

## EXTENDED REPORTS

## Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies

Guy-Robert Auleley, Bruno Giraudeau, Maxime Dougados, Philippe Ravaud

### Abstract

**Objective**—To compare radiographic reading procedures and evaluate their impact on sample size in hip osteoarthritis (OA) longitudinal studies.

**Methods**—Pelvic radiographs performed twice, three years apart, in 104 patients with hip OA were read by a single reader using the Kellgren and Lawrence system, joint space narrowing scale, and joint space width (JSW). Reading procedures were (a) films read as single radiographs, (b) films grouped by patient but read in random order, (c) films grouped by patient and chronologically ordered, all with landmarks for JSW measurements, (d) films read as single radiographs, without landmarks for JSW measurements. JSW was measured at the narrowest point with a 0.1 mm graduated magnifying glass.

**Results**—More Kellgren and Lawrence or joint space narrowing grades were modified respectively with the single (42% and 37%) than with the paired (32% and 23%) or chronologically ordered (34% and 29%) reading procedures. Variability of JSW progression was principally related to mean progression (88.3%) and landmarks (almost 10%). Standardised response means were  $-0.71$  with the paired reading procedure with landmarks,  $-0.68$  with the single reading procedure with landmarks,  $-0.65$  with the single reading procedure without landmarks. With landmarks, 10% more patients would be needed using single than paired reading. Using single reading, 10% more patients would be needed without landmarks than with landmarks.

**Conclusion**—Kellgren and Lawrence grading seems to be influenced by the reading procedure, as is joint space narrowing grading, for assessing hip OA. Paired reading procedure with landmarks for JSW should be recommended in longitudinal studies.

(Ann Rheum Dis 2000;59:422-427)

osteoarthritis (OA).<sup>1-3</sup> Depending on the joint studied, several indices are currently used for assessing radiological progression of OA, including individual radiographic features, composite indices, and quantitative measures.<sup>4</sup> The Kellgren and Lawrence grading system is often used despite its limitations.<sup>5-7</sup> Joint space narrowing, recorded both as measurements of interbone distance or by visual grading, is presently one of the most common variables recommended.<sup>1,2</sup> However, methodological problems remain when using this index. Firstly, when measuring joint space width (JSW) for assessing OA radiological progression, landmarks for measurements may be drawn on radiographs<sup>7,8</sup> or not.<sup>9,10</sup> The implications of drawing landmarks have not yet been studied. Secondly, it has been recommended that the date and identification of the patient of radiographs should not be known when assessing the progression of OA,<sup>11</sup> and the type of blinding differs in studies of OA progression.<sup>7-10,12,13</sup> Thus radiographic OA progression might be assessed with the reader aware of neither the patient's identity nor the chronological order of the radiographs (single reading procedure), aware of the patient's identity but unaware of the chronological order of the radiographs (paired reading procedure), or aware of both the patient's identity and the chronological order of the radiographs (chronologically ordered reading procedure). Yet these radiographic reading procedures seem to have different implications on the power of studies.<sup>14,15</sup> In clinical trials on rheumatoid arthritis, 11.8% and 38% more patients would be needed to detect the same progression difference of joint damage and joint erosion scores, respectively, when reading paired films chronologically ordered rather than reading paired films in random order.<sup>14</sup>

These different reading procedures have not been compared in the assessment of OA progression. Additionally, the impact on the design of OA studies of blindness or drawing landmarks for measuring radiological progression has not been evaluated. The aims of this study were (a) to compare different reading procedures and (b) to evaluate the impact of these reading procedures and using landmarks to measure JSW on sample size, for longitudinal evaluation of hip OA.

