

Clinical assessment of osteoarthritis of the knee

Janet Cushnaghan, Cyrus Cooper, Paul Dieppe, John Kirwan, Timothy McAlindon, Fiona McCrae

Abstract

The repeatability of physical signs used to assess osteoarthritis of the knee has not been systematically examined. The within and between observer variation of 10 commonly used physical signs to determine osteoarthritis of the knee has been assessed here. The results obtained show variation in the repeatability of these signs. For those examining the tibiofemoral joints the repeatability was greater than for those examining the patellofemoral joint. It would therefore seem vital to take note of the repeatability of physical signs in determining the number of subjects to be studied in epidemiological studies and therapeutic studies in osteoarthritis.

Osteoarthritis of the knee is recognised as a major cause of pain and disability.¹ There is, however, no clear agreement on the most appropriate method for diagnosing the condition.² Clinical assessments are often combined with radiographic evaluation for this purpose. Despite the widespread use of clinical assessments in research into the pathogenesis and treatment of osteoarthritis, their reproducibility remains unknown. Precautions to reduce, or even measure, within and between observer variation have not been taken in many published studies of osteoarthritis of the knee. We have therefore investigated the within and between observer variation of 10 commonly used physical signs of knee osteoarthritis.

Patients and methods

Eight patients (six female, two male) were selected for participation in the study from a case register of knee osteoarthritis maintained at the Bristol Royal Infirmary. They were chosen to represent a range of severity of the condition. Their mean age was 64 years and mean disease duration 12.5 years.

Five observers (two consultants, three trainees) participated in the study. They had varying degrees of experience in rheumatology. The observers agreed on the examination techniques to be used to elicit the 10 physical signs, but no training was undertaken to try to improve between observer agreement.

In one three hour session both knees of each patient were examined twice by each observer in constrained random order, supervised by a study controller. At least 30 minutes and several other patient examinations elapsed between the two examinations made by any one observer of any one patient. The following physical signs were graded present or absent: bony swelling, non-bony swelling, tibiofemoral crepitus, patellofemoral crepitus, medial tenderness, lateral tenderness, patellofemoral tenderness, medio-lateral instability, anteroposterior instability. Active ranges of flexion and extension were measured with a goniometer.

Within and between observer reproducibility were estimated for the 10 variables. Within observer reproducibility was assessed (a) using all 80 pairs of observations (to examine overall reliability ignoring the different observers), 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 10 possible combinations of observers (A v B, and B v C, etc).

The level of observer agreement was measured using the kappa (κ) statistic³ for dichotomous variables and by estimating mean differences between paired observations for continuous variables.

Results

Table 1 shows the overall within observer agreement for the nine dichotomous variables on 80 replicate readings. Reproducibility was best for the detection of bony swelling ($\kappa=0.74$)

Table 1 Within observer variation for examination of knee osteoarthritis

Physical sign	Overall agreement		Observer				
	κ	95% CI*	A	B	C	D	E
Bony swelling	0.74	0.52 to 0.56	0.75	0.71	0.75	0.75	0.71
Tibiofemoral crepitus	0.68	0.44 to 0.92	0.54	1.00	0.60	0.48	0.71
Non-bony swelling	0.67	0.45 to 0.89	0.87	0.85	0.87	0.25	0.41
Medial tenderness	0.64	0.40 to 0.88	1.00	0.48	0.47	0.63	0.35
Mediolateral instability	0.55	0.30 to 0.70	0.85	0.20	0.60	0.26	0.70
Anteroposterior instability	0.53	0.29 to 0.77	0.29	0.60	1.00	0.30	0.30
Patellofemoral crepitus	0.50	0.26 to 0.74	0.28	0.71	0.48	0.25	0.61
Lateral tenderness	0.50	0.26 to 0.74	0.31	0.75	0.42	0.75	0.50
Patellofemoral tenderness	0.41	0.30 to 0.52	0.50	0.38	0.38	0.63	0.10

*CI=confidence interval.

The Rheumatology Unit,
Bristol Royal Infirmary,
Bristol BS2 8HW

J Cushnaghan
C Cooper
P Dieppe
J Kirwan
T McAlindon
F McCrae

Correspondence to:
Mrs Cushnaghan.

