Measurement of structural progression in osteoarthritis of the hip: the Barcelona consensus group

R. D. Altman M.D., Professor of Medicine†*, D. A. Bloch Ph.D., Professor of Biostatistics‡, M. Dougados M.D.§, M. Hochberg M.D., Professor of Medicine and Chief of Rheumatology||, S. Lohmander M.D., Ph.D., Professor of Orthopaedics¶, K. Pavelka M.D., Ph.D., Professor of Medicine, Head of Institute and Clinic of Rheumatology#, T. Spector M.D., M.Sc., FRCP†† and E. Vignon M.D.‡‡

† Division of Rheumatology, David Geffen School of Medicine at the University of California at Los Angeles, Los Angeles, CA, USA
‡ Stanford University, Palo Alto, CA, USA
§ Rheumatology B Department, Rene Descartes University, Cochin Hospital, Paris, France
|| University of Maryland, Maryland, USA
¶ Department of Orthopaedics, Lund University Hospital, Lund University, Lund, Sweden
# Charles University, Prague, Czech Republic
†† St. Thomas’ Hospital, London, UK
‡‡ Service de Rhumatologie, Centre Hospitalier Lyon Sud, Lyon, France

Summary

Objective: To outline the best available method of measurement for detecting progression of osteoarthritis (OA) of the hip especially in therapeutic trials.

Method: A Medline search of articles related to progression of hip OA was performed. A group of experts met over a 1.5-day session to review available literature and new research. Specific questions were addressed in order to reach a consensus on measuring progression of OA of the hip.

Results: Of the available surrogate measures, a single yearly standing or reclined antero-posterior plain radiograph of the pelvis with feet internally rotated 15–20°, can be evaluated with the use of an atlas for joint space width (JSW, interbone distance). There should be a minimum JSW upon baseline screening that may be 1 or 2 mm. Digitization of films offers a slight reduction in variability of measurements. Progression of OA can be calculated by measurement of the JSW on paired and blinded films. A reduction of ≥ 0.5 mm is greater than the ‘minimum perceptible difference’ as well as the variation of most imaging techniques, and represents a clinically relevant and significant reduction in the JSW. Narrowing of the superomedial or superolateral JSW may tend to progress more rapidly than other changes. In clinical trials, patients who discontinue the study treatment need to be followed after discontinuation, and an imputation strategy which provides unbiased estimates of both the treatment effect and its variance is an appropriate technique for intent-to-treat analysis.

Conclusion: For the development of new agents intended to prevent, retard, stabilize or reverse the progress of OA of the hip, the radiographic methodology presently available is adequate to detect changes in hip JSW of OA.

© 2004 Published by Elsevier Ltd on behalf of OsteoArthritis Research Society International.

Key words: Osteoarthritis, Radiography, X-ray, Hip, Structure modification, Disease modification, Joint space width.

Introduction

Osteoarthritis (OA) of the hip is a common and potentially disabling condition. Indeed, in the surgical prevention of painful disability it is estimated that annually between 50 and 140 per 100 000 inhabitants of the developed countries receive a total hip arthroplasty for OA. The number of procedures not only reflects the success of the surgical procedure, but also indicates the failure of non-surgical regimens to adequately treat and prevent progression of symptomatic OA.

There is an ongoing effort to change the therapy of OA from symptomatic to structure (disease) modification. Several agents are being evaluated for their potential to alter the course of OA. Guidelines for conduct of clinical trials for structure modification of the hip and knee have been proposed.

More recently, a consensus group summarized the measurements that would be appropriate for the knee. Further guidance for knee OA is expected from the
proceedings of a workshop on imaging held in Bethesda that was cosponsored by the National Institutes of Health, OMERACT (Outcome Measures in Arthritis Clinical Trials), and OARSI (Osteoarthritis Research Society International).

However, there has been no recent conference addressing progression of OA of the hip. Hence, a workshop was organized with a limited number of experts. Specific questions discussed by the panel addressed the methods for determining structural progression of OA of the hip.

**Methods**

A Medline search for OA of the hip for the time period of January 1, 1993 through January 1, 2003 was performed, with reference to disease progression. The search included the following keywords: hip osteoarthritis, hip progression, hip X-ray. Appropriate articles were reviewed and constituted the basis of the panel discussions.

After reviewing the literature, a group of experts met for a 1.5-day session. The purpose of the preliminary half-day meeting was to collate the questions and outline the next day’s agenda. The agenda for the full-day meeting included several presentations of published and unpublished data, with continuous discussion about the questions raised in the preliminary meeting. Indeed, the questions were presented at the beginning of the day and reassessed several times throughout the day.

