

Radiographic assessment of the knee joint in osteoarthritis

C Cooper, J Cushnaghan, J R Kirwan, P A Dieppe, J Rogers, T McAlindon, F McCrae

Abstract

The development of radiographic systems for the grading of osteoarthritis requires knowledge of the reproducibility of their individual component features. This paper reports the reproducibility, both within and between observers, for five commonly used radiographic features of osteoarthritis in the tibiofemoral and patellofemoral compartments of the knee joint. The results suggest that assessments of joint space narrowing, osteophyte, and bony contour in the tibiofemoral compartments are more reproducible than those of sclerosis and cyst. Patellofemoral assessments, with the exception of osteophyte, are considerably less reproducible between observers than tibiofemoral assessments.

Osteoarthritis of the knee joint is a major cause of pain and disability.¹ There is, however, no agreed definition of the condition which may be applied in clinical and epidemiological studies. The most widely adopted method for this purpose is radiographic evaluation of the knee joint. The radiographic diagnosis of osteoarthritis of the knee joint began with the five point grading system of Kellgren and Lawrence,² which was subsequently accepted by the World Health Organisation.³ The widely recognised deficiencies of this scale have led to a search for more effective scoring systems for osteoarthritis at various skeletal sites.^{4,5} Information on the reproducibility, both within and between observers, of these systems is fundamental to their development. We have, therefore, investigated the reproducibility of five radiographic features commonly used to measure osteoarthritis of the knee joint.

Patients and methods

Standing anteroposterior and lateral knee radiographs were obtained from 50 patients (14 men, 36 women) selected from a case register of patients with osteoarthritis of the knee maintained at the Bristol Royal Infirmary. They were chosen to represent a range of severity of the condition. The mean age of the patients was 69.4 years and the mean disease duration 15.9 years. Forty four patients had bilateral disease and three had unilateral disease. The remaining three patients were included on the register, but were felt on review to have minimal radiographic changes in either knee.

Five observers (one consultant, four trainees) participated in the study. They agreed on the criteria for radiographic evaluation during a half

hour training period, following which no further discussion was undertaken to improve observer agreement. The observers scored all 100 knee radiographs on two separate occasions over a three day period. The radiographs were presented to each observer by a study controller in a random order, ensuring that the observer was unaware on the second occasion of his or her initial score.

Five radiographic features were assessed in each of the three compartments (medial and lateral tibiofemoral and patellofemoral) in each knee. They comprised joint space narrowing, osteophyte formation, subchondral sclerosis, cyst formation, and abnormality of the bony contour. Each feature was graded on a three point scale (0=absent, 1=mild, 2=severe) which had been agreed during the training period. Joint space narrowing and sclerosis were assessed by eye alone according to the pre-determined scale.

Within and between observer reproducibility were estimated for the five features in each joint compartment. Within observer reproducibility was assessed (a) using all 500 pairs of observations (to examine overall reliability), and (b) for each individual observer (to examine the range of reliability within different observers). Between observer reproducibility was likewise estimated (a) for the grouped results of the five observers as a whole and (b) for each of the ten possible combinations of observers. The level of observer agreement was quantified using the κ (kappa) statistic, around which 95% confidence limits were calculated.⁶

Results

Figure 1 shows the overall within observer reproducibility of the five radiographic features in the tibiofemoral joint. The reproducibility of assessments made on right and left knees did not differ statistically significantly nor did those

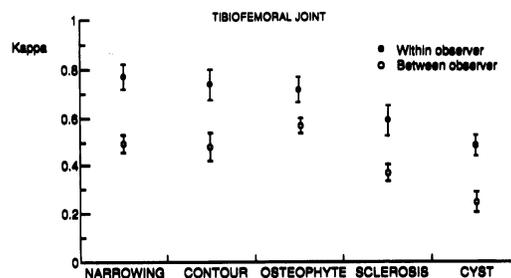


Figure 1 Within and between observer variation for radiographic assessment of the tibiofemoral compartment of the knee joint (bars show 95% confidence intervals around each κ statistic).

The Rheumatology Unit,
Department of Medicine,
Bristol Royal Infirmary,
Bristol BS2 8HW,
United Kingdom
C Cooper
J Cushnaghan
J R Kirwan
P A Dieppe
J Rogers
T McAlindon
F McCrae

Correspondence to:
Dr C Cooper,
Section of Clinical
Epidemiology,
Harwick Building,
Mayo Clinic,
Rochester, MN 55905, USA.