Faculté de Médecine  
Cochin Port-Royal and  
Clinique de  
Rhumatologie, Hôpital  
Cochin, AP-HP, Paris,  
France  
G-R Auleley  
M Dougados  
Ph Ravaud

Centre de Recherche  
Clinique, Faculté de  
Médecine, Tours,  
France  
B Giraudeau

Correspondence to:  
Dr Ph Ravaud, Clinique de  
Rhumatologie, Hôpital  
Cochin, 27 rue du faubourg  
Saint Jacques, 75014 Paris,  
France  
Email: philippe.Ravaud@  
cch.ap-hop-paris-fr

Accepted for publication  
25 February 2000

Structural morphological changes on radiographs are considered the primary outcome variables for assessing the progression of

## Patients and methods

### PATIENTS

A sample of 104 patients fulfilling the American College of Rheumatology clinical and radiographic criteria for the diagnosis of hip OA including hip pain with at least two of the following: joint space narrowing, osteophytes, and erythrocyte sedimentation rate  $<20$  mm/1st h,<sup>16</sup> were selected from a three year randomised, controlled trial. In this randomised trial, other inclusion criteria were: age between 50 and 75 years, presence of daily pain for at least one month in the past three months, absence of secondary hip OA (presence or past history of hip fracture, inflammatory rheumatic disease, osteonecrosis, Paget's disease, etc),<sup>17</sup> a JSW larger than 1 mm at the narrowest point, absence of medial or axial femoral head migration, or both, on radiographs, obtaining written informed consent. To enter our study, patients were selected using the following criteria: available radiographs for three years and absence of hip prosthesis on the target joint.

### RADIOGRAPHS

Each patient had plain pelvic radiograph at entry into the study and at three years. Antero-posterior radiograph was performed with the patient standing on both legs. The patient's feet were  $15^\circ \pm 5^\circ$  internally rotated. The x ray beam was horizontal, perpendicular to the table. The source to film distance was 100 cm. Thus 208 radiographs—that is, two radiographs for each patient were obtained.

### RADIOGRAPHIC ASSESSMENT

Radiographs were assessed by one reader (GRA). For quantitative measurement, the interbone distance at the narrowest point was measured in millimetres using a 0.1 mm graduated magnifying glass laid directly over the radiograph. In addition, joint space narrowing was graded 0–3 using a radiographic atlas.<sup>18</sup> Overall severity of OA was graded using the Kellgren and Lawrence grading scale<sup>19</sup> defined as follows: 0 = normal; 1 = doubtful narrowing of joint space and possible osteophytic lipping; 2 = definite osteophytes and possible narrowing of joint space; 3 = moderate multiple osteophytes, definite narrowing of joint space, some sclerosis, and possible deformity of bone contour; 4 = large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone contour.

### READING PROCEDURES

Four reading sessions were performed one week apart. At each session the reader used a different reading procedure and was unaware

of the results of the previous sessions. Radiographs were read as single, paired, or chronologically ordered depending on whether the patient's identity or time sequence, or both, were known or unknown (table 1).

For the first three reading procedures, hip OA progression was assessed using the Kellgren and Lawrence grading scale, joint space narrowing grading scale, and JSW measurement. JSW was measured with landmarks. Landmarks were drawn before the first reading session. To draw landmarks, the two radiographs of each patient were placed side by side on a light box and landmarks were immediately drawn. The landmarks consisted of two points, one on the distal margin of the condylar cortex for the femoral surface and the other on the margin of the bright radiodense band of the subchondral cortex in the floor of the articular fossa for the acetabulum.

For the single reading procedure all 208 radiographs were read as single. Patient's identity and date of radiographs were masked with adhesive tape using a randomisation list.

For the paired reading procedure, radiographs performed at entry and at three years were grouped by patient. Only the date of radiographs was masked with adhesive tape using another randomisation list. Pairs of radiographs of the same patient (n=104) were read side by side in random order, the chronological order being unknown to the reader.

For the chronologically ordered reading procedure, radiographs were grouped by patient and read by pairs chronologically ordered, the chronological order being known to the reader.

After these first three reading sessions, landmarks were erased.

For the single reading procedure without landmarking the 208 radiographs were read blindly as single and JSW measurements were made without previous landmarks.

