Clinical signs of early osteoarthritis: reproducibility and relation to x ray changes in 541 women in the general population

D J Hart, T D Spector, P Brown, P Wilson, D V Doyle, A J Silman

Abstract

The definition and classification of early clinically apparent osteoarthritis both in clinical situations and in epidemiological surveys remains a problem. Few data exist on the between-observer reproducibility of simple clinical methods of detecting hand and knee osteoarthritis in the population and their sensitivity and specificity as compared with radiography.

Two observers first studied the reproducibility of a number of clinical signs in 41 middle aged women. Good rates of agreement were found for most of the clinical signs tested (kappa=0·54–1·0). The more reproducible signs were then tested on a population of 541 women, aged 45–65, drawn from general practice, screening centres, and patients previously attending hospital for non-rheumatic problems. The major clinical signs used had a high specificity (87–99%) and lower sensitivity (20–49%) when compared with radiographs graded on the Kellgren and Lawrence scale (2+=positive). When analysis was restricted to symptomatic radiographic osteoarthritis, levels of sensitivity were increased and specificity was lowered.

These data show that certain physical signs of osteoarthritis are reproducible and may be used to identify clinical disease. They are not a substitute for radiographs, however, if radiographic change is regarded as the 'gold standard' of diagnosis. As the clinical signs tested seemed specific for osteoarthritis they may be of value in screening populations for clinical disease.

Department of
Rheumatology and
Environmental and
Preventive Medicine,
St. Bartholomew's
Hospital, London
D J Hart
T D Spector

Department of
Rheumatology,
St Bartholomew's
Hospital, London
P Brown
D V Doyle

Department of
Radiology,
The London Hospital,
Whitechapel
P Wilson

ARC Epidemiology
Research Unit,
Manchester University
A J Silman

Correspondence to:
Dr T D Spector,
Department of Preventive
Medicine, St Bartholomew's
Hospital Medical School,
Charterhouse Square,
London EC1M 6BQ

Accepted for publication
14 September 1990

The definition of osteoarthritis remains unclear. For epidemiological purposes the diagnosis of osteoarthritis is traditionally based on the yardstick of radiographic changes developed by Kellgren and Lawrence.[1] It is widely known that many people in the general population have radiographic evidence of osteoarthritis, but remain asymptomatic,[2] and the inclusion of asymptomatic radiographic osteoarthritis as a clinical disease entity is controversial. For large epidemiological studies mass radiography to establish prevalence of the disease may not be practical and may pose ethical problems. If osteoarthritis is to be defined clinically, however, the development of a reliable and reproducible clinical examination is essential, and the American Rheumatism Association have proposed non-radiological criteria for the knee[3] using a consensus approach. Some studies have attempted to assess physical signs in patients with established osteoarthritis[4] but as yet few data exist on the use of clinical criteria in the general population.

Many clinical features are synonymous with osteoarthritis, such as bony swelling of interphalangeal joints, but little work has been done on the reproducibility of these signs and their relation with radiographic changes.

We examined 'normal' women from the general population in an age group with the highest incidence of 'new osteoarthritis' to assess the prevalence and reproducibility of clinical signs and their relation with symptoms and x ray changes. The aim was to determine whether a diagnosis based on clinical signs alone is a useful research tool.

The study was in two parts: firstly, a pilot study to assess the interobserver reproducibility, and then a larger study to assess the prevalence of clinical signs, their relation with radiographs, and how this was altered in patients with symptoms.

Methods

PILOT STUDY

For the first part of the study common clinical signs (which would be quick to assess as part of a population screening study) were selected to test interobserver variation. Forty one unselected women, aged between 45 and 60 (mean age 53), from a local general practice were invited for clinical examination of hands and knees. The presence or absence of the following signs was recorded in the distal and proximal interphalangeal joints, first carpometacarpal joints, and both knees: bony swelling, soft tissue swelling, tenderness on palpation of the joint line, and joint pain on movement. In addition, crepitus was recorded in the knee joint. Each woman was examined by the same two observers, who were unaware of each other's results.

The following signs were negative in all subjects: soft tissue swelling and pain on movement of the distal interphalangeal and proximal interphalangeal joints, and soft tissue swelling of the carpometacarpal joint. Table 1 shows the concordance and kappa (x) statistic for agreement between the two observers for the remaining clinical signs in the 41 women. Concordance was analysed on the basis of correct classification of a joint group rather than for individual joints. Concordance was generally high for all signs, though knee crepitus (65%; x=0·14) was the lowest, and carpometacarpal joint pain on movement (100%; x=1·0) the highest.