Accepted for publication
15 January 1991

made on the medial and lateral tibiofemoral compartments. The results for each of the four compartments have thus been amalgamated. Within observer reproducibility was greatest for joint space narrowing ($\kappa=0.77$, 95% confidence interval 0.72–0.82). Abnormalities of contour and osteophyte formation were only slightly less reproducible than joint space narrowing. Sclerosis and cyst formation, in contrast, were significantly less reproducible than each of the other three features.

Reproducibility was lower between than within observers for all five radiographic features in the tibiofemoral joint. Osteophyte showed the greatest between observer reproducibility, and again, values for joint space narrowing and contour were significantly better than those for sclerosis and cyst formation.

Figure 2 shows the within and between observer reproducibility of the five radiographic features in the patellofemoral joint. Within observer reproducibility was slightly lower than in the tibiofemoral joint. Between observer reproducibility in the patellofemoral joint, however, was significantly lower for all the features except osteophyte. In particular, assessments of patellofemoral joint space and contour were no better than those of sclerosis and cyst.

The table shows the distribution of grades for each radiographic feature in the tibiofemoral and patellofemoral compartments, using the data from the most consistent observer at each site. There was a greater overall occurrence of joint space narrowing and osteophyte than of cyst formation and abnormalities of contour.

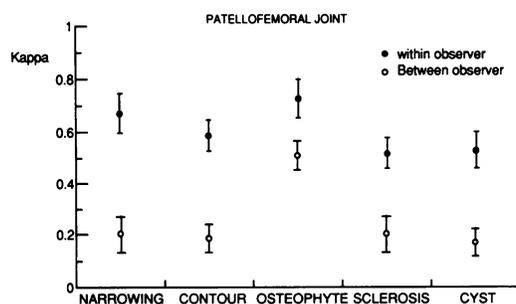


Figure 2 Within and between observer variation for radiographic assessment of the patellofemoral compartment of the knee joint (bars show 95% confidence intervals around each κ statistic).

Distribution of grades for each radiographic feature in the tibiofemoral and patellofemoral compartments, using data from the most consistent observer at each site

Radiographic feature	Grade			Total
	0	1	2	
Tibiofemoral				
Joint space narrowing	67	91	35	193*
Osteophyte	97	70	33	200
Sclerosis	130	51	19	200
Cyst	188	9	3	200
Contour	175	25	N/A†	200
Patellofemoral				
Joint space narrowing	33	20	16	69*
Osteophyte	47	32	21	100
Sclerosis	89	10	1	100
Cyst	73	19	8	100
Contour	79	21	—	100

*Remaining radiographs not classifiable by this observer.
†N/A=not applicable.

The results for the individual observers were also examined separately. It was found that the poor overall reproducibility of some features—for example, tibiofemoral sclerosis and cyst, did not result from inconsistency in the assessments of all five observers, but arose from the poorer performance of three of the five. For tibiofemoral cyst, the two most consistent observers were capable of attaining a concordance characterised by a κ statistic of greater than 0.6, whereas the least consistent observer gave values of less than 0.2.

Discussion

We have assessed within and between observer reproducibility for five radiographic features of osteoarthritis in the tibiofemoral and patellofemoral compartments of the knee joint. Our results suggest that assessments of joint space narrowing, osteophyte, and bony contour in the tibiofemoral compartment are significantly more reproducible, both within and between observers, than assessments of sclerosis and cyst formation. In the patellofemoral compartment, the only feature to show a similar overall reliability was osteophyte formation.

The κ statistic provides a tool with which to assess agreement between sets of observations above that expected by chance alone.⁷ As only three levels of severity were adopted in our radiographic grading of each variable, and discordance of greater than one category was relatively infrequent, we decided against adopting the weighted κ statistic, as this might have led to an over optimistic perception of observer agreement.

Potential bias resulting from patient selection is an important consideration in our study. We included both knees of a series of subjects who were part of a longitudinal study exploring the natural history of osteoarthritis. The 100 study radiographs were chosen by a worker, who did not subsequently perform any of the study assessments, on the basis of global evidence of mild, moderate, and severe disease in one or both knees. It was hoped that the prevalence of radiographic abnormalities in the study sample would thereby fall within the range where the κ statistic is not prevalence dependent. The frequency of certain features such as tibiofemoral cyst formation, however, remained less than 10% for all observers, and the value of the κ statistic might have been artificially depressed.

Certain methodological aspects of this study might also have influenced the results. We specifically chose not to include a radiologist as an observer, as one of our purposes was to assess consistency among the rheumatological staff who view these radiographs in day to day clinical and research settings. Joint space narrowing and sclerosis were not measured by ruler, but were assigned to three categories of severity to enable consistency between features. Investigations of hip osteoarthritis have suggested that measurement by ruler may increase the within observer reproducibility of joint space narrowing.⁸ Finally, 'standard abnormal' radiographs could not be used in our study, as they do not currently exist. Indeed, one future

aim of this work is to construct such a set using reproducibility as a criterion.

