Outcome Measures for Clinical Trials of Disease Modifying Osteoarthritis Drugs in Patients with Hip Osteoarthritis

MAXIME DOUGADOS

As the frequency of hip osteoarthritis (OA) increases as a result of the aging population, this disorder will increasingly become a major health problem. Consequently, it will be important to optimize treatment and evaluate interventions that may prevent or delay progression of the disease. Several outcome measures have been recommended for use in studies to evaluate potential disease modifying drugs for OA (DMOAD). Variables related to symptoms are reliable, clinically relevant outcome measures, but they are subjective and may be influenced by treatments that do not alter the metabolism of articular cartilage or that have a direct effect on other joint tissues, such as analgesics. Consequently, symptomatic effects are considered to be secondary outcomes in DMOAD trials.

Structural variables usually assess the rate and extent of cartilage breakdown [e.g., a change in radiographic joint space width (JSW) or in cartilage volume, as measured by magnetic resonance imaging (MRI), or in the incidence and extent of cartilage surface defects, as assessed by arthroscopy or MRI]. Such outcome variables are accurate, have high intrinsic validity, and are usually considered as the primary outcome to be assessed in studies of DMOAD. However, the clinical relevance of the results obtained remains debatable.

At this time, only a few trials have been published that have evaluated the effects of a treatment on the structural progression of hip OA[1,2]. There are numerous reasons for this, including the large scale and high costs of such studies, which are due to marked variability in the natural history of the disease and the absence of a validated, universally accepted structural outcome measure. Therefore, during the past few years, several methods of assessing structural variables in DMOAD trials have been proposed. Assessment of JSW in the plain radiograph by measurement of the minimum interbone distance has been standardized and is valid and reliable[3-5]. Standardization of radioanatomic positioning of the hip in serial examinations is much less problematic than that of the knee. Although other techniques are being developed, quantification of joint space narrowing (JSN) by serial measurement of hip JSW is currently accepted as the most sensitive technique for determining progression of structural damage of the hip and has been used as an outcome measure in the few published trials of primary or secondary prevention.

Obtaining a dichotomous outcome variable for use in trials of hip OA. Evaluation of the interbone distance in radiographs provides a continuous variable. At a group level, the results are usually presented as the mean change in JSW. This is a powerful parameter for statistical analysis of the results of a clinical trial but is very difficult for the clinician to interpret because it does not present the results as "the percentage of patients with or without a key event," "time to a key event," or "number of patients needed to treat to prevent a key event," an approach currently applied in studies of therapies for other disorders, such as osteoporosis (using fracture as an endpoint) or coronary heart disease (using myocardial infarction as an endpoint).

Interest exists, therefore, in identifying a valid dichotomous outcome variable that reflects the natural history of hip OA, i.e., that would permit presentation of the results of a randomized controlled trial and the percentage of patients with clinically relevant progression. We have considered 2 approaches: (1) artificially creating a key event by dichotomizing the variable, "change in JSW"; and (2) using the incidence of an event that is generally considered to reflect disease severity, such as total hip arthroplasty (THA), to differentiate between changes in the active treatment group and the control group in a clinical trial.

A great deal of effort has gone into the evaluation of changes in symptoms in therapeutic trials. Participants in the 1996 OMERACT meeting[6] considered which variables should be collected in clinical trials of OA drugs. For assessment of the effects of therapy on symptoms, 3 domains were considered to be "most important" (i.e., 90% of the participants agreed upon the importance of each): pain, function, and patient global assessment. In addition, for studies that are at least one year in duration, it was recommended that joint structure be evaluated (e.g., by radiography), because of concerns about safety (not to assess the effects of the drug on structural change, which was considered not to be feasible in this relatively short time span). Responder criteria have since been developed based upon analyses of the results of 10 clinical trials of symptomatic therapies, permitting classification of clinical responses as "yes" or "no," i.e., as a dichotomous variable[7].

From the Service de Rhumatologie, Hôpital Cochin, Paris, France.
M. Dougados, MD, Professor of Rheumatology.
Address reprint requests to Dr. M. Dougados, Rheumatology, Hospital Cochin, 27 Rue du Faubourg St-Jacques, Paris 75014, France.
E-mail: maxime.Dougados@cch.ap-hop-paris.fr

