Impact of the joint space width measurement method on the design of knee osteoarthritis studies

Olivier Bruyère1,2,3, Yves E. Henrotin3, Aline Honoré1,2, Lucio C. Rovati4, Laurence Seidel1,2, Charles Dardenne1,2, and Jean-Yves L. Reginster1,2,3,5

1WHO Collaborating Center for Public Health Aspects of Osteoarticular Disorders, Liège, 2Department of Public Health and Epidemiology, University of Liège, 3Bone and Cartilage Research Unit, University of Liège, Liège, Belgium, 4Department of Clinical Pharmacology, Rotta Research Laboratorium, Monza, Italy, 5Georgetown University Medical Center, Washington, DC, USA

ABSTRACT. Background and aims: Recent guidelines recommend measurement of articular loss over several years, determined by conventional X-rays, as the principal outcome measure in clinical trials of potential structure-modifying drugs in osteoarthritis (OA). The aim of this study was to assess the impact of the joint space width measurement method on sample size calculation in knee OA studies. Methods: Standard knee X-rays were taken in 212 patients with knee OA at baseline and after 3 years of follow-up. Mean joint space width (JSW) was measured with an in-house computer-assisted method. Minimum JSW, measured with a graduated magnifying lens, was taken as external standard. After calculation of the intra- and inter-observer reproducibility of the JSW, sensitivity to change was assessed using the standardized response mean (SRM). The number of patients needed to identify a mean significant difference of 0.5 mm in joint space narrowing between the placebo and the treated group, after 3 years of follow-up, was then calculated. Results: JSW measured with the computer-assisted technique showed better intra- and inter-observer reproducibility than when using the magnifying lens. JSW values measured with our computer-assisted method were significantly correlated with JSW values obtained using the magnifying lens (r=0.87, p<0.001). The SRM were 0.44 and 0.40 for the computer-assisted method and magnifying lens, respectively. The number of patients needed was 131 per group using the computer-assisted method, and 104 using the magnifying lens. Conclusions: Our method of measurement of JSW may be of potential use in longitudinal studies evaluating the effect of structure-modifying drugs in OA, due to its high level of precision and efficiency. However, although sensitivity to change is markedly better with the digitized method compared with the graduated magnifying lens, we recommend the measurement of mean and minimum JSW in structure-modifying OA trials.

©2003, Editrice Kurtis

INTRODUCTION

Osteoarthritis (OA) is a disease leading to progressive deterioration in the structure and function of synovial joints (1). The main patient-related outcomes are pain and disability, which result in an important community healthcare burden. Standard X-rays are still regarded as the most appropriate tool to identify the anatomic changes in joint structure leading to the diagnosis of OA (2). Recent guidelines from regulatory authorities and expert consensus panels recommend the principal outcome measure in clinical trials of potential structure-modifying drugs in OA to be measurement of articular loss (joint space narrowing) over several years, determined by conventional X-rays (3-5). These documents recommend using digital image analysis to measure both minimum and mean joint space width (JSW) but no single method is regarded as the gold standard (6). Such an “ideal” method should be user-friendly, accurate, reproducible, sensitive to changes, and efficient. For a prospective study evaluating a potential structure-modifying drug, high precision and accuracy would allow sample size and consequently the cost of trials, to be reduced. However, few comparisons of different methods to assess JSW in long-term prospective studies have been performed.

The objective of the present paper is to critically assess a new digital image analysis method measuring the mean

Key words: Digital image analysis, knee, joint space width, OA diagnosis, osteoarthritis.

Correspondence: J-Y. Reginster, M.D., Ph.D., Bone and Cartilage Research Unit, University of Liège, Quai G. Kurth 45, 4000 Liège, Belgium.
E-mail: jyreginster@ulg.ac.be

Received May 31, 2002; accepted in revised form October 2, 2002.
JSW and joint space area (JSA) from knee X-rays of osteoarthritic patients. We therefore measured the analytical properties of the method and evaluated its ability to detect significant changes in JSW and JSA. We used the measurement of the minimum joint space width with a previously validated graduated magnifying lens (7) as external standard.

