJ, Nash P, Ritchlin C, Taylor W. Assessment of patients with psoriatic arthritis: a review of currently available measures. Arthritis Rheum 2004;50:24-35.
4. Helliwell PS, Marchesoni A, Peters M, Barker M, Wright V. A re-evaluation of the osteoarticular manifestations of psoriasis. Br J Rheumatol 1991;30:339-45.
5. Gladman DD, Shuckett R, Russell ML, Thorne JC, Schachter RK. Psoriatic arthritis (PSA) — an analysis of 220 patients. QJM 1987;238:127-41.
6. Fournie B, Crognier L, Arnaud C, et al. Proposed classification criteria of psoriatic arthritis. Rev Rheum Engl Ed 1999;66:446-56.
7. Rothschild BM, Pingitore C, Eaton M. Dactylitis: implications for clinical practice. Semin Arthritis Rheum 1998;28:41-7.
8. Olivieri I, Barozzi L, Favaro L, et al. Dactylitis in patients with seronegative spondylarthropathy. Assessment by ultrasonography and magnetic resonance imaging. Arthritis Rheum 1996;39:1524-8.
9. Kane D, Greaney T, Bresnihan B, Gibney R, FitzGerald O. Ultrasonography in the diagnosis and management of psoriatic dactylitis. J Rheumatol 1999;26:1746-51.
10. McGonagle D, Gibbon W, Emery P. Classification of inflammatory arthritis by enthesitis. Lancet 1998;352:1137-40.
11. Brockbank JE, Stein M, Schentag CT, Gladman DD. Dactylitis in psoriatic arthritis: a marker for disease severity. Ann Rheum Dis 2005;64:188-90.
12. Anderson JJ, Felson DT, Meenan RF, Williams HJ. Which traditional measures should be used in rheumatoid arthritis clinical trials? Arthritis Rheum 1989;32:1093-9.
13. Ritchie DM, Boyle JA, McInnes JM, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. QJM 1968;37:393-406.
14. Kember P, Ainsworth L, Brightman P. A hand anthropometric survey of British workers. Cranfield, UK: Cranfield Institute of Technology; 1981.
15. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76:378-82.
16. Gladman DD, Cook RJ, Schentag C, et al. The clinical assessment of patients with psoriatic arthritis: results of a reliability study of the Spondyloarthritis Research Consortium of Canada. J Rheumatol 2004;31:1126-31.
17. Clegg DO, Reda DJ, Mejias E, et al. Comparison of sulfasalazine and placebo in the treatment of psoriatic arthritis. A Department of Veterans Affairs Cooperative Study. Arthritis Rheum 1996;39:2013-20.