Randomized controlled trials (RCTs) are widely accepted as the most reliable method of determining the effectiveness of specific therapies. The design, conduct, and analysis of clinical trials aim at providing valid results, which implies that the treatment effect reported represents its true direction and magnitude and that the trial minimizes bias, which can distort results.

Hip and knee osteoarthritis are together a major cause of disability and can be treated either by nonpharmacological treatment, such as surgery, rehabilitation, joint lavage, acupuncture, behavioral interventions, or spa therapy, or pharmacological treatment, such as oral drug, intra-articular injection, or topical treatments.

Randomized controlled trials have been developed essentially in the context of pharmacological treatment (ie, oral drugs; intra-articular injection; and topical, intramuscular, and intravenous treatments), but assessment of the effectiveness of nonpharmacological treatments (ie, surgery, arthroscopy, joint lavage, rehabilitation, acupuncture, and education) presents specific issues.

Objectives To compare the quality of articles of nonpharmacological and pharmacological treatments of hip and knee osteoarthritis and to identify specific methodological issues related to assessment of nonpharmacological treatments.

Design and Setting We searched MEDLINE and the Cochrane Central Register of Controlled Trials for articles of randomized controlled trials published between January 1, 1992, and February 28, 2002, in 28 general medical and specialty journals with high impact factors and assessing nonpharmacological and pharmacological treatments in patients with hip or knee osteoarthritis.

Main Outcome Measures The quality of the methods reported in the selected articles was assessed by 2 independent reviewers using the Jadad scale, the Delphi list, and guidelines found in the Users’ Guides to the Medical Literature. Investigators also used a checklist of items developed by the authors to analyze study characteristics.

Results A total of 110 articles were included in the analysis; 50 (45.5%) assessed nonpharmacological treatments and 60 (54.5%) assessed pharmacological treatments. Reports of nonpharmacological treatments had a lower global quality score than did reports of pharmacological treatments as measured by the Jadad scale (mean [SD] score, 1.4 [1.3] vs 3.0 [1.3]) and the Delphi list (mean [SD] score, 5.2 [1.5] vs 7.5 [1.1]). Lack of reporting adequate random sequence generation and intention-to-treat analyses were found in both nonpharmacological and pharmacological articles. Nonpharmacological treatments were less often compared with a placebo than were pharmacological treatments (28.0% of articles vs 71.7%). Compared with pharmacological articles, nonpharmacological articles less often described blinding of patients (26.0% vs 96.7%), care providers (6.0% vs 81.7%), and outcome assessors (68.0% vs 98.3%). Care providers’ skill levels could influence treatment effect in 84.0% of nonpharmacological articles vs 23.3% of pharmacological articles.

Conclusions In this analysis of reports of hip and knee osteoarthritis therapy, nonpharmacological articles scored lower than pharmacological articles in terms of quality. Assessments of nonpharmacological treatments must take into consideration additional methodological issues.
vestigators can standardize the dosage, measure compliance, produce identical placebos, and blind patients, care providers, and outcome assessors. In nonpharmacological treatment, it is often technically or ethically difficult to perform a sham intervention, and the blinding of patients and care providers is frequently impossible, whereas the placebo effect of nonpharmacological treatment is probably important. For example, reports of RCTs assessing joint lavage describe the effectiveness of this treatment in knee osteoarthritis. However, these studies were performed without a sham intervention in the control group, and patients were not blinded. These results are inconsistent with those of another RCT evaluating joint lavage that used a sham intervention in the control group and blinded patients and outcome assessors. These conflicting results could be linked to the choice of the control intervention or to the results’ variability. Moreover, contrary to pharmacological treatment, in nonpharmacological treatment, care providers are an integral part of the treatment; the success of the treatment depends on the care providers’ skills, experience, and enthusiasm. For example, hip and knee arthroplasty outcomes are well known to depend on surgeons’ experience and hospital surgical volume. Finally, nonpharmacological treatment is usually complex and difficult to standardize, and technical modification may occur as the procedure evolves. These methodological issues are usually not taken into account in assessment of the quality of articles evaluating nonpharmacological treatment.

The goals of this study were to compare the quality of reports of nonpharmacological treatment and pharmacological treatment of hip and knee osteoarthritis and to identify specific methodological issues in assessment of nonpharmacological treatment.