### STATISTICAL ANALYSIS

To compare the reading procedures when assessing hip OA progression by Kellgren and Lawrence and joint space narrowing grading scales, we constructed histograms of changes in grades (values at three years minus values at entry).

The effects of reading procedures when assessing hip OA progression by JSW measurements were assessed using the same methods. Comparisons of paired, chronologically ordered, single reading procedures with landmarks and single reading procedure without landmarks were performed using descriptive statistics. These included means and standard deviations of the differences between the values of JSW measured on the last and the first radiographs of the same patient (values at three years minus values at entry) and histograms showing the progression of JSW.

A principal component analysis was also performed. The object of this analysis was to take the four values of change in JSW measurements (that is, one value for each reading procedure) for each patient and to find combinations of these to produce mutually uncorrelated indices (named principal compo-

Table 1 Sessions and procedures of reading

Reading sessions	Landmarks	Patient identity	Time sequence	Reading procedures
First	Present	Blinded	Blinded	With landmarks
Second	Present	Known	Blinded	Single
Third	Present	Known	Known	Paired
				Chronologically ordered
Fourth	Absent	Blinded	Blinded	Without landmarks
				Single

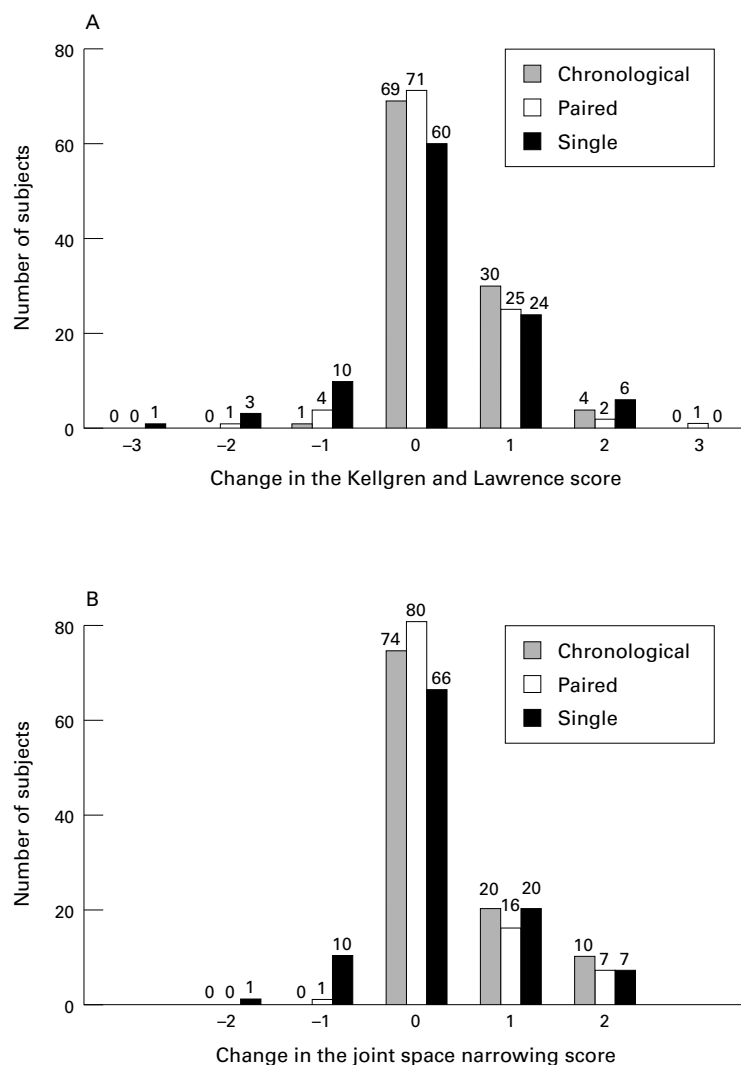


Figure 1 Distribution of changes in Kellgren and Lawrence score and joint space narrowing score using three reading procedures in 104 patients with hip osteoarthritis.