**METHODS**

**Patients**

The study population, more extensively described elsewhere (8), was composed of 212 individuals, of both sexes. They came from a double-blind, placebo-controlled study evaluating the symptomatic and structural effects of a structure-modifying drug in knee OA, over a period of 3 years. Of these 212 patients, 139 completed the 3-year study (71/106 in the placebo group, 68/106 in the treated group). Knee OA was diagnosed according to the clinical and radiological criteria of the American College of Rheumatology (9). Baseline demographics of the population are listed in Table 1.

**Procedures**

X-ray acquisition. Standard radiographs were taken for each knee at baseline and after 3 years. Patients stood with their knees fully extended, with the posterior aspect of the knees in contact with a vertical cassette in a cassette holder. The lower limbs were internally rotated until the patella was centralized over the lower end of the femur. The feet were positioned a small distance apart. Foot maps were used to reposition patients at the time of subsequent X-rays. The horizontal X-ray beam was centered on the joint space and parallel to the tibial plateau. Fluoroscopy was used to correct lower limb positioning and X-ray beam alignment. The focus-film distance was 110 centimeters. There was one knee for each radiograph.

**Joint space measurement**

Mean joint space width. Radiographs were digitized and image analysis was performed according to a new in-house technique. Radiographs were placed on a conventional horizontal light box and viewed with a high-definition camera. The image was captured in an Olivetti PC-M300 microcomputer, giving a spatial resolution of 720x560 pixels with 256 levels of grey. A millimeter-squared grid was used to assess the joint space width. This grid was superimposed on the radiographs, and placed with its horizontal axis tangential to the tibial plateau and its vertical axis centered on the middle of the edge of the external tibial spine. Two different grids were compared to select the area for calculation of the mean joint space. The large grid (Fig. 1) used a box of 2 cm x 1.5 cm, placed on the horizontal axis, 2.5 cm externally from the vertical axis. The small grid (1.5 cm x 1.5 cm) was placed 3 cm externally from the vertical axis. The computer mouse was used to identify the proximal and distal joint limits within the grid and the outline of the joint. The proximal and distal limits were not marked before measurement. Then the computer automatically calculated first the joint space area (JSA) and then the mean joint space width (JSW) of the medial compartment of the tibio-femoral joint. That is, the JSW corresponds to the ratio between the JSA and the length of the inferior proximal and distal limit.

**Minimum joint space width.** The minimum joint space width, e.g., the narrowest point of the medial compartment of the tibio-femoral joint, was assessed by visual determin-
nation using a 0.1 mm graduated magnifying lens (Peak Light Scale Lupe 7X), as previously described (7).

All baseline X-rays (N=212) were analyzed following Altman et al. (10). This method allowed to each X-ray a global score corresponding to the sum of 13 values, grading the severity of OA from 0 to 3 for 13 radiographic features. To test the inter- and intra-observer reproducibility of our computer-assisted JSW measurement with the different grids, we used ten X-rays graded 0 for the medial compartment joint space narrowing item of the Altman et al. atlas (10), ten graded 1 or 2, and ten graded 3. Two independent assessors calculated the JSW and JSA with the LG and SG for each of the 30 radiographs. These measures were repeated, blindly, of the X-ray sequence, after one week. We selected the grid with the best inter- and intra-observer reproducibility, to measure the mean JSW of all X-rays taken at baseline and at the end of the study. For all these radiographs, we also measured the minimum JSW with a graduated magnifying lens (7).

Statistical analysis
The intra- and inter-observer reproducibility of the JSW, measured both with digitized image analysis (2 grids) and the graduated magnifying lens, was assessed by the coefficient of variation (CV), calculated using the method described in Jonsson et al. (11). To determine the CV for each modality, means were in fact calculated across patients for each individual measurement. Individual patient variances were then calculated by applying the measurement values of the first and second examinations. A global within-patient standard deviation (SD) was calculated from the individual patient variances. The CV was then calculated for each joint measurement by dividing the pooled within-patient SD by the cross-patient mean. An overall CV was then calculated for each modality by averaging the CV for each individual joint measurement. Similarly, the CV were calculated to assess the reproducibility of the JSA.

The strength of the association between JSW and JSA measured with the two grids was quantified using the Pearson correlation coefficient (PCC). The same coefficient was also used to quantify the association between JSW measured by digitized image analysis and by the graduated magnifying lens.

