Relevant change in radiological progression in patients with hip osteoarthritis. I. Determination using predictive validity for total hip arthroplasty

J. F. Maillefert1,4, A. Gueguen2, M. Nguyen1, L. Berdah3, M. Lequesne5, B. Mazières6, E. Vignon7 and M. Dougados1

1Rene´ Descartes University, Cochin Hospital, Institut de Rhumatologie, Paris, 2INSERM U88, St Maurice Hospital, Paris, 3NEGMA Laboratories, Toussus le Noble, 4General Hospital, Dijon, 5Leopold Bellan Hospital, Paris, 6Rangueil Hospital, Toulouse and 7Lyon-Sud Hospital, Pierre Bénite, France

Abstract

Objective. To determine a cut-off point above which a change in joint space width (JSW) could be considered as relevant in patients with hip osteoarthritis (OA) on the basis of predicted need for subsequent total hip arthroplasty (THA).

Methods. A multicentre, prospective, longitudinal, 5-yr follow-up study was performed. A pelvic radiograph was obtained at entry and after 1 and 2 yr. For each film, the narrowest JSW was measured using a 0.1 mm graduated magnifying glass. The absolute and relative differences between baseline and 1 and 2 yr of follow-up were calculated. We determined the cut-off points above which an absolute or relative decrease in JSW between baseline and 1 and 2 yr of follow-up could be considered relevant on the basis of the predicted need for THA during the remaining years of the study. The need for THA was categorized as ‘yes’ or ‘no’. Thereafter, for each observed change in JSW (0.1 per 0.1 mm or 1% per 1%), the sensitivity and specificity for subsequent THA were calculated. The choice of cut-off was based on maximal sensitivity and specificity, using the graphic representation of correct classification probabilities. In this way it was possible to obtain the best measured JSW threshold with maximal true positive and minimal false positive results.

Results. A total of 423 and 385 patients met the criteria for analysis using the decrease in JSW between baseline and 1 and 2 yr respectively. The best cut-off points were absolute decreases in JSW of 0.2 and 0.4 mm and relative decreases in JSW of 15 and 20% after 1 and 2 yr respectively, with corresponding ranges of sensitivity and specificity of 68–75 and 67–78%.

Conclusion. This work determined the cut-off above which a change in JSW could be considered clinically relevant in patients with hip OA, on the basis of predicted subsequent need for THA. For validation, similar studies should be conducted in other countries with different health-care systems.

Key words: Relevant change, Predictive validity, Total hip arthroplasty, Hip osteoarthritis, Joint space width, Outcome measure.
However, this evaluation results in a continuous variable and does not discriminate between patients with and those without relevant change. The percentage of patients having radiological disease progression is an important end-point in trials and is easier to understand than mean changes. In addition, having an established definition of progression and non-progression allows comparison of data from different trials. Because JSW is evaluated routinely in daily practice, such a cut-off point may also facilitate the monitoring of OA patients. Thus, the determination of a cut-off point above which a change in JSW could be considered relevant in hip OA should be useful.

Several approaches can be used to determine a cut-off above which a change in imaging variables could be considered relevant [5]. Distribution-based models describe features of the distribution of the measure in the population. In particular, the smallest detectable difference (SDD) is based on measurement error and thus on reproducibility [6]. In other words, a cut-off point determined by the use of SDD would help ensure that the changes observed were true changes rather than random fluctuation related to variability in the measurement procedure. Authority-based models are based on expert opinion, i.e. on the intuitive clinician’s global assessment on the basis of gained experience and knowledge. Predictive models are driven experimentally and are used in evidence-based approaches.

In a previous study, we determined the SDD on the basis of intraobserver progression score measurement error in the absolute progression of JSW in hip OA [7]. In the present study, we aimed to determine a cut-off above which a change in JSW could be considered relevant in patients with hip OA, based on prediction of further total hip arthroplasty.

Methods

Study design
A multicentre, prospective, longitudinal, 5-yr follow-up study was designed. The study protocol was approved by the ethics committee of the Hôpital Cochin, Paris, France.

Inclusion criteria
These have been described previously [8]. Briefly, outpatients visiting a rheumatologist and fulfilling the American College of Rheumatology criteria for the diagnosis of hip OA [9] were enrolled in the study after written informed consent had been obtained. Other inclusion criteria were age between 50 and 75 yr, daily hip pain for at least 1 month during the last 3 months. Exclusion criteria were joint space width <1 mm at the narrowest point, radiographic medial or axial femoral head migration, secondary hip OA defined by the presence of a history of hip fracture, inflammatory rheumatic disease, osteonecrosis of the femoral head and Paget’s disease.