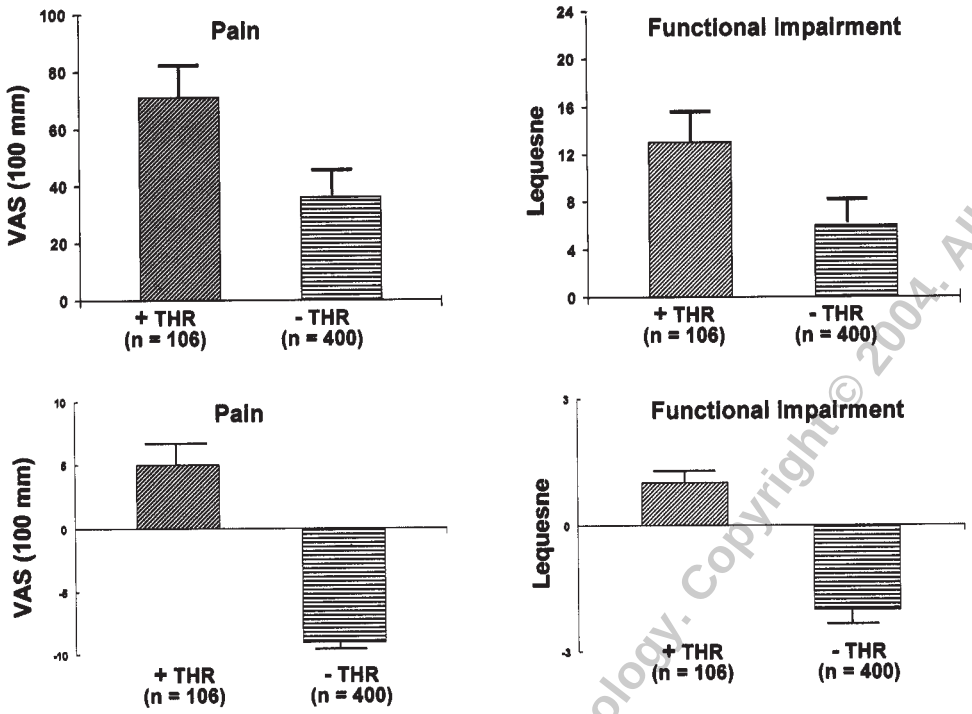

Figure 1. Face validity of "requirement for performance of total hip arthroplasty" as an outcome measure. The upper panels indicate last values recorded for hip pain and function prior to surgery. Lower panels indicate changes in hip pain and function, relative to the baseline scores, in patients who underwent THA and those who did not in the 3 year period after enrollment in the ECHODIAH study. From Dougados, *et al*. J Rheumatol 1999;26:855-61.

In addition to symptoms, many other variables might be considered as outcome measures in clinical trials. Further, it is possible to combine variables related to symptoms and those related to structure-modification to achieve a single endpoint.

We have asked whether the time at which the patient fulfils criteria that would ordinarily satisfy requirements for THA in a patient with OA might be a suitable endpoint. Taking the perspective of the clinical rheumatologist, we considered that pharmacologic modification of disease progression would result in prevention, or at least a delay in a decision to perform joint replacement surgery. The use of "requirement for total joint arthroplasty" (TJA) as an outcome measure had several advantages: face validity, simplicity, reproducibility, sensitivity to change, and discriminant capacity.

Consider the following: Does the requirement for surgery reflect the severity of OA, rather than, e.g., the salary of the surgeon or his willingness to perform the surgery? We recently examined the relationship between the symptomatic and structural severity of hip OA and the requirement for THA in the ECHODIAH study, a 3-year placebo-controlled randomized trial of diacerein in 507 patients with hip OA[7]. Among those who underwent surgery, scores for joint pain and function deteriorated prior to the operation: in contrast, among those who did not have surgery, pain and function improved during the 3-year period of observation (Figure 1).

A similar analysis was performed with respect to severity of structural changes of OA (Figure 2). Among patients who underwent THA, mean JSW in the last radiograph obtained prior to surgery was 0.6–0.7 mm; among those who completed the 3-year trial without surgical intervention, JSW was 2.1 mm (Figure 2). The rate of narrowing, as reflected by a change in JSW, was > 1 mm per year among those who underwent surgery, but only 0.2 mm per year among patients who did not have surgery, indicating that subjects who underwent THA had not only more severe symptomatic OA but more rapidly progressive structural change, thus establishing the face validity of THA as an outcome measure.

The simplicity underlying use of this outcome measure may be deceptive, however. Judgment by the physician that the patient's condition warrants surgery is not necessarily tantamount to actual performance of the procedure: e.g., a certain percentage of such patients will refuse the procedure, perhaps influenced by the opinions of family or others.