nents). This lack of correlation between principal components then allowed better understanding of the differences emerging from the four reading procedures. Moreover, principal components were ordered so that the first component explained the largest amount of total variability, the second component explained the second largest amount of total variability, and so on. Estimations of intraclass correlation coefficients (ICCs) of pairs of reading procedures were derived in the framework of a two way random effect model. Approximate 95% confidence intervals were estimated by the Fleiss and Shrout result.<sup>20</sup>

Table 2 Progression of joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis followed up for three years

	Reading procedure			
	Chronologically ordered with landmarks	Paired with landmarks	Single with landmarks	Single without landmarks
Mean (mm)	-0.60	-0.59	-0.58	-0.47
Standard deviation (mm)	0.85	0.83	0.85	0.73
Standardised response mean*	-0.70	-0.71	-0.68	-0.65

\*Standardised response mean is the ratio of mean and standard deviation of change.

Standardised response mean (SRM)—that is, a responsiveness statistic indicating the magnitude of change (mean difference) in comparison with the standard deviation of change,<sup>21</sup> was estimated for chronologically ordered, paired, and single reading procedures with landmarks or single reading procedure without landmarks. As it has been shown that high responsive instruments lead to lower sample size requirements,<sup>22</sup> the effects of these reading procedures on sample size requirements when using a paired Student's *t* test were evaluated by comparing the corresponding SRM.

## Results

At baseline, patients had a mean (standard deviation) age of 62.3 (7.2) years, weight 69.9 (11.8) kg, and height 165.5 (9.1) cm.

When assessed with the single reading procedure, the Kellgren and Lawrence grade changed in 44 patients (42%) compared with change in 35 (34%) and 33 (32%) patients when assessed with the chronologically ordered or paired reading procedures, respectively (fig 1). Of note, on the Kellgren and Lawrence scale 14 patients had an improvement with the single reading procedure against only one with the chronologically ordered reading procedure. When assessed with the single reading procedure, the joint space narrowing grade changed in 38 patients (37%) compared with change in 30 (29%) and 24 (23%) patients when assessed with the chronologically ordered or paired reading procedures, respectively. Although an improvement in Kellgren and Lawrence or joint space narrowing grading scales by more than one grade (change towards a lesser grade) was rarely seen (<10%), the frequency of progression differed according to the radiographic feature and reading procedure. Improvement on the Kellgren and Lawrence and joint space grading scale was more often seen with the single reading procedure than with the other reading procedures. Except for two patients, those in whom improvement was seen on the Kellgren and Lawrence grading scale differed from those in whom improvement was seen on the joint space narrowing grading scale. For these two patients, improvement was seen only with the single reading procedure.

Table 2 and fig 2 show JSW progression according to the reading procedure. The progression was less with the single reading procedure without landmarks (-0.47 mm) than with the other reading procedures (at least -0.58 mm for the single with landmarks). Progression standard deviation was also lower with the single reading procedure without landmarks (0.73 mm) than with the other reading procedures (0.85 mm for the single with landmarks).

Figure 3 shows the results of principal component analysis. The two first principal components represented almost 98% of the total variability of JSW progression. The first principal component was responsible for 88.3% of the total variability and corresponded to the progression of JSW. The second principal