Accepted for publication
9 November 1989

Table 2 Between observer variation for examination of knee osteoarthritis

Physical sign	Overall agreement		Observer combinations									
	κ	95% CI*	A/B	A/C	A/D	A/E	B/C	B/D	B/E	C/D	C/E	D/E
Tibiofemoral crepitus	0.64	0.48 to 0.80	0.54	0.74	0.71	0.33	0.60	0.85	0.82	0.74	0.46	0.67
Bony swelling	0.55	0.40 to 0.70	0.74	0.75	0.61	0.40	0.48	0.59	0.31	0.61	0.61	0.47
Lateral tenderness	0.43	0.28 to 0.58	0.31	0.42	0.38	0.41	0.59	0.75	0.43	0.38	0.23	0.50
Medial tenderness	0.40	0.26 to 0.54	0.28	0.64	0.52	0.28	0.31	0.42	0.02	0.31	0.26	0.35
Patellofemoral pain/tenderness	0.35	0.20 to 0.50	0.50	0.25	0.88	0.47	0.25	0.63	0.50	0.31	0.38	0.61
Non-bony swelling	0.28	0.12 to 0.44	0.43	0.47	0.29	0.35	0.20	0.59	0.21	0.29	0.39	-0.16
Patellofemoral crepitus	0.24	0.10 to 0.38	0.03	0.12	0.31	0.25	0.20	0.47	0.64	-0.12	0.10	0.39
Mediolateral instability	0.23	0.06 to 0.40	0.60	0.33	0.33	0.39	0.20	-0.09	0.60	-0.11	0.00	-0.11
Anteroposterior instability	0.00	-0.11 to 0.11	0.09	0.29	0.38	0.06	0.20	-0.14	-0.20	0.54	-0.14	0.48

*CI=confidence interval.

Table 3 Within observer variation for flexion and extension of the knee

	Observers				
	A	B	C	D	E
Flexion					
Group mean (degrees)	123	114	117	113	112
SD	19	18	20	18	20
Mean difference between paired readings (degrees)	2.1	0.6	-2.4	1.6	-4.0
95% CI*	(-10.5 to 14.7)	(-11.3 to 12.5)	(-17.5 to 12.7)	(-9.0 to 12.2)	(-16.1 to 8.1)
Critical difference (degrees)	17.8	16.8	21.4	14.9	17.1
Extension					
Group mean (degrees)	4.4	-0.6	0.9	2.5	1.2
SD	4.2	8.2	6.5	3.8	3.3
Mean difference between paired readings (degrees)	-0.9	1.92	1.0	0.8	0.6
95% CI*	(-6.2 to 4.4)	(-10.1 to 13.1)	(-7.9 to 9.9)	(-4.2 to 5.7)	(-13.4 to 14.7)
Critical difference (degrees)	7.5	17.0	12.6	7.1	19.8

*CI=confidence interval.

and worst for the detection of patellofemoral tenderness ($\kappa=0.41$). For certain signs—for example, bony swelling, the within observer agreement was consistent among the five observers. For others, such as anteroposterior instability, there was wide variation between the results of the five observers.

Table 2 shows that for any particular variable the degree of overall within observer agreement was higher than the degree of between observer agreement. Overall between observer agreement was greatest for tibiofemoral crepitus ($\kappa=0.64$) and least for anteroposterior instability ($\kappa=0.00$). Table 2 also shows the 10 estimates of the κ statistic, which were available for each variable and which assessed each possible combination of the five observers.

In the assessment of ranges of knee movement there was no statistically significant systematic difference within observers for flexion or extension (table 3). There was, however, considerable variation in the difference between paired observations, as indicated by the width of the 95% confidence intervals around the mean differences shown in table 3. The numbers of subjects in the study were insufficient for similar statistical analysis of between observer variation for flexion and extension. The critical difference⁴ for knee flexion varied from 14.9 to 21.4 degrees, and for knee extension from 7.1 to 19.8 degrees.

Discussion

The results of this study suggest considerable variation both within and between observers in their interpretation of 10 commonly used physi-

cal signs of knee osteoarthritis. This variation is considerably lower for certain signs—for example, bony swelling, than for others such as those assessing the patellofemoral joint.