Receiver Operating Characteristic (ROC) analysis (12) was used to assess the sensitivity and specificity of the computer-assisted method, in order to identify a joint space narrowing of 0.5 mm, assessed with the magnifying lens method.

To evaluate the sensitivity of the two methods (in- house digital image analysis and magnifying lens) to detect early changes over 3 years, we calculated the standardized response mean (SRM) in the placebo group (mean change divided by the SD of the change) (13, 14). A SRM >0.8 is considered to reflect a high potential to detect changes. SRM of 0.5 and 0.2 are linked to moderate and low potentials to detect changes, respectively (15). We also assessed the overall efficiency of each method by evaluating the number of patients needed, in order to identify a mean difference of 0.5 mm in joint space narrowing between the placebo and treated groups.

As proposed by Pocock (16): \( N = \frac{(2\delta^2)}{(\mu_1 - \mu_2)^2} \times f(\alpha, \beta) \) where \( \delta \) is the standard deviation in the population at baseline (1.44 for mean JSW, 1.28 for minimal JSW), \( (\mu_1 - \mu_2) \) is the change in the mean response (0.5 mm); and \( f(\alpha, \beta) \) is a function of \( \alpha \) and \( \beta \) errors, with \( \alpha=0.05 \) and \( \beta=0.2 \). The arbitrary cut-off of a decrease of at least 0.5 mm in JSW after 3 years was based on the results of Lequesne et al. (4) in which a difference of 0.5 mm in joint space narrowing between an active drug and a comparator was suggested for the calculation of a sample size in a disease-modifying OA study.

RESULTS
As shown in Table 2, the JSW measured with our in- house computer-assisted method (large grid) consistently showed better intra-observer reproducibility than the magnifying lens. The small grid consistently reduced the reproducibility of the technique, and this was also the case for JSA measurement. X-rays with a low Altman grade (lower severity of OA) were analyzed with better reproducibility than those with a high Altman grade. Similar results were observed for inter-observer reproducibility (Table 3). The JSW values measured with our computer- assisted method were significantly correlated both with the JSA values (r=0.987, p<0.0001 for the small grid; r=0.999, p<0.0001 for the large grid) and with JSW values obtained from analyses using the magnifying lens (r=0.89, p<0.0001 for the small grid; r=0.87, p<0.001

Table 2 - Intra-observer coefficient of variation (CV) for JSW, expressed in % (SD).

<table>
<thead>
<tr>
<th></th>
<th>Grade 0 (N=10)</th>
<th>Grade I or II (N=10)</th>
<th>Grade III (N=10)</th>
<th>All patients (N=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small grid</td>
<td>4.63 (3.27)</td>
<td>3.72 (1.14)</td>
<td>7.94 (7.54)</td>
<td>4.96 (3.57)</td>
</tr>
<tr>
<td>Large grid</td>
<td>4.34 (6.02)</td>
<td>2.24 (1.05)</td>
<td>4.19 (3.65)</td>
<td>3.51 (4.50)</td>
</tr>
<tr>
<td>Magnifying lens</td>
<td>2.63 (1.93)</td>
<td>5.75 (9.34)</td>
<td>9.30 (11.03)</td>
<td>5.26 (7.49)</td>
</tr>
</tbody>
</table>

138 Aging Clin Exp Res, Vol. 15, No. 2
Table 3 - Inter-observer coefficient of variation (CV) for JSW, expressed in % (SD).

<table>
<thead>
<tr>
<th>Grid Type</th>
<th>Grade 0</th>
<th>Grade I or II</th>
<th>Grade III</th>
<th>All patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N=10)</td>
<td>(N=10)</td>
<td>(N=10)</td>
<td>(N=30)</td>
</tr>
<tr>
<td>Small grid</td>
<td>6.24 (8.8)</td>
<td>2.69 (2.7)</td>
<td>8.19 (4.79)</td>
<td>5.26 (6.93)</td>
</tr>
<tr>
<td>Large grid</td>
<td>3.91 (6.20)</td>
<td>1.62 (0.85)</td>
<td>4.77 (4.24)</td>
<td>3.20 (4.69)*</td>
</tr>
<tr>
<td>Magnifying lens</td>
<td>7.99 (8.59)</td>
<td>6.45 (12.19)</td>
<td>10.19 (8.72)</td>
<td>7.85 (10.36)</td>
</tr>
</tbody>
</table>

*p<0.05 compared with magnifying lens.

for the large grid). Since the large grid showed better intra- and inter-observer reproducibility, we used it to measure the mean JSW on the whole set of X-rays, at baseline and at the end of the study.