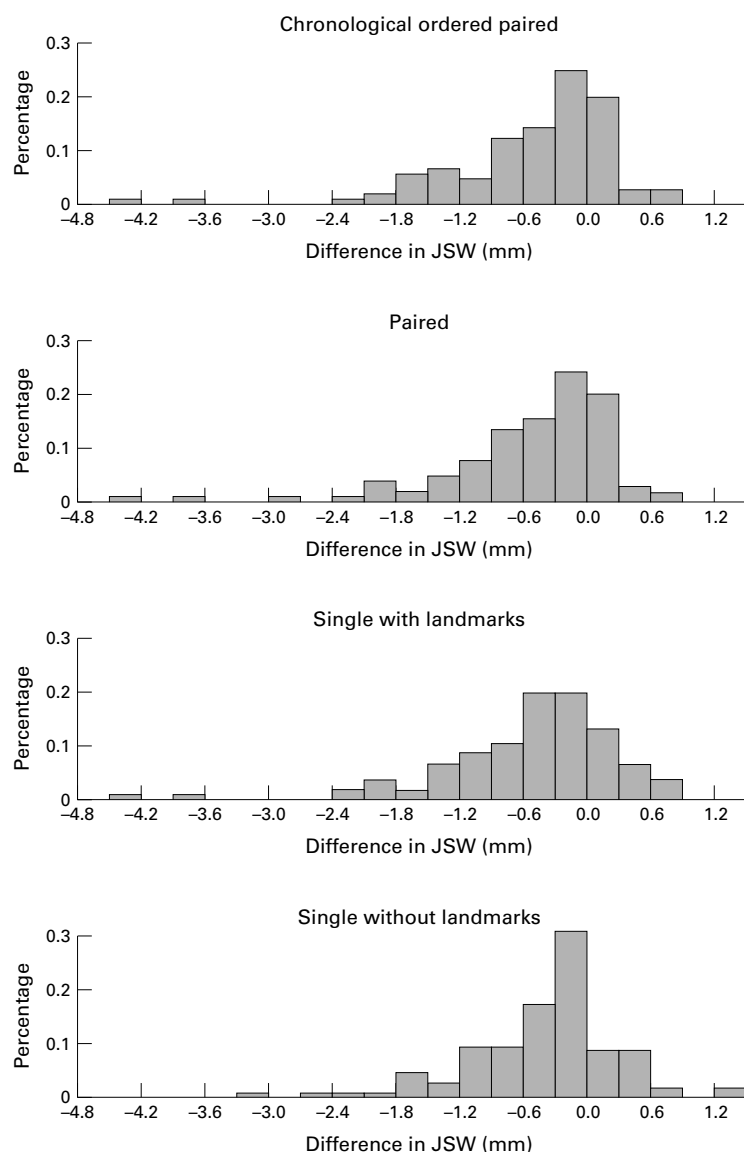


Figure 2 Distribution of the progression in joint space width (JSW) using four reading procedures in 104 patients with hip osteoarthritis.

component contrasted the reading procedures using landmarks with the procedure without landmarks and was responsible for about 10% of the total variability.

Table 3 gives details of the ICCs and their corresponding 95% confidence intervals. There was a high agreement between reading procedures performed with landmarks as ICCs reached 0.96. The agreement between each of the three reading procedures with landmarks and the one without landmarks was less, though the corresponding ICCs remained at about 0.70.

The paired reading procedure of JSW measurements with landmarks had the highest SRM (-0.71), whereas the reading procedure without landmarks had an SRM of -0.65 (table 2). The ratio between estimated sample sizes when comparing two reading procedures actually amounts to the square of the ratio of the corresponding SRM. This ratio equalled 1.10 or 1.06 when comparing the paired reading procedure with landmarks or the chronologi-

cally ordered reading procedure with landmarks with the single reading procedure with landmarks, respectively. Therefore, 10% or 6% more patients would be needed when using the single reading procedure with landmarks than when using the paired reading procedure with landmarks or the chronologically ordered reading procedure with landmarks, respectively. When comparing the paired reading procedure with landmarks with the chronologically ordered reading procedure with landmarks, this ratio equalled 1.03. That is, when using the chronologically ordered reading procedure with landmarks, 3% more patients would be needed.

The ratio between estimated sample sizes when using the single reading procedure with and the single reading procedure without landmarks equalled 1.10. That is, 10% more patients would be needed than when using the single reading procedure without landmarks in longitudinal studies.