What about the reproducibility of TJA as an outcome measure? Ideally, evaluation of this would require the presentation of a single patient with OA to several physi-

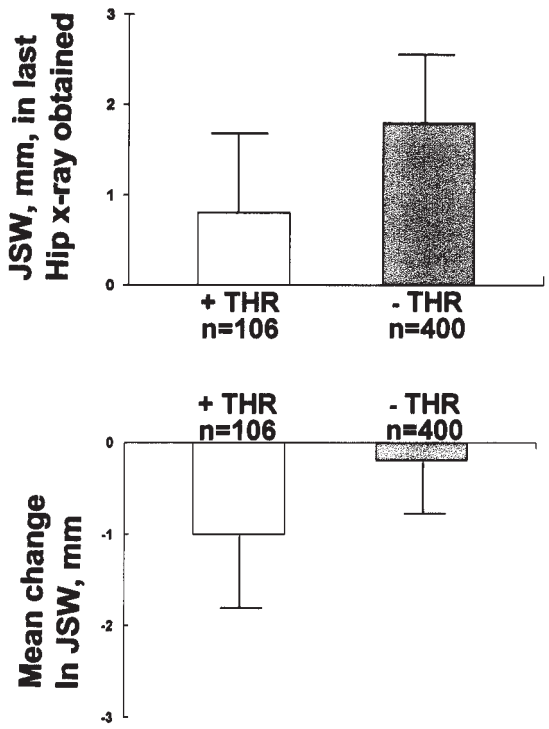

Figure 2. Face validity of "requirement for performance of total hip arthroplasty" as an outcome measure. Upper panel shows last value for joint space width (JSW) recorded prior to surgery, in mm. Lower panel shows change in JSW, in mm, relative to the baseline value, during the 3 year period after enrollment in the ECHODIAH study. From Dougados, *et al*. J Rheumatol 1999;26:355-61.

cians, each of whom would evaluate the patient and render a judgment with regard to the recommendation for TJA. Because this is impractical, we took an alternative approach: in the ECHODIAH trial, a multicenter national study, we examined the percentage of subjects with hip OA enrolled in the study who subsequently underwent THA in relation to the geographic region of France in which they resided. We found that the probability that a patient with hip OA who was enrolled in this trial would undergo THA was comparable in all regions of the country, suggesting that the indications of an orthopedic surgeon on the French Riviera with respect to performance of THA are the same as those of a surgeon in the North of France.

With respect to sensitivity to change our study indicates that the number of subjects who underwent THA each year was 8% of the total over the 3-year period of the study. That percentage has remained unchanged through the seventh year of followup. We have been able to identify subjects who were at greatest risk for TJA. To increase the number of subjects undergoing THA from 8% per year to 12–14% per year, one need only focus on patients with specific characteristics — women, age 65, with superolateral hip OA.

The discriminant capacity of an outcome measure reflects its responsiveness, or its sensitivity, i.e., its ability to discriminate between the effect of treatment and placebo. In the ECHODIAH trial, we demonstrated a statistically significant effect of diacerein on change in JSW: fewer subjects in the diacerein treatment group than in the placebo group showed progression of JSN after 3 years, although there were no significant differences between the 2 groups with respect to symptoms.

Although the primary object of the ECHODIAH study was not to achieve a decrease in the incidence of THA, the results indicated that 15% of subjects taking diacerein, and approximately 20% of those who received placebo, underwent THA during treatment or within 6 months after the discontinuation of treatment. Although this difference was not statistically significant, it permits calculation of the reduction of absolute risk and relative risk and of the number needed to treat (NNT). These analyses indicated that 1 of every 19 patients who entered this clinical trial would be spared THA after 3 years of treatment with diacerein. Is this clinically relevant? Certainly, it is easier for the clinician or patient to decide whether that is a clinically relevant difference than a change in JSW in millimeters, with a standard deviation of, e.g., 0.60 or 0.80. The importance of NNT in studies that show a statistically insignificant difference between treatment groups is, however, unclear.

Thus, TJA fulfills the validity criteria discussed above. It permits assessment of the effect of treatment on symptoms and structural change related to OA. However, it is affected by parameters that have nothing to do with the disease itself, such as comorbidity (the decision to perform THA in a patient with renal failure or recent myocardial infarction will not be the same as in a patient who is in good health except for her arthritis). It is also affected by the willingness of the patient to undergo the procedure and is influenced heavily by the healthcare system. For 2 patients with identical disease characteristics, the likelihood of having TJA may be different in Iraq or Mali than in Spain or Italy. Indeed, the healthcare system may be the most important factor accounting for the fact that one patient may undergo surgery while another, with seemingly identical disease, may not.

Taking the above into account, it is possible to utilize as an outcome measure the time at which the patient meets criteria that meet the threshold for an indication for THA, rather than the actual performance of surgery. Our colleagues in cardiology utilize this methodology in assessing cardiac transplantation for patients with heart failure, chiefly because of great variability in the availability of the surgical procedure from site to site. Further, effective pharmacologic interventions that might delay the surgical procedure may be better evaluated by considering their effect on the time at which the patient fulfils a set of criteria for surgery, rather than the time to actual performance of the procedure, which, as indicated above, may be confounded by a large number of variables[8,9].