Structural evaluation
An anteroposterior weight-bearing radiograph of the pelvis with the lower limbs in 15° internal rotation was obtained at entry and after 1 and 2 yr. All films were collected and analysed by one of us (ML) [10, 11]. Films were blinded with respect to patient identity and date. A randomization list was used for blinding with respect to chronology. The films of a single patient were placed side by side on a light-box and the narrowest JSW (identical for all films) was selected and measured (interbone distance) using a 0.1 mm graduated magnifying glass. The evaluation of the intra-observer reliability using 30 pairs of films, assessed twice after a 1-month interval, showed an intra-class coefficient of correlation of 0.840 (95% confidence interval 0.693–0.920).

Decision for surgery
This was made by the patient, the rheumatologist and the surgeon with no reference to radiographic change (although the rheumatologist and the surgeon had the opportunity to evaluate the last follow-up X-rays) and/or analysis of objective symptomatic variables.

Analysis
Absolute and relative changes in JSW between baseline and 1 and 2 yr of follow-up evaluations were calculated. Analyses were performed in order to propose relevant thresholds, allowing us to transform the continuous variable (absolute and relative changes in JSW) into a dichotomous variable: progression or no progression.

The patients were classified as requiring total hip arthroplasty (THA) or not requiring THA. Thereafter, different possible thresholds were evaluated: for each change observed in measured JSW (0.1 per 0.1 mm or 1% per 1%), we measured the sensitivity (percentage of patients with a decrease in measured JSW above the threshold, among the patients who underwent THA during the follow-up) and the specificity (percentage of patients with a decrease in measured JSW below the threshold, among the patients who did not have THA during the follow-up) for further THA. The choice of cut-off was based on maximal sensitivity and specificity, using the graphic representation of correct classification probabilities. In this way we were able to obtain the best measured JSW threshold with maximal true positive and the minimal false positive results [12].

Two analyses were performed, producing four different thresholds. The first analysis aimed to determine a cut-off point above which an absolute or relative decrease in JSW between baseline and the 1-yr follow-up could be considered relevant on the basis of the prediction of THA during the following 4 yr. The second analysis aimed to determine a cut-off point above which an absolute or relative decrease in JSW between baseline and the 2-yr follow-up could be considered relevant on the basis of prediction of THA during the following 3 yr. In some instances, the interval between baseline and follow-up radiographs was not 12 or
24 months. Only the pairs of films separated by intervals of 12 ± 3 and 24 ± 6 months were used for analysis.

For each of the four thresholds, Kaplan–Meier estimates of the cumulative probabilities of not having THA (according to the decrease in JSW above or below the threshold) were obtained.

Results

Five hundred and eight patients were recruited initially. In the first analysis, 39 patients (7.7%) underwent THA during the follow-up period for further THA. The best measured JSW thresholds, with the maximal (0.1 mm per 0.1 mm or 1% per 1%), the sensitivity (percentage of patients with a decrease in measured JSW above the threshold, among the patients without THA during the follow-up period) for further THA. The best measured JSW thresholds, with the maximal true positive and minimal false positive results, were selected.

In the first analysis, 39 patients (7.7%) underwent THA during the first year. Among the remaining 469 patients, 46 were excluded because they had an interval between baseline and follow-up that differed from 12 ± 3 months. Among the 423 remaining patients, 422, whose main baseline characteristics are shown in Table 1, were observed over the following 4 yr. In these 422 patients, the mean JSW was 2.31 ± 0.81 (S.D.) mm at baseline and 1.99 ± 1.02 mm after 1 yr of follow-up (mean absolute and relative changes −0.33 ± 0.57 mm and −17 ± 29% respectively). During the remaining 4 yr of the study, THA was performed in 149 of the 422 patients (35.3%). Thresholds of −0.2 mm and −15%, i.e. a decrease in JSW of at least 0.2 mm (15%), were the most relevant, with sensitivity of 75 and 68% (−0.2 mm) and 74 and 78% (−15%) respectively (Table 2).