Listed below are the results of the panel discussion, that is, a consensus which is based on the best knowledge of the experts with the support of the available evidence.

For purposes of this review, ‘joint space width’ (JSW) will refer to interbone distance and ‘mean JSW’ will refer to the average JSW from the pelvic brim to the medial extent of the acetabular roof. The joint space area is defined as the area of the interbone distance for the area between the pelvic brim to the medial extremity of the acetabular roof, or to the medial rim of the articular surface of the acetabular roof. The reduction in JSW can be rated over time as the ‘rate of narrowing of the JSW.’ To be meaningful, the JSW narrowing rate should reflect the same length of follow-up for all study subjects, and is often annualized.

**Results**

**BACKGROUND**

Guidelines have been developed for the conduct of clinical trials in OA of the hip. There appears to be two types of therapeutic trials for OA: (1) pain and/or function, and (2) structure (disease) modification. The medical literature is replete with trials directed at symptom modification. The concept of structure modification is relatively new, and no agent is universally accepted as being able to alter progression of OA. Part of the problem in assessing structure modification concerns the use of a validated outcome measure. Many studies on structure modification have been conducted in animals to determine joint preservation. Although useful, information from animal models is not easily transferred to humans. Since access to joint specimens in humans is limited, there is a need for a surrogate outcome measure that reflects joint changes.

**MEASURES OF PROGRESSION**

Among the surrogates for hip OA structural progression, several potential candidates could be considered as putative outcome measures for structure modification (Table 1).

<table>
<thead>
<tr>
<th>Potential measures of progression of osteoarthritis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
</tr>
<tr>
<td>- Arthroscopy</td>
</tr>
<tr>
<td>- Computed tomography</td>
</tr>
<tr>
<td>- Magnetic resonance imaging</td>
</tr>
<tr>
<td>- Radiography</td>
</tr>
<tr>
<td>- Ultrasound</td>
</tr>
<tr>
<td>Process</td>
</tr>
<tr>
<td>- Bone scanning</td>
</tr>
<tr>
<td>- Laboratory markers</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td>- Time to joint replacement</td>
</tr>
<tr>
<td>- Patient-relevant symptoms and function*</td>
</tr>
</tbody>
</table>

*Not discussed in the text.

After reviewing the literature, the consensus of the group of experts was that each of the candidates has potential; however, at this time only radiography is validated adequately to be regarded and used as a primary endpoint for change in OA of the hip.

**Arthroscopy**

Arthroscopy of the hip is not as advanced as arthroscopy of other joints, such as the knee. Hip arthroscopy is performed for loose bodies, labral injuries, to identify focal chondral lesions or for "clean-out" in late stages of Legg–Calve–Perthes disease. Because of the anatomy, only the periphery and the iliofemoral joint can be examined, the latter requiring distraction. However, labral disruptions that can be detected arthroscopically, may predict progression.

**Bone scanning**

Bone scanning can reveal increased localization of radionuclide in and about the hip related to subchondral remodeling and synovitis. When OA of the hip is severe enough, the bone scan may be difficult to separate from later stages of avascular necrosis of bone. Although there are some data on bone scanning for hand and knee, the panel could find no study examining bone scanning for progression of OA of the hip.

**Computed tomography**

Computed tomography (CT) provides detail about the subchondral bone with the potential to measure JSW. No publications could be found that have performed cross-sectional or longitudinal studies of CT in OA of the hip.

**Laboratory markers**

Laboratory markers for progression of OA of the hip have not yet been validated and there is less information for the hip than for OA of the knee. Of the limited information available for the hip, C-reactive protein, cartilage oligomeric matrix protein (COMP), YKL-40, matrix metalloproteinase-3 and -9, and tissue inhibitor of metalloproteinase-1 (TIMP-1) show the greatest promise. In a 1-year longitudinal study, COMP (but not bone sialoprotein) correlated with the yearly narrowing of the mean JSW. These data were not adjusted for ethnicity or sex. In another 1-year longitudinal study, TIMP-1 at over 600 ng/ml (but not hyaluronic acid serum levels) were predictive of narrowing of the JSW. Both of the longitudinal studies used the radiograph as their reference for OA progression. Some of the studies attempted to detect
A composite index has been proposed that includes symptom, structural and therapeutic variables. Further testing of this score suggests that it may be used to determine which patients should be referred for joint replacement. The same group suggested changes in JSW to be the best predictor of the need for hip arthroplasty; hence, time to joint replacement could be a surrogate for narrowing of the JSW; it also suggested that a reduction in JSW of 0.2 mm or more (15%) in a year, or 0.4 mm (20%) in 2 years has acceptable sensitivity and specificity for decision on joint replacement surgery. Explicit criteria have also been suggested, but to date there has been no further validation of the utility of these criteria.