Table 1 Comparison between two observers of clinical signs in 41 patients

	DIP*			PIP*			CMC*				Knee					
	BS*	T*	TP*	BS	T	TP	BS	T	P*	TP	BS	STS*	T	P	TP	C*
Prevalence (No (%))	9 (22)	5 (12)	5 (12)	6 (15)	2 (5)	2 (5)	2 (5)	13 (32)	10 (24)	15 (37)	3 (7)	2 (5)	9 (22)	5 (12)	12 (29)	3 (7)
Concordance (%)	90·2	100	97·5	95·1	100	100	100	95·1	100	97·5	90·2	97·5	97·5	92·7	95·1	65·0
Kappa (−1 to +1)	0·68	1·0	0·85	0·75	1·0	1·0	1·0	0·87	1·0	0·92	0·10	0·25	0·74	0·85	0·92	0·14

*DIP=distal interphalangeal; PIP=proximal interphalangeal; CMC=carpometacarpal; BS=bony swelling; T=tenderness; TP=tenderness or pain on movement; P=pain on movement; STS=soft tissue swelling; C=crepitus.

MAIN STUDY

An examination was undertaken of 541 women. aged between 45 and 65 (mean age 54), drawn from (a) the age-sex register of a large general practice in Chingford, East London; (b) volunteers from a cancer screening study, and (c) patients previously attending hospital for non-joint related problems. Women were contacted by post and a reminder letter sent six weeks later. The crude response rate for the survey was 70%. Adjustment for those no longer at the address produced a response rate of 87%. Women were seen by a single trained observer, who made a systematic clinical examination. Standard radiographs of hands and antero-posterior weightbearing views of knees were requested for all women and were available in 420 subjects (78%). After examination the women also filled out a questionnaire about joint symptoms, including details of pain, stiffness, and swelling in a joint lasting more than a month, and the onset and duration of these symptoms.

All radiographs were scored by two observers, who were unaware of each other's results, using Kellgren and Lawrence's *Atlas of Standard Radiographs*,[5] grade 2 or more being considered positive. (In the Kellgren and Lawrence scoring system 0=none; 1=doubtful; 2=minimal: definite small osteophyte(s) with minimal narrowing; 3=moderate:moderately sized osteophyte(s) or definite small osteophyte(s) with moderate narrowing; and 4=severe:large osteo-

phyte(s) or small or moderate osteophyte(s) with severe narrowing.) When there was disagreement the films were read jointly and reassessed. Radiographic joint scores were taken as the highest score for that joint group. The results are presented bilaterally for the joint groups, and classification depends on at least unilateral involvement.

ANALYSIS

The concordance of observer agreement was calculated with a simple unweighted x statistic.[6] Sensitivity, specificity, and positive predictive value were calculated to assess performance of clinical signs in detecting radiographic change.

Results

For the main study the clinical examination was extended to 541 women. Table 2 shows the prevalence of reported symptoms, clinical signs, and x ray changes in the group. In the knee joint symptoms of pain (on most days for a month) were more commonly (38%) recorded than in the other joint groups, but clinical signs of osteoarthritis were the least noted (11·5%). This site was also one of the least common sites with radiographic changes (19%). The distal interphalangeal joint was one of the most commonly symptomatic joints (25%) and was often found to have clinical (26%) and x ray changes of osteoarthritis (44%). The proportion of patients with radiographic or clinical signs who had symptoms is also shown. The percentage agreement (positive or negative) for symptoms with clinical signs was 71% for the knee, 84% for the distal interphalangeal and proximal interphalangeal joints, and 89% for the carpometacarpal joint.

Tables 3–6 show the comparison of clinical signs with x ray changes for each of the four joints sites. The most commonly seen and reproducible clinical signs of osteoarthritis (table 1) were studied. In general, signs had higher specificity than sensitivity. Tenderness and pain on movement were very specific, but not sensitive for osteoarthritis. Bony swelling was more sensitive in the digits than in the knee and carpometacarpal joint. The analysis was repeated confined to those joints which were reported to be symptomatic, and the results are also shown. The data show that by selecting only women with symptoms increased sensitivity is obtained at the expense of specificity. The analysis was repeated for Kellgren and Lawrence grade 2+ v grade 0 excluding Kellgren and Lawrence grade 1+ as these might not be regarded as true negatives. This slightly increased specificity and positive predictive

Table 2 Symptoms, clinical signs, and radiographic changes in 541 women. Results are given as number (percentage)

	DIP*	PIP*	CMC*	Knee
Reported symptoms	137 (25)	112 (21)	87 (16)	207 (38)
Clinical signs	140 (26)	90 (17)	84 (16)	62 (11)
No (%) with symptoms	95 (68)	59 (66)	56 (67)	57 (92)
Radiographic change†	185 (44)	62 (15)	133 (32)	71 (19)‡
No (%) with symptoms	76 (41)	20 (32)	49 (37)	40 (56)

*DIP=distal interphalangeal; PIP=proximal interphalangeal; CMC=carpometacarpal.
†Based on 420 women.
‡Only 366 knee radiographs available.