The sensitivity and specificity of our computer-assisted method for JSW measurement, when aiming at detecting a relevant value (0.5 mm) of joint space narrowing, was 90% and 47% respectively. The area under the ROC curve, evaluating the discriminatory value of our technique, was 0.85 (Fig. 2).

Over the 3-year period, the mean decrease observed in JSW for the computer-assisted technique and measuring the grid were [mean (SD)] 0.45 mm (1.02 mm) and 0.40 mm (1.00 mm) respectively. The corresponding values of SRM were 0.44 and 0.40. Based on the values of mean JSW [5.82 mm (1.44)] and minimal JSW [3.89 mm (1.28 mm)] measured at baseline and the formula proposed by Pocock, the number of patients needed to identify a relevant difference (0.5 mm) between the placebo and the treated groups for joint space narrowing, over three years, was 131 per group when using our in-house computer-assisted image analysis system, and 104 per group when using the graduated magnifying lens. In the initial study, the number of complete patients with this cut-off of worsening after 3 years was 27/71 (38.0%) and 30/71 (42.3%) with the computer-assisted method and the graduated magnifying lens, respectively.

DISCUSSION

The measurement of JSW on standard X-rays is currently considered to be the most appropriate outcome measure to detect structural progression of knee OA. In daily practice and clinical research, JSW is assessed through different methods, e.g., rulers with mm divisions, calipers, graduated magnifying lenses and image digitalization (17-19). Magnetic resonance imaging, while potentially more accurate and precise than conventional X-rays, is still considered for investigational use only by regulatory authorities and, furthermore, currently suffers from a clear lack of accessibility (3). Other imaging techniques, such as computer tomography and ultrasonography, have not yet been fully validated for this specific purpose (3). Recent regulatory guidelines actually recommend using a digitalized image analysis system for the measurement of both minimal and mean JSW (5, 6, 29). However, none of them identifies a single method as the "gold standard" compared with the others. Such an ideal method should be user-friendly, highly reproducible, and sensitive to changes. With the aim of contributing to the identification of such a method, we developed an in-house technique for computer-assisted measurement of JSW, which relies on careful identification of the region of interest for mean JSW measurements, defined by precise and accurate limits. We chose the tibial plateau for the horizontal axis and the edge of the external tibial spine for the vertical axis of our region of interest, due to the well-defined and consistent visibility of these anatomical landmarks on conventional X-rays. In the process of identification of our region of interest, i.e., where JSW could be measured with the highest reproducibility, we compared grids of two different sizes and selected the larger one, which allowed us to reach the lowest CV when repeated measurement of JSW were performed, both by the same investigator and by different observers. Our large grid is more likely to give a full picture of joint space narrowing within a large portion of the joint. The intra-observer CV for JSW measurement with our computer-assisted method using the large grid (3.51%) was in the

Figure 2 - Sensitivity and specificity of graduated magnifying lens in detecting a relevant joint space narrowing (0.5 mm) in mean JSW, assessed by computer.
same range as the CV obtained in our previous paper (8) (1.82%), using a previously validated digitized image analysis system (21) and the same method for assessment of reproducibility.

As expected, the best coefficient of variation was observed in X-rays from patients exhibiting the smallest narrowing joint space. Although the calculation of JSW is guided by the computer, determination of the edge of the bone is made manually, which may be more difficult and less accurate in cases of intense bone remodeling and in the presence of osteophytes, as seen in the most severe cases of OA.