## Discussion

In this study we assessed the radiographic progression of hip OA on pelvic plain radiographs taken three years apart in 104 patients with hip OA. Single, paired, and chronologically ordered reading procedures resulted in a difference in the number of patients changing grades on Kellgren and Lawrence or joint space narrowing grading scales. The single reading procedure more often showed change on the Kellgren and Lawrence scale (42% of patients) and on the joint space narrowing scale (37% of patients) than each of the other reading procedures. Measurement of JSW progression on single radiographs without landmarks would require 10% more patients than on single radiographs with landmarks in longitudinal studies.

The data reported in this study emphasise the consequences of using insensitive methods for measuring radiological OA progression in terms of sample size and, therefore, of the cost of studies. For example, in longitudinal studies, measurement of JSW progression on single radiographs with landmarks would necessitate 10% more patients than on paired radiographs with landmarks.

Although several recommendations for assessing OA progression exist,<sup>1-3</sup> the need for standardised methodology remains a challenge.<sup>23</sup> The Kellgren and Lawrence grading scale has been shown to be poorly responsive in knee OA.<sup>24</sup> Clearly, the results of this study suggest that this index is also dependent on the reading procedure in assessing hip OA progression. The change in the Kellgren and Lawrence scale was seen more often with the single reading procedure than with the paired reading procedures. Such findings have been reported elsewhere for other progression scores when reading paired radiographs was compared with reading single radiographs.<sup>25</sup>

Individual features, such as the joint space narrowing scale, have been presumed to be more informative in prospective studies.<sup>26</sup> Altman *et al* rated joint space narrowing on pairs of radiographs as most important in



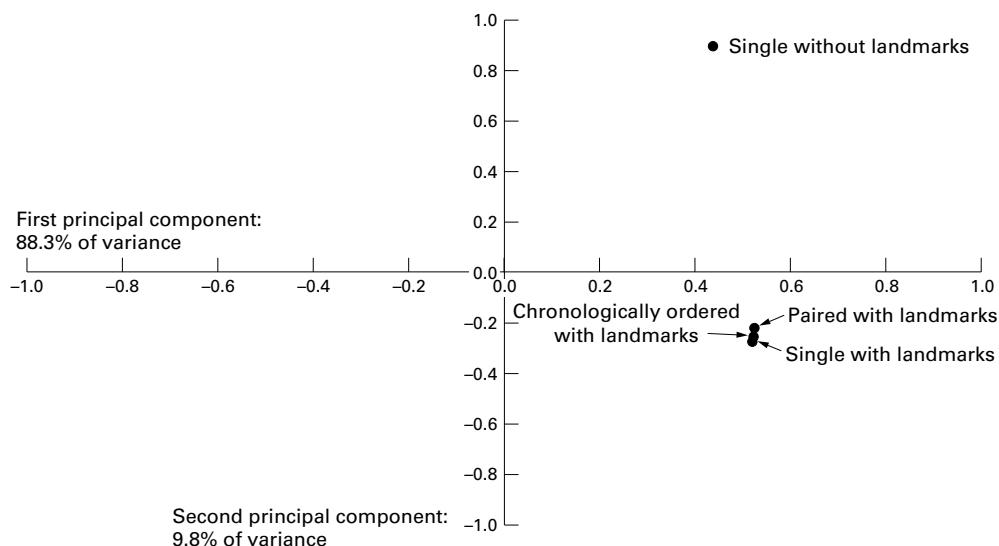


Figure 3 Principal component analysis on change in joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis. The first principal component accounts for 88.3% of the total variability. All the four coefficients are positive and nearly equal, thus allowing the assertion that this component is a measure of global time variation value. The second principal component opposes the single reading procedure without landmarks to the three other reading procedures. This component, which accounts for 9.8% of total variability, then reflects the advantage of using landmarks.

determining progression of hip OA by identifying the correct time sequence of radiographs.<sup>7</sup> We did not find that the results with the joint space narrowing scale were less influenced by reading procedures than results with the Kellgren and Lawrence scale for assessing hip OA progression. Our results actually show that the change in the joint space narrowing scale is also seen more often with the single reading procedure. Differences in methodologies used in these studies may explain these contradictory conclusions. While Altman *et al* used the average score of joint space narrowing rated as a percentage of narrowing by three different readers,<sup>7</sup> we, in contrast, used a single reader who graded joint space narrowing using an atlas.<sup>18</sup>