Table 3 Comparison of clinical signs with x ray changes (grade 2+ Kellgren and Lawrence) in the distal interphalangeal joint

	x Ray changes positive—grade 2+		
	Bony swelling	Tenderness	Pain on movement
Sensitivity (%)			
A*	49	7	1
S*	82	17	3
Specificity (%)			
A	90	97	99
S	49	83	94
Predictive value (%)			
Positive A	79	62	40
S	78	68	50
Negative A	69	57	56
S	55	32	31

*A=all patients; S=only patients with symptoms.

Table 4 Comparison of clinical signs with x ray changes (grade 2+ Kellgren and Lawrence) in the proximal interphalangeal joint

	x Ray changes positive—grade 2+		
	Bony swelling	Tenderness	Pain on movement
Sensitivity (%)			
A*	40	8	5
S*	75	20	10
Specificity (%)			
A	87	99	99
S	49	93	95
Predictive value (%)			
Positive A	35	45	50
S	32	44	40
Negative A	89	86	86
S	87	79	78

*A=all patients; S=only patients with symptoms.

Table 5 Comparison of clinical signs with x ray changes (grade 2+ Kellgren and Lawrence) in the carpometacarpal joint

	x Ray changes positive—grade 2+			
	Bony swelling	Tenderness	Pain on movement	Tenderness or pain on movement
Sensitivity (%)				
A*	19	26	22	36
S*	41	49	57	73
Specificity (%)				
A	98	92	96	91
S	100	43	61	39
Predictive value (%)				
Positive A	83	60	73	66
S	100	65	76	72
Negative A	72	88	73	76
S	44	29	40	41

*A=all patients; S=only patients with symptoms.

Table 6 Comparison of clinical signs with x ray changes (grade 2+ Kellgren and Lawrence) in the knee joint

	x Ray changes positive—grade 2+			
	Bony swelling	Tenderness	Pain on movement	Tenderness or pain on movement
Sensitivity (%)				
A*	6	17	14	20
S*	10	30	23	33
Specificity (%)				
A	98	88	95	86
S	97	74	88	69
Predictive value (%)				
Positive A	44	25	38	25
S	50	28	38	26
Negative A	81	82	82	82
S	76	76	77	76

*A=all patients; S=only patients with symptoms.

values but not to any appreciable extent. As grade 1 may not be considered completely 'normal' we also repeated the analysis shown but used grade 1+ or more as a definition of osteoarthritis v grade 0. No major differences were seen.

Discussion

We found that certain clinical signs of osteoarthritis of the hands and knees in the normal population are easy to assess and reproducible. These signs when applied to a larger population have high specificity but low sensitivity when tested against the Kellgren and Lawrence grading scale for radiographic osteoarthritis. The sensitivity was improved at the expense of specificity when only patients with symptoms were included.

Our results were based on white women aged 45–65, predominantly of social classes II-IV, and thus can be generalised to most women in the United Kingdom of the same age. The

prevalences of osteoarthritis defined radiographically are similar to those of other population surveys.[7] [8]

We found only one other study which had looked at clinical signs in the population—that by Claessens et al,[9] in which they studied the knee joint in an older population aged 45–85. Results were similar for tenderness and pain on movement, though they found bony swelling more useful, possibly because of the inclusion of a higher age group.

Non-response bias is unlikely to be a major problem as the overall response rate was high (80%). Most women were selected from general practitioner registers in a randomised fashion, and thus selection bias is not a likely factor as over 95% of possible patients register with a general practitioner in this area. A slight response bias in obtaining radiographs in the clinically negative group might have arisen if the women who were symptomatic were more likely to agree to radiography than those who were asymptomatic. As the numbers of responders were high any effect would have been slight, however.