We found better reproducibility with our computer-assisted method of measurement than with the magnifying lens. This cannot be linked to any fault in our use of the magnifying lens, since our intra-observer reproducibility of 5.16% is within the same range as previous reports from others (3.8%) using a similar device (22). As for the computer-assisted analyses, the better intra- and inter-observer reproducibility for the lens was observed for low-grade OA X-rays. In another study (19) comparing four measuring instruments for assessment of JSW in knee OA (ruler, caliper, graduated magnifying glass, digitized assessment), it was reported that, although intra- and inter-observer reproducibility was high with all measuring instruments, the reader's component accounted for 0% of total variance for the ruler, 2% for the digitized method, 16% for the caliper, and 18% for the graduated magnifying glass. With other especially prepared computer program, Buckland-Wright et al. (23) reported a CV of 6.4% for variability of repeated manual measurements of the medial compartment of the JSW from conventional extended-view radiographs. Computerized measurement alone did not reduce the error variation in JSW measurement (CV=6.2%), probably because of lack of standardization of the knee position. However, when the automated, magnification-corrected measurement system was applied to JSW measurement on radiographs of normal knees in the semiflexed position, the CV reached 3.2%.

We assessed the efficiency of the two measurement techniques by calculating the required number of patients to be included in a clinical trial, to show what is usually considered as a relevant difference (0.5 mm) in joint space narrowing over a 3-year period. While the graduated magnifying lens technique (minimum JSW) required a sample of at least 104 patients per group to show such a difference, 131 patients in each group would be needed with the computer-assisted method (mean JSW). However, it must be borne in mind that the minimum JSW may be more sensitive in detecting changes in cartilage than the mean JSW (23). In the main paper from which this placebo-treated population was extracted (8), we used another computer-assisted method to measure JSW and joint space narrowing over time, and found a mean JSW at baseline in the whole group of 5.32 mm (SD 1.33 mm). With this rather similar value and applying the same calculation as we did to determine the appropriate sample size in the present study, the number of patients needed to find a relevant difference of 0.5 mm between placebo and treated groups, after 3 years, would have been 112 in each group, based on the same power calculations. Based on this comparison, our new in-house method of JSW assessment appears to be at least as efficient as that previously published and used in our former study.

Another way of reducing the size of the sample to be included in a clinical trial is, obviously, to optimize the positioning of the patient (5) or the X-ray acquisition procedure, e.g., by using semi-flexed (24) or Schuss views (25), fluoroscopy (22), macroradiographs (22, 26, 27) or magnification correction (22). These procedures, together with an accurate method of measurement of JSW, do promise the best compromises in evaluating the progression of OA lesions. As a possible limitation of our study, we should acknowledge that the weight-bearing radiographic view that we used may not have been the most accurate to assess JSW or sensitive enough in detecting changes in joint space narrowing. So, although the CV could probably be lower, the inter- and intra-observer differences between JSW assessed by the graduated magnifying lens and by computer-assisted technique would still exist. The same is true for longitudinal follow-up, in which SRM could probably be higher. Nevertheless, in prospective studies, a frequent criticism of weight-bearing radiographic view are changes in patient positioning due to changes in symptoms or to OA in other sites of the lower limb. However, we believe that it is unlikely that symptom changes observed in one group affected the results, given the mild to moderate disease and symptom conditions at baseline and throughout the initial study from which these results came (8). Similarly, the relationship between symptom and structure modification is of poor magnitude and clinical relevance.

CONCLUSIONS

We have developed an in-house method for computer-assisted assessment of JSW of the OA knee joint. Our method appears to be highly reproducible and also offers an opportunity to assess JSA. Our method of measurement of JSW may be of potential use in longitudinal studies evaluating the effect of structure-modifying drugs in OA, due to its high level of precision and efficiency. Assessment of minimum JSW by a graduated magnifying lens, a widespread and cheap technique of assessing JSW, provides reproducibility and precision within the same order of magnitude as that observed when using a computer-assisted method of measurement of mean JSW. However, the sensitivity to change is markedly better for the digitized method, and it is therefore recommended for longitudinal
Joint space width measurement in knee osteoarthritis

trials. Moreover, mean JSW assessment is less sensitive to radiographic procedures or patient positioning errors, and may be assessed when the interbone distance is very small (25). However, in accordance with previous guidelines (5) and on the basis of the present results, we recommend the measurement of mean and minimum JSW in structure-modifying OA trials.

REFERENCES