### Ultrasound

The panel could find no trials examining ultrasound for assessment of OA progression of the hip. It is uncertain if the methodology has become sensitive enough to assess thickness or volume of articular cartilage.

### Radiographic technique and processing

Hips radiographs can be obtained in the standing or reclined position. In a study of 30 patients, weight-bearing radiographs of the hip were narrower than the extended view only for those with a maximum JSW < 2.5 mm; this suggests weight-bearing radiographs may be preferable for prospective structure-modifying trials. However, for the consensus among the group of experts was that there may not be any significant advantage to standing, as radiographs obtained while reclined in normals appear similar (Table II). However, since normal hips are not OA hips, the choice remains between standing and supine positions. If the radiograph is obtained standing, the feet should be internally rotated 10°-20°, and a foot map should be used. Variations in foot rotation increases the variability in JSW measurements. If the radiograph is obtained while reclined, there should be a slight medial stress (i.e., internal rotation or inversion of the foot) to press the femoral head toward the acetabulum. The X-ray beam should be centered to the upper pole of the symphysis pubis and both hips included in a single film.

### Time to hip surgery

Comparison of survival curves between treatment groups from study entry for time to hip replacement surgery is a measurable surrogate, and attractive as an endpoint for its simplicity. Objectives to using time to surgery as an endpoint include the variation among countries on the frequency of joint replacement surgery, variations in access, and variable clinical criteria for hip replacement being used.

Many factors influence the decision for hip surgery, such as: symptoms; change in X-ray; patient's needs, requirements and willingness; co-morbidities; waiting lists, social; economic; societal; insurance coverage; etc. These criteria are difficult to standardize since there is considerable variation among surgeons as to when joint replacement becomes indicated and unfortunately, there are no generally accepted guidelines for hip joint replacement surgery other than joint pain and an abnormal radiograph.

Hence, it appears necessary to determine at what point joint replacement would be appropriate. Lequesne et al. developed the algofunctional index for such a purpose. Although not validated for that purpose, a score in the range of 10–14 out of a maximum of 24 was considered as an indication for surgery.

<table>
<thead>
<tr>
<th>Table II: Hip JSW on pelvic radiographs: comparison of standing with supine technique. Ninety-two consecutive patients due for total hip arthroplasty were included, and a standing AP view added to the standard protocol. Both radiographs were taken with 100 cm film-focus distance. One hundred and seventy-three hips were available for comparison.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Standing (mm)</td>
</tr>
<tr>
<td>Supine (mm)</td>
</tr>
</tbody>
</table>

mm, millimeters; Min, minimum; Max, maximum; SD, standard deviationData from Ingvarsson et al. 79
50 patients. The combination of views detected narrowing of JSW in about one-third of the patients suggesting that the two views may be of benefit in prospective trials; however, this has not been tested and verified to date.

In summary, antero-posterior radiographs of the pelvis that includes both hips with feet rotated in 15–20° should be obtained utilizing a foot map with slight advantage of weight-bearing over reclined position in more severely narrowed hips.

**Measurements**

Once obtained, films should be forwarded to a central location for reading. There should be an initial reading of the screening film for confirmation of eligibility to enter the study (i.e., confirmation of diagnosis and stage of OA). For purposes of classification, the use of an atlas provides a set of standards. Once selected, the same atlas should be used throughout the trial. The reading of individual features is more sensitive to change than the global Kellgren & Lawrence score.

In summary, studies should include central readings of radiographs using individual radiographic features graded from a published atlas. Consensus is that radiographs should be obtained yearly, although there have been no studies to examine the time needed to detect minimal change. This recommendation is based on published trials, discussed below. Obtaining annual radiographs is felt to minimize loss of information with intent-to-treat analysis of discontinuations. It is to be noted that by intent-to-treat analysis, we mean that all patients receiving at least one therapy are included in the final analysis. If statistical imputation strategies are to be used for missing data, then these strategies should be specified in the statistics plan of the protocol.

**Reading of images**

A coordinating center collates and codes the radiographs. The code should be alpha-numeric, and should blind the reader to the demographic information. The panel of experts felt that all reading systems contain some bias. It may be that the least bias would occur when the reader(s) gathers and reads each subject's set of films together, without knowledge of their sequence. For rheumatoid arthritis, there has been a recommendation that the sequence of reading radiographs be known to the reader. The expert group's consensus was that sequence should be unknown to the reader(s), since there is risk for greater bias in OA, and progression is not inevitable.