For the reader there was no advantage in having information on patient identity or even on time sequence for measuring JSW in this study. Perhaps, such information would have been important if there had been a dramatic progression in JSW. However, measuring JSW on paired films would reduce the sample size required compared with the other reading procedures with landmarks. Although the purpose of blindness in clinical studies is, as far as possible, to reduce bias introduced by the investigator's knowledge of a previous measurement for a given subject,<sup>27, 28</sup> the blindness effect seems smaller with a more objective outcome measure, such as JSW measurements, than with outcome measures requiring observers'

judgment, such as Kellgren and Lawrence or joint space narrowing scales.

More interesting is the influence of drawing landmarks when assessing JSW progression. Landmarks have been recommended for measuring JSW in OA clinical trials.<sup>3</sup> According to descriptive statistics, reading procedures with landmarks clearly seemed similar in our study. The presence of landmarks accounted for almost 10% of the variability in JSW progression, and the reading procedures with landmarks clearly contrasted with the reading procedure without landmarks. We considered SRM for comparison of the respective effect on sample size requirements because not only the SD of JSW progression but also the mean of JSW progression differed according to the reading procedure. JSW progression was more homogeneous and at the same time the progression was less with the reading procedure without landmarks than with reading procedures with landmarks. It may be that JSW measurements are not performed in the same sites with these respective reading procedures. Therefore, 10% more patients would be needed in longitudinal studies with a single reading procedure without landmarks than with a single reading procedure with landmarks.

A limitation of this study is that we did not evaluate paired and chronologically ordered reading procedures without landmarks for JSW measurements. We considered that evaluating

Table 3 Intraclass correlation coefficients for change in joint space width measurements using four reading procedures in 104 patients with hip osteoarthritis

Reading procedure	Reading procedure		
	Paired with landmarks	Single with landmarks	Single without landmarks
Chronologically ordered with landmarks	0.963 (0.947 to 0.975)	0.958 (0.938 to 0.971)	0.698 (0.583 to 0.785)
Paired with landmarks		0.963 (0.946 to 0.975)	0.715 (0.606 to 0.798)
Single with landmarks			0.695 (0.580 to 0.782)

these reading procedures without landmarks would have a limited interest and would give results close to those with landmarks. Nor were differences between repeated readings of each procedure by the same reader assessed. For reading procedures using landmarks, these differences may be expected to be almost the same if these reading procedures are compared by pairs.

Although other studies are needed to confirm these findings, the paired reading procedure with landmarks seems to be the most appropriate procedure and should be recommended for measuring JSW progression in hip OA clinical studies.

