

Observer variation in the examination of knee joints

J. S. MARKS,¹ M. K. PALMER,² M. J. BURKE,³ AND P. SMITH³

From the Department of Rheumatology, Withington Hospital,¹ and Christie Hospital and Holt Radium Institute,² Manchester, and Rheumatism Research Centre, Manchester Royal Infirmary³

SUMMARY The knees of 20 patients with rheumatoid arthritis were each examined on two occasions on the same morning by three observers. Assessments were made of joint warmth, synovial thickening, effusion, instability, quadriceps power, bony enlargement, range of movement, and knee circumference, and were graded on a scale of 0-4. Analysis of the results shows close intraobserver agreement for all measurements but considerable interobserver variation for all subjective measurements.

Clinical examination of joints is important in the assessment of patients with rheumatic disorders. The reliability of any method of assessment depends on the error associated with its use and in order to quantify this it is necessary to document the intra- and the interobserver variation. The existence of such variability has been established for respiratory diseases (Gough *et al.*, 1952), cardiovascular disorders (Rose *et al.*, 1964), and in the interpretation of radiographs (Kellgren, 1956; Macrae *et al.*, 1971), but there have been few studies on the reliability of clinical assessment in the rheumatic diseases apart from the measurement of digital joint size with jewellers' rings (Hart and Clark, 1951), grip strength (Lee *et al.*, 1974), spinal mobility (Moll and Wright, 1971), and of knee circumference with a tape measure (Nicholas *et al.*, 1976).

We report a study of observer errors in the clinical assessment of inflammation in the knees of patients with rheumatoid arthritis.

Methods

Twenty female patients with rheumatoid arthritis were studied. One patient had a plaster knee cylinder so that a total of 39 knees were examined. None of the observers (M.J.B., J.S.M., P.S.) had previously seen any of the patients who were examined in a randomised order on two occasions by each observer on the same morning.

Assessment was made on each occasion of warmth, synovial thickening, joint effusion, anteroposterior and lateral instability, quadriceps lag, and bony enlargement. These observations were graded as absent

0, doubtful 1, mild 2, moderate 3, severe 4. The midpatella circumference was recorded with a tape measure and the range of movement was measured by visual observation without the use of a goniometer.

Results

Intraobserver variation was determined by comparing the joint scores for each feature measured by an observer on the two occasions on which the same knee was examined. Table 1 shows the number of knees out of 39 examined which were given the same grade on both occasions (N), or which differed by two grades or more on both occasions (n), and the correlation coefficient (r). As absolute figures were obtained for joint circumference and range of movement the correlation coefficients only have been calculated.

Table 1 *Comparing occasions for each observer*

	Observer 1			Observer 2			Observer 3		
	N	n	r	N	n	r	N	n	r
Warmth	17	2	0.47	33	2	0.87	32	2	0.83
Thickening	35	0	0.91	34	0	0.90	34	1	0.79
Effusion	28	0	0.86	36	0	0.97	26	5	0.69
AP instability	22	3	0.54	26	2	0.69	27	5	0.60
Lateral instability	23	1	0.82	31	3	0.91	30	3	0.86
Quadriceps lag	25	4	0.81	36	2	0.91	32	0	0.95
Bony enlargement	29	0	0.81	18	3	0.65	26	3	0.42
Range of movement			0.91			0.97			0.97
Midpatella circumference			0.96			0.97			0.96

N = number of knees out of 39 given the same grade on 2 occasions.
n = number of knees in which the grade differed by 2 grades or more on 2 occasions.

Accepted for publication September 25, 1977
Correspondence to Dr J. S. Marks, Rheumatism Research Centre, The Royal Infirmary, Manchester M13 9WL

Table 2 Comparing observers

	Observer 1 v observer 2			Observer 1 v observer 3			Observer 2 v observer 3		
	N	n	r	N	n	r	N	n	r
	Warmth	17	9	0.52	19	9	0.57	25	8
Thickening	20	1	0.54	16	3	0.19	21	4	0.57
Effusion	18	6	0.52	18	4	0.65	25	8	0.57
AP instability	17	4	0.52	21	2	0.50	20	7	0.43
Lateral instability	25	5	0.77	19	6	0.69	23	8	0.73
Quadriceps lag	29	8	0.66	27	4	0.83	26	8	0.72
Bony enlargement	13	12	0.58	4	13	0.28	12	8	0.26
Range of movement			0.90			0.91			0.92
Midpatella circumference			0.96			0.94			0.93

Interobserver variation was determined by comparing the difference in joint scores between two observers on the first occasion in which the same knee was examined. Table 2 shows the number of knees which the two observers gave the same grade (N) or differed by two grades or more (n), and the correlation coefficient (r).

Discussion

The results of this study show very close intra-observer agreement for all measurements and in particular for range of movement, knee circumference, synovial thickening, and joint effusion. There was considerable interobserver variation for all subjective measurements which was greatest for bony enlargement and least for lateral instability and

quadriceps lag. There was, however, close inter-observer agreement for the objective measurements, namely range of movement and knee circumference.

These findings indicate that in serial assessments of the state of a joint it is of crucial importance that the observations should be made by the same observer. Providing this criterion is met it appears that these easily performed clinical measurements are reliable and of value in clinical trials.

We are most grateful to Dr James Sharp for helpful advice throughout all stages of this study.

References

- Gough, J., Fletcher, C. M., Gilson, J. C., and Oldham, P. D. (1952). Discussions on the diagnosis of emphysema. *Proceedings of the Royal Society of Medicine*, **45**, 576-586.
- Hart, F. D., and Clark, C. J. M., (1951). Measurement of distal swelling in rheumatoid arthritis. *Lancet*, **1**, 775-777.
- Kellgren, J. H. (1956). A study of observer differences in the reading of hand films. *Annals of the Rheumatic Diseases*, **15**, 55-60.
- Lee, P., Baxter, A., Dick, W. C., and Webb, J. (1974). An assessment of grip strength measurement in rheumatoid arthritis. *Scandinavian Journal of Rheumatology*, **2**, 71-77.
- Macrae, I. F., Haslock, D. I., and Wright, V. (1971). Grading of films for sacroiliitis in population studies. *Annals of the Rheumatic Diseases*, **30**, 58-66.
- Moll, J. M. H., and Wright, V. (1971). Normal range of spinal mobility. *Annals of the Rheumatic Diseases*, **30**, 381-386.
- Nicholas, J. J., Taylor, F. H., Buckingham, R. B., and Ottonello, D. (1976). Measurement of circumference of the knee with an ordinary tape measure. *Annals of the Rheumatic Diseases*, **35**, 282-284.
- Rose, G. A., Holland, W. W., and Crowley, E. A. (1964). A sphygmomanometer for epidemiologists. *Lancet*, **1**, 296-300.