While reading, the assessor should “break” from reading every hour to prevent fatigue. Readings should include test-retest and a calculation of inter- and intra-observer variability when there is more than one reader. The reader should be seated and the viewing box should be horizontal.

In relation to the number of readers, in a study of 118 twin pairs, it was determined that there should be a minimum of one skilled reader of the radiographs. The conclusion was that for less experienced assessors, two may be needed. Indeed, most published trials have included one highly skilled reader. In contrast, in an examination of 40 sets of hip radiographs, read by eight skilled readers, it was found that the most consistent results occurred with three readers.

For OA of the hip, four reading techniques for a single reader were compared in a study of 104 patients with two radiographs separated by 3 years. Abnormal images were more frequently detected by Kellgren & Lawrence grading with paired readings and landmarks identified on the film. In a study that compared two different grading methods, the overall number of OA hips was the same, but they overlapped only some of the same individuals, without demonstrating superiority of one of the methods.

In summary, a central reading center should use blinded radiographs to patient and sequence. A very experienced reader is acceptable, but readings may be more consistent with up to three readers. Consistency of interpretation needs to be established for all trials with rereading of radiographs.

**What and how to measure**

Of the possible anatomic features of the hip (Table III), measurement of the JSW at the narrowest point appears most consistent when examining progression. There are several methodologies for determining JSW that have been proposed (Table IV).

The distance between bony plates correlates with the articular cartilage thickness, and can be read with a graduated magnifying glass as described by Lequesne. Joint space surface area and mean JSW were measured using a computer program that analyzed digitized frontal weight-bearing pelvic radiographs. Three separate films for each of 20 hips were reviewed five times by two observers. The inter-observer coefficient of variation for joint space area and mean JSW was 5% and 4%, respectively. The correlation coefficient with the Lequesne magnifying eyepiece technique was $r = 0.89$, suggesting computerized readings may be useful in clinical trials.

Alternative digitizing programs have been developed that automatically find the articular margins of the hip to produce measurements of the minimum JSW and joint space area.

**Table III**

<table>
<thead>
<tr>
<th>Radiographic features of osteoarthritis of the hip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Osteophytes</td>
</tr>
<tr>
<td>Superior femoral</td>
</tr>
<tr>
<td>Inferior femoral</td>
</tr>
<tr>
<td>Inferior acetabular</td>
</tr>
<tr>
<td>Superior acetabular (difficult to interpret)</td>
</tr>
<tr>
<td>Joint space narrowing</td>
</tr>
<tr>
<td>Superolateral</td>
</tr>
<tr>
<td>Superior</td>
</tr>
<tr>
<td>Superomedial</td>
</tr>
<tr>
<td>Subchondral sclerosis</td>
</tr>
<tr>
<td>Femoral</td>
</tr>
<tr>
<td>Acetabular</td>
</tr>
<tr>
<td>Subchondral cysts</td>
</tr>
<tr>
<td>Femoral</td>
</tr>
<tr>
<td>Acetabular</td>
</tr>
<tr>
<td>Buttressing</td>
</tr>
</tbody>
</table>

**Table IV**

<table>
<thead>
<tr>
<th>Methods of measuring JSW</th>
</tr>
</thead>
<tbody>
<tr>
<td>General reading</td>
</tr>
<tr>
<td>o Kellgren &amp; Lawrence grade</td>
</tr>
<tr>
<td>o Scale (e.g., 0 = normal; 3 = complete loss of joint space)</td>
</tr>
<tr>
<td>Caliper</td>
</tr>
<tr>
<td>o Hand held</td>
</tr>
<tr>
<td>o Digitized</td>
</tr>
<tr>
<td>Magnifying glass</td>
</tr>
<tr>
<td>o mm scale in eyepiece</td>
</tr>
<tr>
<td>Digitized image</td>
</tr>
<tr>
<td>o Assessor-marked landmarks</td>
</tr>
<tr>
<td>o Computer-assisted landmarks</td>
</tr>
</tbody>
</table>
Radiographs of a subset of the ECHODIAH study cohort (discussed below) were measured both by using the eyepiece and by computerized analysis of hip film radiographs that were digitized\textsuperscript{49,52}. Both methods were judged efficient, with the computerized technique showing less variation.

An alternative to the measurement of the JSW with a graduated eyepiece is the use of an electronic caliper. This was tested and found efficient in 100 hip radiographs, including 50 patients with OA\textsuperscript{54}.