Although the association between these radiographic features has been documented in patients with osteoarthritis of the knee, clinical observations have often pointed to differences in their order of appearance and variation in their clustering.⁵⁻⁹ In spite of this, the earliest grading system of osteoarthritis, one which assumes a sequence of events in the progression of the disorder and attributes major significance to osteophyte, continues to be used in clinical and epidemiological studies.¹⁰⁻¹² Improvement or revision of this system requires information on the reproducibility of its component radiographic features. Such data are only now beginning to accumulate. As part of an investigation aiming to improve the assessment of radiographic progression in osteoarthritis, Altman *et al*⁴ evaluated the reproducibility of joint space narrowing, spurs, alignment, attrition, and sclerosis in 32 pairs of knee radiographs. Their findings for narrowing and osteophyte (spur) were similar to those reported here, although assessment of contour was more reproducible in our study. Kallman *et al*⁷ reported the reproducibility of a series of assessments on hand radiographs from patients with osteoarthritis. Although the reproducibility for all features tended to be greater for the hand than the knee, the higher ranking of narrowing and osteophyte was again observed.

This is the first study attempting to evaluate the radiographic features of osteoarthritis in the patellofemoral joint. Isolated patellofemoral joint osteoarthritis is a recognised disease, although its clinical and epidemiological characteristics remain obscure.¹³ The frequently observed discrepancy in epidemiological studies between knee pain and radiographic evidence of osteoarthritis¹⁴ may result, in part, from a failure to examine lateral knee radiographs. Our findings suggest that the consistency of assess-

ment between observers for features at this site other than osteophyte, needs to be improved. Alterations in the radiographic technique, perhaps with the inclusion of skyline views of the joint,¹⁵ and training of observers, represent potential means to improve reproducibility at this site.

This work was supported by the Arthritis and Rheumatism Council. We are indebted to Dr I Watt for assistance in definition of the radiographic features.

- 1 Felson D. Epidemiology of hip and knee osteoarthritis. *Epidemiol Rev* 1988; 10: 1-28.
- 2 Kellgren J H, Lawrence J S. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957; 16: 494-501.
- 3 *The epidemiology of chronic rheumatism. Atlas of standard radiographs*. Vol. 2. Oxford: Blackwell Scientific, 1963.
- 4 Altman R D, Fries J F, Bloch D A, *et al*. Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987; 30: 1214-25.
- 5 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.
- 6 Fleiss J L. The measurement and control of misclassification error. In: Fleiss J L, ed. *Statistical methods for rates and proportions*. Chichester: Wiley, 1981: 140-53.
- 7 Cohen J. A coefficient of agreement for nominal scales. *Educational Psychology Measurements* 1960; 20: 37-46.
- 8 Croft P, Cooper C, Wickham C, Coggan D. Defining osteoarthritis of the hip for epidemiological studies. *Am J Epidemiol* 1990; 32: 514-22.
- 9 Wood P H N. Osteoarthritis in the community. *Clin Rheum Dis* 1976; 2: 495-507.
- 10 Acheson R M, Chan Y K, Clemett A R. New Haven survey of joint diseases. XII. Distribution and symptoms of osteoarthritis in the hands with reference to handedness. *Ann Rheum Dis* 1970; 29: 275-85.
- 11 Felson D T, Naimark A, Anderson J, Kazis L, Castelli W, Meenan R F. The prevalence of knee osteoarthritis in the elderly. The Framingham osteoarthritis study. *Arthritis Rheum* 1987; 30: 914-8.
- 12 Van Saase J L C M, Van Romunde L K J, Cats A, Vandenbroucke J P, Valkenburg H A. Epidemiology of osteoarthritis: Zoetermeer survey. Comparison of radiological osteoarthritis in a Dutch population with that in 10 other populations. *Ann Rheum Dis* 1989; 48: 271-80.
- 13 Ahlback S. Osteoarthritis of the knee. A radiographic investigation. *Acta Radiol Suppl (Stockh)* 1968; 277: 1-29.
- 14 Hochberg M C. Epidemiologic association of pain in osteoarthritis of the knee: data from the National Health and Nutrition Examination Survey and the National Health and Nutrition Examination—I Epidemiologic Follow-up Survey. *Semin Arthritis Rheum* 1989; 18 (suppl 2): 4-9.
- 15 Laurin C A, Dussault R, Levesque H P. The tangential x-ray investigation of the patellofemoral joint. X-ray technique, diagnostic criteria and their interpretation. *Clin Orthop* 1979; 144: 16-26.