In the second analysis, 81 patients (15.9%) underwent THA during the first 2 yr. Among the 427 remaining patients, 42 were excluded as having an interval between baseline and follow-up X-rays different from 24 ± 6 months. Among the 385 other patients, 384, whose main baseline characteristics are shown in Table 1, were observed during the following 3 yr. In these 384 patients, the mean JSW was 2.35 ± 0.81 (S.D.) mm at baseline and 1.91 ± 1.07 mm after 2 yr of follow-up (mean absolute and relative changes −0.44 ± 0.61 mm and −22.7 ± 32% respectively). During the remaining 3 yr of the study, THA was performed in 111 of the 384 patients (28.9%). Thresholds of −0.4 mm and −20%, i.e. a decrease in JSW of at least 0.4 mm and 20%, were the most relevant, with sensitivity and specificity of 68 and 67% (−0.4 mm) and 70 and 68% (−20%) respectively (Table 2).

Figure 1 shows one of the graphic representations of correct classification probabilities and Fig. 2 shows one of the Kaplan–Meier estimates of the cumulative probabilities of not having THA, according to the decrease in JSW below or above a determined threshold.

**Table 1. Main baseline characteristics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Demographic data</th>
<th>Sex (male, female; no.)</th>
<th>Body mass index (kg/m²; mean ± S.D.)</th>
<th>Pain (mm; mean ± S.D.)</th>
<th>Lequesne’s index (mean ± S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>First analysis (422 patients)</td>
<td>176, 246</td>
<td>25.8 ± 3.5</td>
<td>43.8 ± 20</td>
<td>7.6 ± 2.5</td>
<td></td>
</tr>
<tr>
<td>Second analysis (384 patients)</td>
<td>216</td>
<td>25.7 ± 3.4</td>
<td>43 ± 19.3</td>
<td>7.4 ± 2.4</td>
<td></td>
</tr>
</tbody>
</table>

The table shows the main baseline characteristics of the 422 patients included in the first analysis (determination of the cut-off point above which an absolute or relative decrease in JSW between baseline and 1 yr could be considered relevant on the basis of predicted THA during the subsequent 4 yr) and the 384 patients included in the second analysis (determination of a cut-off point above which an absolute or relative decrease in JSW between baseline and 2 yr of follow-up could be considered relevant on the basis of predicted THA during the subsequent 3 yr).

*Measured on a 100-mm visual analogue scale.

*A scale from 0 to 24 reflecting the effect of OA on daily activities.

**Table 2. Cut-off points above which an absolute or relative decrease in JSW between baseline and 1 and 2 yr of follow-up could be considered relevant on the basis of prediction of THA during the subsequent 4 and 3 yr respectively**

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity for subsequent THA</th>
<th>Specificity for subsequent THA</th>
<th>Positive predictive value</th>
<th>Negative predictive value</th>
<th>No. of patients with a decrease in JSW below the threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute decrease in JSW between baseline and 1 yr</td>
<td>0.2 mm</td>
<td>75%</td>
<td>68%</td>
<td>66.9%</td>
<td>79.2%</td>
</tr>
<tr>
<td>Relative decrease in JSW between baseline and 1 yr</td>
<td>15%</td>
<td>74%</td>
<td>78%</td>
<td>68.5%</td>
<td>85.4%</td>
</tr>
<tr>
<td>Absolute decrease in JSW between baseline and 2 yr</td>
<td>0.4 mm</td>
<td>68%</td>
<td>67%</td>
<td>50%</td>
<td>80.5%</td>
</tr>
<tr>
<td>Relative decrease in JSW between baseline and 2 yr</td>
<td>20%</td>
<td>70%</td>
<td>68%</td>
<td>47.9%</td>
<td>85.1%</td>
</tr>
</tbody>
</table>

The thresholds were obtained from the graphic representation of correct classification probabilities, showing for all possible changes in JSW (0.1 mm per 0.1 mm or 1% per 1%) the sensitivity (percentage of patients with a decrease in measured JSW above the threshold, among the patients who underwent THA during the follow-up period) and specificity (percentage of patients with a decrease in measured JSW below the threshold, among the patients who did not have THA during the follow-up period) for further THA. The best measured JSW thresholds, with the maximal true positive and minimal false positive results, were selected.
Discussion

This study determined the cut-off point above which a change in JSW could be considered relevant on the basis of prediction of THA. To our knowledge, thresholds in JSW progression in patients with hip OA, determined using this method, have not been available previously. This cut-off point could be used to differentiate between patients with and without a structurally relevant change in therapeutic trials and clinical practice.