In a study comparing JSW on films read manually or digitized, the smallest detectable difference of JSW was 0.78 mm when measured manually and 0.67 mm when measured by computer, for a non-statistical difference\textsuperscript{55}. The smallest detectable difference was less (0.47 mm) by computer when calculating the average JSW. The smallest detectable difference was 0.5 mm in the ECHODIAH study\textsuperscript{53}.

At this time, there are no prospective studies examining images that were obtained by the new digitized techniques that no longer use film images.

In summary, sequential measurements of JSW by interbone distance is the recommended method for detecting progression. Measurement of JSW can be performed with a calibrated eyepiece, an electronic caliper, or by computer. Consistency of interpretation needs to be established for any method used.

The radiographic JSW

Normal

The superior aspect of the normal hip demonstrates a JSW of 2–5 mm. In a study of 78 normal hip radiographs, there was no difference in JSW between right and left sides, but women had a narrower mean JSW than men (women: 6.4 ± 1.0 mm vs men: 7.2 ± 1.0 mm; mean ± SD; \( P = 0.001 \))\textsuperscript{48}. Another review of 118 normal radiographs measured the average JSW to be 3.6 ± 0.6 mm in the right or left hip, without an age-related decline in JSW\textsuperscript{56}. However, there was an age-related decline in JSW in another study, even when OA of the hip was excluded\textsuperscript{57}. Although the JSW of hips of women were narrower than men, this difference was not significant when adjusting for height\textsuperscript{58}. In summary, most normal hip radiographs are between 2 and 5 mm at its narrowest point.

Abnormal

The prevalence of hip OA in Iceland has been assessed\textsuperscript{48}. Reclined pelvic scout X-ray films (\( N = 1517 \)) obtained prior to double contrast barium enemas were reviewed for the presence of hip OA, using a minimal JSW of 2.5 mm or less as a criterion for hip OA. The overall prevalence was 11\%, rising each decade to 35\% in those \( \geq 85 \) years old. In a follow-up study, 294 hips were randomly selected from 3002 radiographs. Intra- and inter-observer reliability was examined\textsuperscript{59}. Minimum hip JSW was measured with a millimeter ruler and global assessment made with the use of an atlas (Kellgren & Lawrence 2–4). Only 166 (66\%) were diagnosed as OA by both systems. Inter- and intra-observer reliability was greater with JSW measurement than with global assessment. The use of a qualitative “greater or lesser” JSW was even less consistent utilizing Kellgren & Lawrence readings.

In summary, most studies reliably use hip JSW to assist in the diagnosis of OA.

Range of JSW for inclusion in a trial

In a measurement of the narrowest point, OA was considered when the hip JSW was \( \leq 2.5 \)\textsuperscript{58} or \( \leq 3 \) mm\textsuperscript{59}.

It may be that structure preservation cannot be achieved when there has been a loss of cartilage beyond a critical thickness. Although that thickness is not known, it appears that a minimum JSW is needed. The consensus of the group was that a minimum joint space of 2 mm would be ideal. However, because of recruitment problems, a lesser joint space may be considered, perhaps down to 1 mm. Hence, measurement of disease progression is possible. A hip JSW of less than 2 mm appears to be near end-stage disease and should not be included in clinical trials of structure modification. However, although 2.5 mm may be a relevant lower limit for inclusion in a structure-modifying trial, 1.5 mm correlated better with clinical symptoms\textsuperscript{39}.

On the other hand, if the JSW is over 5 mm, although there may be some degree of OA there is probably going to be slow progression of the disease. The exception might be if the JSW is narrower than that of the contralateral hip by at least 0.5 mm (note: this value is empirical and not tested).

In summary, the experts agreed that at this time, the best radiographic reflection of progressive joint disease is the change in JSW, more accurately described as interbone distance. The measurements should be taken at the narrowest point on an antero-posterior radiograph, and that same point measured in all films (minimum JSW). The minimum JSW at the initiation of a study should be at least 2.0 mm with a Kellgren & Lawrence of at least grade 3.

Progression

The actual pattern of loss of JSW in the population is not known. In a study of 463 patients with OA of the hip, the average JSW decreased from 2.2 ± 0.8 mm to 1.7 ± 1.0 mm after 2 years\textsuperscript{60}. In an editorial, Lequesne reported rates of progression as between 0.22 and 0.30 mm/year from three studies\textsuperscript{61}.