We have recently analyzed results we obtained when we applied a set of criteria considering the time to indication for THA in the ECHODIAH study cohort[9]. We applied a very simple model that considered structural parameters, the patient's perspective, and the use of concomitant symptomatic therapy. Table 1 depicts the grading system we proposed, based on the results of a multivariate analysis. Radiographs were performed annually; other variables were obtained at each visit, e.g., "Were nonsteroidal antiinflammatory drugs taken on more than half the days in the previous 3 month period?" We weighted structural change less heavily than symptoms. We found that if a patient had a score < 40, the probability of undergoing THA during the study was less than 10%, if the score was ≥ 40, the likelihood of THA exceeded 50% (Figure 3). We do not regard a cutoff of 40 as an essential breakpoint, but merely consider this approach to be a starting point for discussion.

Finally, I would like to emphasize that we are not defining appropriate criteria for THA. International groups [Bone and Joint Decade, EuroHip, Outcome Measures in Rheumatology Clinical Trials (OMERACT)] that include orthopedic surgeons and rheumatologists are currently addressing that problem. Nor have we addressed the advantages and benefits of THA; others are doing that. Rather, it has been our purpose to attempt to develop workable outcome measures that will permit the evaluation of new therapies for OA.

Table 1. Criteria used to define the "time to indication for total hip arthroplasty." From Maillefert JF, *et al*. J Rheumatol 2002; 29:347-52.

Variable	Scoring	Score
Joint space width on last available radiograph	> 2 mm	0
	≥ 2 > 1 mm	20
	≤ 1 mm	35
Patient's overall assessment (none-mild-moderate-severe-very severe)	≤ 2	0
	> 2	15
Lequesne index	≤ 2	0
	> 2	25
NSAID intake in the 3 mo prior to the last visit	≤ 1 day/week	0
	> 1 day/week	15
Analgesic intake in the 3 mo prior to the last visit	≤ 1 day/week	0
	> 1 day/week	10

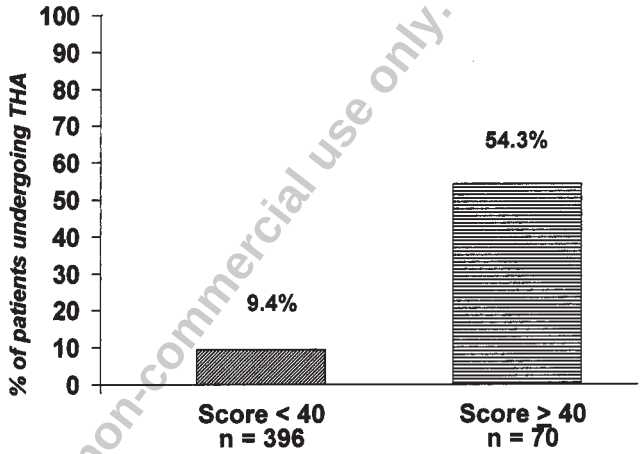

Figure 3. Frequency with which total hip arthroplasty was performed in patients with cutoff scores < 40 and ≥ 40, based upon the criteria listed in Table 1 to define an "indication for THA" among patients in the ECHODIAH study. From Maillefert, *et al*. J Rheumatol 2002;29:347-52.

REFERENCES

1. Dougados M, Nguyen M, Berdah L, et al. Evaluation of the structure-modifying effects of diacerein in hip osteoarthritis: ECHODIAH, a three-year, placebo-controlled trial. Arthritis Rheum 2001;44:2539-47.
2. Lequesne M, Maheu E, Cadet C, Dreiser RL. Structural effect of avocado/soybean unsaponifiables on joint space loss in osteoarthritis of the hip. Arthritis Care Res 2002;47:50-8.
3. Auleley GR, Duche A, Drape JL, Dougados M, Ravaud P. Measurement of joint space width in hip osteoarthritis: influences of joint positioning and radiographic procedure. Rheumatology 2001;40:414-9.
4. Auleley GR, Giraudeau B, Dougados M, Ravaud P. Radiographic assessment of hip osteoarthritis progression: impact of reading procedures for longitudinal studies. Ann Rheum Dis 2000;59:422-7.
5. Conrozier T, Lequesne M, Favret H, et al. Measurement of the radiological hip joint space width. An evaluation of various methods of measurement. Osteoarthritis Cartilage 2001;9:281-6.
6. Bellamy N, Kirwan J, Boers M, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799-804.
7. Pham T, Van der Heijde D, Lasserre M, et al. Outcome variables for osteoarthritis clinical trials: The OMERACT-OARSI set of responder criteria. J Rheumatol 2003;30:1648-54.
8. Dougados M, Gueguen A, Nguyen M, et al. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol 1999;26:855-61.
9. Maillefert JF, Gueguen A, Nguyen M, et al. A composite index for total hip arthroplasty in patients with hip osteoarthritis. J Rheumatol 2002;29:347-52.