Because of the characteristics of the population, it might be difficult to extrapolate the results of this study to a general population of patients with hip OA. This study focused on a particular subgroup of patients, all of whom had, for example, painful, active disease. However, such characteristics are those commonly observed both in daily practice and in therapeutic trials evaluating the effects of particular drugs in hip OA. The assessment of JSW seems to be related to the radiological procedure used and, in particular, to patient positioning. For example, weight-bearing may affect JSW in hip OA [13, 14]. In this study, the radiological procedure was standardized: all X-rays were taken with the patient standing with the lower limbs in 15° internal rotation. The threshold might have been different if another standardized procedure had been used. It must also be pointed out that individuals with baseline values below a threshold determined using the absolute decrease in JSW can never be regarded as having deteriorated [15]. Thus, the thresholds determined using the absolute decrease in JSW should be preferred in trials examining structurally severe hip OA or, in daily practice, in monitoring patients with such severe disease.

In this study, several thresholds were determined. Obtaining thresholds for the prediction of THA using several intervals between pairs of films, i.e. 1 and 2 yr, makes it possible to use them in trials of different duration or to perform intermediate analyses in 2-yr
follow-up trials. The accuracy of the proposed thresholds appeared to be good. The cut-off points determined with the relative decrease tended to have greater accuracy than those obtained with the absolute decrease in JSW. However, this method does not take account of the severity of structural deterioration at baseline. Moreover, in patients with severe baseline structural deterioration, a decrease of 15 or 20% in JSW corresponds to absolute decreases that are far less than measurement error. Thus, we recommend the use of one of the cut-off points determined using the absolute decrease in JSW rather than the relative decrease. The cut-off point determined using the absolute decrease in JSW was lower than the SDD (0.6 mm) [7]. Increasing the threshold to 0.6 mm would improve specificity but reduce sensitivity for the prediction of subsequent THA. Despite poorer accuracy, the threshold determined using the absolute decrease in JSW between baseline and the 2-yr follow-up (0.4 mm) is preferable to that obtained by the use of the absolute decrease in JSW between baseline and the 1-yr follow-up (0.2 mm).

A model based on the prediction of subsequent THA was used. This model has several advantages: it is probably clinically relevant, because THA is performed only in patients with severe disease; the variable ‘THA or no THA’ is easy to ascertain; and changes in JSW have been shown to be highly predictive of the subsequent need for THA [8]. However, the choice of subsequent THA as the gold standard has some disadvantages. First, this choice is subjective: the goals of treatment in OA include not only the prevention of THA but also the relief of pain and the attainment and maintenance of optimal joint function, thus reducing disability and handicap and maintaining the risk of progression [16]. Thus, the need for THA has been proposed as a ‘hard’ outcome measure in trials, but not as a replacement for other outcomes [8]. Secondly, the indications for THA might vary among centres, surgeons and patients, and might be affected by factors that are not closely related to the clinical presentation of the disease [17, 18]. Thirdly, despite the fact that the decision regarding surgery was made with no reference to radiographic change, it was potentially made with reference to the last assessment of radiographic damage, thus might have introduced some bias. Thus, the thresholds might have been different if they had been determined with other sets of patients in other countries. It is not possible to proceed differently in the absence of validated sets of criteria for considering THA. Such criteria have been proposed recently [19], but as validation by others is still lacking they were not used in the present study. Because the decision with regard to THA in the present study took into account the opinions of the patient and at least two specialized practitioners (the rheumatologist and the surgeon) and, because the study used a multicentre design, some disparities in surgical decisions should have been taken into account. Economic considerations should not have affected the results as the great majority of the French population is insured by the National Health Service. However, for validation, similar studies should be conducted in other countries with different health-care systems.

This work determined a cut-off point above which a change in JSW could be considered as clinically relevant in individuals with hip OA on the basis of prediction of subsequent THA. This cut-off point could be used to differentiate patients with and without a structurally relevant change, both in therapeutic trials and in clinical practice. However, for validation, similar studies should be conducted in other countries with different health-care systems.

Acknowledgement

This work was supported in part by NEGMA Laboratories.

References

12. Bruynesteyn K, van der Heijde D, Boers M et al. Minimum clinically important difference in radiological...