A reduction in JSW can also be expressed as loss of mean JSW\textsuperscript{52}. A reduction of 0.43 ± 0.43 mm yearly might be expected, when using a computerized technique. Note the wide variation of the measurement.

It has also been proposed that a reduction in JSW can be expressed as rate of narrowing of the JSW (JSN, JSW narrowing rate), often annualized. For example, for measurement made at two time points, a person’s annual rate of JSN equals the difference from the initial to the final measurement divided by the time between measurements; it is then either scaled up or down to reflect an annualized rate. However, as discussed in the Biostatistics section below, there is little data to support loss of JSW over time as being linear rather than episodic. Hence, this strategy is problematic.

Another method is to compute Kaplan–Meier curves for “number of days since baseline without progressing more than a clinically and statistically meaningful amount” (e.g., 0.5 mm)\textsuperscript{3}. As discussed in the Biostatistics section below, this is the recommended method for the primary comparison of JSN between treatment groups when missing data are ignored.

In summary, the temporal pattern of rate of loss of JSW in OA of the hip is unknown. As above, annual radiographs are recommended. Outcomes of Kaplan–Meier curves should be used for the primary comparison between groups when missing data are ignored.

How can those at risk for future progression be selected.

At present, there is no single anatomic feature or set of features that will efficiently predict progression or rapid progression of OA of the hip.

In a study of 272 hips with OA, 80\% were eccentric and 20\% were concentric\textsuperscript{63}. The eccentric femoral heads
migrated superolaterally or superomedially. This group underwent total joint replacement about 4 years after the onset of symptoms. Concentric femoral heads migrated medially, causing progressive thinning of the medial wall of the acetabulum, with a slower progression.

In a study of 1578 hips with OA, the distribution of change was superolateral 59%, medial 26%, and global 15%64. Approximately 20% were bilateral and 55% were women.

In a study of 70 hip joints, superomedial or concentric (medial) migration of the femoral head was most sensitive to change60.

In summary, studies are unclear as to whether superomedial or superolateral hip OA disease are more likely to progress. There are no published studies demonstrating that either demographic or clinical measures at the baseline of a study can predict progression.

**What represents progression.** Radiographic features have been examined for their ability to reflect progression. In a prospective study of 463 patients with OA of the hip by American College of Rheumatology criteria, several parameters including osteophytes, osteosclerosis, subcondral cysts, femoral head migration, Kellgren & Lawrence grade, subjective change in JSW and JSW in mm were assessed for progression60. Progression was defined as a 0.5 mm reduction in JSW using a magnifying glass. Wide distribution of loss of JSW, subcondral bone production, and severity of loss of JSW were predictive of progression.

The minimum clinically important difference for detecting change of JSW in radiographic progression of OA of the hip was determined to be 0.4 mm; this was determined in a study of 298 patients followed for 3 years65. The mean loss of JSW for individuals was 0.63 ± 0.74 mm.

From a mathematical perspective, the mean JSW was felt to be more sensitive to change than minimum JSW in a retrospective review of 69 digitized images of the hip with a decrease of 0.43 ± 0.43 mm per year66.

Loss of JSW was determined in the contralateral hip of a cohort of 99 patients postoperative from total hip replacement surgery66. Within 20 months, an initial median JSW of 3.48 mm declined at a rate of 0.10 mm/year. The authors determined that 15% of the hips had an accelerated decline of >0.2 mm/year. They could not identify an associated feature that would predict the more rapid progression.

In a similar study of 61 patients followed for 61 months, the mean JSW narrowed 0.43 ± 0.43 mm/year67.

Progression was felt to be defined as ≥0.6 mm/year in a 1-year study of 30 patients68. Predictors of progression on a larger cohort were felt to be: JSW of ≤2 mm (relative risk, RR 2.11), superolateral migration of the femoral head (RR 4.25), women (RR 2.51), Lequesne functional index >10 (RR 2.60), and age >65 years old (RR 1.90).

In summary, progression of ≥0.5 mm in JSW at any hip site appears clinically relevant. For minimum JSW, this value represents the minimally perceptible difference: it takes into account error in reading and variation in imaging technique (e.g., positioning).