Clinical and epidemiological studies of knee osteoarthritis require reproducible methods for the diagnosis of the condition. The American Rheumatism Association (ARA) has recently proposed a set of criteria for its definition.⁵ In an analysis of 23 clinical and laboratory features thought to warrant inclusion, five were found to possess the greatest sensitivity and specificity in the discrimination of osteoarthritis from other rheumatic disorders of the knee. These included age over 50 years, crepitus, bony enlargement, duration of early morning stiffness, and radiographic osteophytosis. It is noteworthy that the two physical signs among these five features were found in our study to possess high reproducibility, both within and between observers. This might explain, in some part, their higher discriminative ability in the ARA study.

For certain physical signs, although overall reproducibility was high, individual observers performed poorly. Some of these between individual differences are a result of random variation in the κ statistic. Some, however, may represent true differences between examiners in their reproducibility. This points to a potential role for training to reduce observer variation. Standardisation of examination technique and training have been shown to reduce such variation in examination for joint tenderness in rheumatoid arthritis.⁶

Measurement error is an important contributor to the large sample sizes required in trials

using clinical end points. The critical difference of around 20 degrees for knee flexion in this study implies that any trial using that measure as an end point will need to be of sufficient size and duration to show changes of greater than that value in order to attain biological significance. Reducing observer variation through training represents a potentially useful means of restricting the size of such trials.

Observer variation in clinical evaluation of the knee joint has been assessed in two previous studies.^{7,8} Marks *et al* investigated the reproducibility of nine physical signs in patients with rheumatoid arthritis.⁷ Although differences in analytical methods make comparison with our results difficult, within observer variation was found to be low for most signs. Spector *et al* reported a κ statistic of 0.62 for the between observer variation in assessment of knee tenderness or pain on movement.⁸

Some caution is required in the interpretation of our findings. We performed all of the assessments on a single day to avoid bias from fluctuations in the disease. Inevitably, therefore, the number of patients adopted in the study was small and the precision with which we could assign reproducibility (as evidenced by the 95% confidence intervals around κ statistics) was limited. This was particularly the case when the results of individual observers or pairs of observers were examined separately. Further, the results of any study of reproducibility will be influenced by the selection of subjects and observers included. We used patients with a range of severity of osteoarthritis, who were representative of typical attenders at a rheuma-

tology outpatient department. Our observers comprised two consultants and three trainees in rheumatology.

In conclusion, it is possible to assess both within and between observer variation in clinical examination of the knee joint for osteoarthritis. Our results suggest that such variation is higher for signs relating to the patellofemoral joint and lower for most tibiofemoral indices, most notably bony swelling and crepitus. Until more clearly defined protocols are derived for the eliciting of these clinical signs it would seem prudent for clinical investigators to assess the reproducibility of their own measurements.

This work was supported by the Arthritis and Rheumatism Council.

- 1 Hadler N M. Osteoarthritis as a public health problem. *Clin Rheum Dis* 1985; 11: 175-85.
- 2 McAlindon T, Dieppe P. Osteoarthritis: definitions and criteria. *Ann Rheum Dis* 1989; 48: 531-2.
- 3 Cohen J A. Coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960; 20: 37-46.
- 4 Kirwan J R, Byron M A, Winfield J, Altman D G, Gumpel J M. Circumferential measurements in the assessment of synovitis of the knee. *Rheumatology and Rehabilitation* 1979; 18: 78-84.
- 5 Altman R, Asch E, Bloch D, *et al*. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. *Arthritis Rheum* 1986; 29: 1039-49.
- 6 Klinkhoff A V, Bellamy N, Bombardier C, *et al*. An experiment in reducing interobserver variability of the examination for joint tenderness. *J Rheumatol* 1988; 15: 492-4.
- 7 Marks J S, Palmer M K, Burke M J, Smith P. Observer variation in the examination of knee joints. *Ann Rheum Dis* 1978; 37: 376-7.
- 8 Spector T D, Brown P, Hart D, Wilson P, Silman A J, Doyle D. The clinical diagnosis of osteoarthritis in the general population [Abstract]. *Br J Rheumatol* 1988; 27 (suppl 2): 17.