- 1 Dieppe PA. Recommended methodology for assessing the progression of osteoarthritis of the hip and knee joints. *Osteoarthritis Cartilage* 1995;3:73-7.
- 2 Group for the respect of ethics and excellence in science for the registration of drugs used in the treatment of osteoarthritis. *Ann Rheum Dis* 1996;55:552-7.
- 3 Recommendations from a task force of the Osteoarthritis Research Society: design and conduct of clinical trials in patients with osteoarthritis. *Osteoarthritis Cartilage* 1996;4:217-43.
- 4 Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. *J Rheumatol* 1997;24:786-91.
- 5 Spector TD, Hochberg MC. Methodological problems in the epidemiological study of osteoarthritis. *Ann Rheum Dis* 1994;53:143-6.
- 6 Murphy Jr WA, Altman RD. Updated osteoarthritis reference standard. *J Rheumatol* 1995;22(suppl 43):56-9.
- 7 Altman RD, Fries JF, Bloch DA, Cartens J, Cooke TD, Genant H, *et al.* Radiographic assessment of progression in osteoarthritis. *Arthritis Rheum* 1987;30:1214-25.
- 8 Ravaud P, Giraudeau B, Auleley G-R, Drape J-L, Rousselin B, Paolozzi L, *et al.* Variability of knee radiographing: implication for definition of radiological progression in medial knee osteoarthritis. *Ann Rheum Dis* 1998;57:624-9.
- 9 Dieppe PA, Cushnaghan J, Shepstone L. The Bristol "OA500" study: progression of osteoarthritis (OA) over 3 years and the relationship between clinical and radiographic changes at the knee joint. *Osteoarthritis Cartilage* 1997;5:87-97.
- 10 Speake SC, Szebenyi B, Dieppe P. Quantitative serial assessment of patellofemoral joint osteoarthritis from lateral knee radiographs. *Osteoarthritis Cartilage* 1995;3(suppl A):89-96.
- 11 Dieppe P, Altman RD, Buckwalter JA, Felson DT, Hascall V, Lohmander LS, *et al.* Standardization of methods used to assess the progression of osteoarthritis of the hip or knee joints. In: Kuettner KE, Goldberg VM, eds. *Osteoarthritis disorders*. Rosemont, IL: American Academy of Orthopedic Surgeons, 1994:481-96.
- 12 Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al.* Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. *Ann Rheum Dis* 1996;55:356-62.
- 13 Ledingham J, Dawson S, Preston B, Milligan G, Doherty M. Radiographic progression of hospital referred osteoarthritis of the hip. *Ann Rheum Dis* 1993;52:263-7.
- 14 Ferrara R, Priolo F, Cammisia M, Bacarini L, Cesare A, Pasero G, *et al.* Clinical trials in rheumatoid arthritis: methodological suggestions for assessing radiographs arising from the GRISAR study. *Ann Rheum Dis* 1997;56:608-12.
- 15 van der Heijde D, Boonen A, Boers M, Kostense P, van der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology* 1999;38:1213-20.
- 16 Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;34:505-14.
- 17 Altman RD. The classification of osteoarthritis. *J Rheumatol* 1995;22(suppl 43):42-3.
- 18 Altman RD, Hochberg M, Murphy Jr WA, Wolpke F, Lequesne M. Atlas of individual radiographic features in osteoarthritis. *Osteoarthritis Cartilage* 1995;3(suppl A):3-70.
- 19 Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16:494-501.
- 20 Fleiss JL, Shrout PE. Approximate interval estimation for a certain intraclass correlation coefficient. *Psychometrika* 1978;43:259-62.
- 21 Fortin PR, Stucki G, Katz JN. Measuring relevant change: an emerging challenge in rheumatologic clinical trials. *Arthritis Rheum* 1995;38:1027-30.
- 22 Tuley MR, Mulrow CD, McMahan CA. Estimating and testing an index of responsiveness and the relationship of the index to power. *J Clin Epidemiol* 1991;44:417-21.
- 23 Dieppe P, Brandt KD, Lohmander S, Felson DT. Detecting and measuring disease modification in osteoarthritis. The need for standardized methodology. *J Rheumatol* 1995;22:201-3.
- 24 Ravaud P, Giraudeau B, Auleley G-R, Chastang C, Poiraudou S, Ayral X, *et al.* Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. *J Rheumatol* 1996;23:1756-64.
- 25 Fries JF, Bloch DA, Sharp JT, McShane DJ, Splitz P, Bluhm GB, *et al.* Assessment of radiological progression in rheumatoid arthritis. A randomized, controlled trial. *Arthritis Rheum* 1986;29:1-9.
- 26 Hart DJ, Spector TD. Radiographic criteria for epidemiologic studies of osteoarthritis. *J Rheumatol* 1995;22(suppl 43):46-8.
- 27 Altman DG. *Practical statistics for medical research*. London: Chapman and Hall, 1994.
- 28 Armitage P, Berry G. *Statistical methods in medical research*. 3rd ed. Oxford: Blackwell Scientific, 1994.