**THERAPEUTIC TRIALS (LESSONS LEARNED)**

Rejholc studied the effects of an intramuscular glycosaminoglycan polysulfuric acid complex on 112 patients with OA of the hip using a matched pair design for the placebo group69. After 10 years of therapy, there were 83 pairs available for evaluation. X-rays were graded by the Kellgren & Lawrence system. The grade of severity of the treated group was less than their matched pair control group, whereas they were the same at baseline. Other measures, including mobility, climbing a 15-step staircase, anti-inflammatory drug consumption, development of OA in the contralateral hip, the number receiving total hip replacement, and other secondary measures statistically favored the treated group. JSW was measured in 35 pairs after 16-year follow-up with a 47.5% reduction in JSW for the treated group vs a 74.5% reduction in JSW in the matched pair placebo group. Radiographs were obtained in the standing position. The report did not specify whether the radiographs were obtained or read centrally, or how reproducible the radiographs were. There was no indication of blinding of radiographs or the skill of the reader or readers. There was no test–retest reported. Most significantly, the radiographs were not available for re-examination for validation of the results.

Pavelka *et al.* re-studied the intramuscular glycosaminoglycan polysulfuric acid complex on 400 patients with OA of the hip or knee over a 5-year period70. Radiographs were obtained yearly in the standing position by a single technician with one X-ray machine, visually matching the follow-up radiographs to the baseline. The radiographic JSW was measured at the narrowest point using the Lequesne technique with an eyepiece that contained a calibrated mm scale62. Films were paired by patient, but blinded as to sequence. Of the 117 patients with OA of the hip, overall JSW decreased over the 5 years by 0.21 ± 0.8 mm (mean ± SD) and 0.22 ± 0.8 mm for the treated and control groups (P = 0.53). The potential of patient selection bias was suggested by subset analysis: if one selected out those with an initial Kellgren & Lawrence grade ≥2–4 and a JSW of ≥1 mm, the treated group had a trend toward less JSW narrowing despite the small numbers of patients in these cohorts (treated, N = 25, −0.23 mm; placebo, N = 21, −0.52 mm; P = 0.11). There is concern about such selection bias: patients were included that did not progress, potentially hiding a beneficial effect of the treatment. In contrast to the prior study, radiographs are still available for re-examination.

Doudagos *et al.* performed a 3-year study of orally administered diacerein in a cohort of 507 patients with OA of the hip by American College of Rheumatology (ACR) criteria53. Patients had weight-bearing hip radiographs obtained at baseline and yearly thereafter. Minimum JSW was measured by a central reader, unaware of the sequence of radiographs or treatment. There were four primary efficacy variables: reduction of JSW of ≥0.5 mm in the intent-to-treat population; reduction of JSW of ≥0.5 mm in the patients completing the study (completers); mean JSW change in the intent-to-treat population; mean JSW change in the completers (Table V). The pitfalls of an intent-to-treat analysis

**Table V**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Diacerein</th>
<th>Placebo</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients with ≥0.5 mm JSW loss (number [percent]):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intent-to-treat</td>
<td>112 (51%)</td>
<td>136 (60%)</td>
<td>0.036</td>
</tr>
<tr>
<td>Completer</td>
<td>62 (47%)</td>
<td>86 (62%)</td>
<td>0.007</td>
</tr>
<tr>
<td>Mean JSW loss, in mm (mean ± SD):</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intent-to-treat</td>
<td>0.39 ± 0.81</td>
<td>0.39 ± 0.75</td>
<td>NS</td>
</tr>
<tr>
<td>Completer</td>
<td>0.18 ± 0.25</td>
<td>0.23 ± 0.23</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Populations: intent-to-treat: 446 patients (diacerein = 221; placebo = 225); completer: 269 patients (diacerein = 131; placebo = 138). NS = not significant.
utilizing the last observation carried forward (LOCF) are illustrated and discussed in this study, that being the only primary outcome that was not statistically different between study groups (see Biostatistics section below).

SPECIAL CLINICAL POPULATIONS

The etiology of idiopathic OA of the hip is—by definition—not known, but it is suspected that the risk factors for different subsets of the population are different. Studies may wish to target special populations because of pattern of joint disease or more consistent predictability of progression.

There appears to be a relatively small subset of patients with OA of the hip who have rapidly progressive hip OA. Lequesne has empirically defined rapid progression as a loss of JSW of >2 mm/year. This progression occurs over a period of months, rather than years. It would be clinically relevant to detect this group: early detection would allow for more close observation and potentially separate therapeutic protocols. In a study of 136 patients referred for joint replacement, rapid radiographic progression was more commonly associated with superior migration or an articular bone response. In a study of 61 patients over 81 months, rapid progressive loss of JSW seemed to relate to older age and the absence of osteophytes. In contrast, no progression was noted with indeterminate, medial or axial migration, protrusion or mild OA. These criteria for rapid progression have not yet been validated in prospective trials.