The generally low sensitivities for the clinical signs in relation to Kellgren and Lawrence grading can be interpreted in two ways. Firstly, that clinical signs are a poor test for a positive radiograph or, alternatively, that radiographs are a poor test of early osteoarthritis. Which of these interpretations is followed depends on the acceptance of the Kellgren and Lawrence grading scale as the standard. Although traditionally used in research for over 30 years, a few studies have questioned its pre-eminence. Several recent preliminary studies have suggested that radiographic grading scales for knee osteoarthritis based predominantly on joint space are more reproducible than those based on the presence of osteophytes as in Kellgren and Lawrence.[10] [11] For osteoarthritis of the hand Kallman et al suggested that new grading methods looking at features of osteoarthritis such as osteophytes, sclerosis, and joint space separately are as reproducible as the Kellgren and Lawrence grading and may be more realistic for pathological processes.[12] Doubts about the central role of osteophytes in grading have also been raised in longitudinal studies[13] and also in the poor correlation between x ray changes and anatomical findings in skeletons.[14]

Whether osteoarthritis should be defined radiographically or clinically is a controversial issue. A subcommittee of the American Rheumatism Association (ARA) proposed preliminary criteria for osteoarthritis of the knee based on a history of clinical examination.[15] Although this has been tested in osteoarthritic patients, with good results,[3] the usefulness of the criteria in the general population remains to be validated. The ARA classification has been criticised by some authors for its oversimplicity.[16] Nevertheless, the development of clinical criteria for osteoarthritis appears to be essential for future work in the field and the criteria would seem to be a step in the right direction.[17] Our data have shown that the clinical examination can be reliably used in population surveys as a means of defining osteoarthritis.

If, however, one accepts the Kellgren and Lawrence grading scale as the standard for diagnosis of osteoarthritis then the clinical examination in the current study will not serve to replace radiography. The clinical examination is, however, good at excluding patients without radiological disease, and if ethical issues are a problem in a population survey radiography could be directed at those who are clinically positive. This method would pick up half of all radiographically positive cases by subjecting only the clinically positive minority of the population to radiography. The high specificity of the tested clinical signs for osteoarthritis suggests that they may be useful in their own right in screening populations for clinical disease.

1 Kellgren J H, Lawrence J S. Radiological assessment of osteoarthrosis. *Ann Rheum Dis* 1957; 16: 494–502.

2 Lawrence J S, Bremner J M, Brier F. Osteoarthrosis: prevalence in the population and relationship between symptoms and x-ray changes. *Ann Rheum Dis* 1966; 25: 1–23.

3 Altman R, Asch E, Bloch D, *et al*. Development of criteria for the classification and reporting of osteoarthritis. *Arthritis Rheum* 1986; 29: 1039–49.

4 Cushnaghan J, Cooper C, Dieppe P, Kirwan J, McAlindon T, McCrae F. Clinical assessment of osteoarthritis of the knee. *Ann Rheum Dis* 1990; 49: 768–70.

5 Kellgren J H, Lawrence J S. *The epidemiology of chronic rheumatism. Atlas of standard radiographs*. Vol 2. Oxford: Blackwell Scientific, 1963.

6 Fleiss J L. *Statistical methods for rates and proportions*. 2nd ed. Chichester: Wiley, 1981.

7 Van Saase J, Van Romunde L, Cats A, Vandenbrouke J, Valkenburg H. Epidemiology of osteoarthritis: Zoetermeer survey. *Ann Rheum Dis* 1989; 48: 271–80.

8 Lawrence J S. *Rheumatism in populations*. London: Heinemann, 1977.

9 Claessens A, Schouten J, Van Den Ouweland F, Valkenburg H. Do clinical findings associate with radiographic osteoarthritis of the knee? *Ann Rheum Dis* 1990; 49: 771–4.

10 Dacre J E, Herbert K E, Perret D, Huskisson E C. The use of digital image analysis for the assessment of radiographs in osteoarthritis. *Br J Rheumatol* 1988; 27 (suppl 1): 46.

11 Cooper C, Cushnaghan J, Kirwan J, *et al*. Radiographic assessment of the knee joint in osteoarthritis. *Br J Rheumatol* 1990; 29 (suppl 1): 19.

12 Kallman D A, Wigley F M, Scott W W, Hochberg M, Tobin J D. New radiographic grading scales for osteoarthritis of the hand. *Arthritis Rheum* 1989; 32: 1584–91.

13 Danielsson L, Hernborg J. Clinical and radiological study of knee joints with osteophytes. *Clin Orthop* 1970; 69: 302–12.

14 Rogers J, Watt I, Dieppe P. Comparison of visual and radiographic detection of bony changes at the knee joint. *BMJ* 1990; 300: 367–8.

15 ARA diagnostic subcommittee for OA. An approach to developing criteria for the clinical diagnosis and classification of OA. *J Rheumatol* 1983; 10: 180–3.

16 McAlindon T, Dieppe P. Osteoarthritis: definitions and criteria. *Ann Rheum Dis* 1989; 48: 531–2.

17 ACR diagnostic subcommittee for OA. Osteoarthritis: definitions and criteria. *Ann Rheum Dis* 1990; 49: 201–3.