The relation of obesity to hip OA is less clear than for the knee. However, there may be a relationship between obesity and progression as suggested by an increase in the risk of total joint replacement from the Nurses' Health study. Of the 568 women who reported total hip replacement, those with a body mass index (BMI) of ≥35 kg/m² had a relative risk of 2.6 over the reference population (with a BMI <22 kg/m²). The relative risk for eventual hip replacement was 5.2 if the BMI was ≥35 kg/m² when the nurses were aged 18 years.

BIOSTATISTICS

With current radiographic techniques, structure-modifying trials for OA of the hip will need to be of 2 or more years' duration. For various reasons, long duration trials are likely to yield a significant percent of subjects not completing the trial. This presents some statistical challenges, particularly related to the primary intent-to-treated analysis, wherein all patients randomized to treatment or control are included in the comparison of groups.

Two methods are traditionally employed to compare the progression of a treatment group to a control group: (1) the comparison between groups of the mean of changes in JSW from study entry to the end of the study; (2) the comparison between groups of the proportion of patients developing JSN (with how much loss of JSW is considered relevant pre-specified in the protocol—many have suggested a loss of ≥0.5 mm of JSW to be significant). Anyone achieving a 0.5 mm reduction in JSW at any time in the protocol would thus be considered a therapeutic failure.

A third option is to use mean values for the treatment groups of each subject's percent reduction of JSW from baseline. In that circumstance, one could empirically set a difference of 25% or 50% reduction from baseline in JSW between treated and control groups as being a relevant outcome.

Yet a fourth option is to compare the mean values of the treatment and control group of patients' annualized JSN rate. The difference in means is usually statistically assessed with a two-sample t-test, and the difference in proportions with a chi-square test.

Any of the four above methods of defining outcomes to compare groups can provide valid statistical inferences while ignoring the missing data. Validity is based on the subjects in the study for the same length of follow-up, the drop-out rate is low and occurs for the same reasons across treatment groups. With the above outcomes, biased conclusions are likely to result when large amounts of missing data are present, especially regarding early discontinuations. If one of these methods is employed, then an intent-to-treat analysis requires missing data to be replaced with "imputed" values. How the imputations are done can greatly influence study conclusions.

For example, in one of the several strategies for statistical analysis of results used in a recent study, the "missing data strategy" replaced all missing data after a subject discontinued treatment with his or her last JSW measurement. Since the pattern of JSW loss is not simply constant (i.e., unchanging over time), early discontinuations led to distorted results with use of this strategy. Additionally, such a filled in data set with a subject's missing data imputed to equal his or her last JSW measurement, artificially reduces variability between "observations" and gives the analyst "sample" sizes larger than reality. Both reduced variability and increased sample sizes contribute to smaller than entitled to standard errors of outcomes, and may lead to optimistically small P-values and biased conclusions.

For the four outcomes discussed above, a more defensible missing data strategy is based on introducing multiple probabilistically plausible imputations of the missing responses. Each set of imputations results in a complete data set. Under mild assumptions, the imputations are unbiased. From the multiple imputations, one obtains estimates of study outcomes and of their uncertainty, allowing valid statistical inferences to be made on an intent-to-treat basis. The multiple imputation strategy employed should use all of the repeated radiographic data available on patients.

The comparison of the Kaplan–Meier curves between treatment and control groups for "number of days since baseline without progressing more than 0.5 mm" is a statistically valid intent-to-treat method when the missing data are ignored. This approach provides a method for dealing with varying durations of follow-up to achieve the threshold value of JSN and appropriately deals with cases dropping out of the study before the threshold value occurs ("censored data") as shown in the diacerein study. Statistical comparisons of the Kaplan–Meier curves are most often based on the log-rank test.

Postscript

Measuring the progression of OA of the hip presents challenges. Assuredly, newer techniques will be developed that will more accurately measure the hip as a joint and not just the area normally occupied by articular cartilage. However, the new methods will need validation, and validation takes time. As products are developed that have the potential to alter the course of OA of the hip, we must test them with the best methods available. The consensus group feels that the information synthesized above summarizes the state-of-the-art in measuring progressive OA of the hip.
References


32. Conrozier T, Lequesne MG, Tron AM, Mathieu P, Berdah L, Vignon E. The effects of position on the


44. Lequesne M, Cadet C, Auleley GR. New technique and reproducibility of the manual measurement of the osteoarthritic (OA) hip joint space (JS) (Abstract) Osteoarthritis Cartilage 2000;8:156.


66. Goker B, Doughan AM, Schnitzer TJ, Block JA. Quantification of progressive joint space narrowing in osteoarthritis of the hip: longitudinal analysis of the