METHODS

Search Strategy


Retrieved articles were assessed by 1 of us (I.B.), who screened the titles and abstracts to identify the relevant studies. Articles were included only if the study was identified as an RCT, published as a full-text article, and assessed nonpharmacological treatment or pharmacological treatment of hip or knee osteoarthritis. Case series, uncontrolled studies, and articles published as abstracts only, editorials, news, or correspondence sections were excluded. Articles were screened for duplicate publication (ie, the same trial with results from different lengths of follow-up published twice), and, in these cases, only the more recent article was selected for inclusion.

Evaluation of Study Quality

Two independent reviewers (I.B. and F.T.) assessed the quality of the methods in the selected articles using the Jadad scale and the Delphi list because of their validity and widespread use. They also used the Users’ Guides to the Medical Literature. However, to our knowledge, no quality assessment tool specific to nonpharmacological treatment is available. Therefore, the reviewers also assessed articles using a checklist of items developed by the authors to target methodological issues when assessing nonpharmacological treatment. (This checklist is available at http://www.bichat.inserm.fr/fichierpfd/emi0357/checklistofitems1.pdf.) With the checklist, data were obtained for year of publication, funding sources (public or private), number of centers involved, type of treatment assessed (oral drug administration, topical treatment, intra-articular injection, surgery, arthroscopy, joint lavage, acupuncture, rehabilitation, or behavioral intervention), and whether data in the CONSORT (Consolidated Standards of Reporting Trials) diagram were reported in a flowchart or in the text. Information about study design, randomization mode, and appropriateness of allocation concealment was collected. Randomization sequence generation was considered adequate if selection bias was prevented by use of,
for example, a table of random numbers, random numbers generated by computer, coin tossing, or shuffling cards. Allocation was considered adequately concealed if patients and the investigators who enrolled the patients could not foresee assignments because of use of, for example, centralized randomization, pharmacy control, opaque sealed envelopes, or num-bered or coded bottles or containers.

Reviewers examined whether treatment was individualized, which supposed that the treatment’s dosage or mode was modified according to patient tolerance or comorbidity and to clinical efficacy. Reviewers evaluated whether, in their opinion, the intervention was described with enough detail to be reproducible. In situations in which care providers could influence the success of the treatment, reviewers analyzed whether the learning curve and the care providers’ experience were taken into account and whether care providers were trained.

Compliance with repeated interventions (ie, treatment necessitating iterative interventions, such as drug treatment or physiotherapy rather than surgery) was evaluated. Quantitative and qualitative compliance were distinguished. Quantitative compliance assessed, for example, whether a patient attended all the physiotherapy sessions or took all of the oral drugs prescribed, and qualitative compliance assessed whether a patient correctly performed the intervention (eg, if the home-based exercises performed by a patient were in accordance with the program prescribed). The method used to measure compliance (eg, drug dosage, pill counts, patient questioning, or diaries) was recorded.

Reviewers examined the control intervention used and whether the potential placebo effect of each intervention was similar, in their opinion. For example, reviewers considered that the placebo effect of use of a nonsteroidal anti-inflammatory drug and that of an indistinguishable placebo were similar. However, the potential placebo effect of joint lavage, performed under aseptic conditions in an operating theater with use of local anesthesia and infusion of 1 L of saline solution, compared with a single intra-articular injection is probably different. Information concerning the similarity of concomitant treatment in each group (unintended additional care provided to either comparison group) and the occurrence or likelihood of contamination (the intervention provided to the control group) between the different treatment groups was also considered.

Reporting of blinding of patients, care providers, and outcome assessors and whether blinding was tested were also studied. Blinding refers to keeping participants, care providers, and outcome assessors unaware of the assigned intervention so that they are not influenced by that knowledge.

Reviewers examined whether a sample size justification or an evaluation of the power of the study a posteriori was reported and whether an intention-to-treat analysis was reported and/or performed. Finally, we aimed at following the approach of the Users’ Guides to the Medical Literature to subjectively appraise the validity of the study. According to these guides, the final assessment of validity is never a “yes” or “no” decision but a continuum, ranging from strong studies that avoid bias to weak studies that likely yield a biased estimate of effect. For this purpose, reviewers gave a subjective evaluation of the study’s quality on a numerical rating scale ranging from 1 to 10 by answering the question, “To what extent were systematic errors or bias avoided in this report?” Since this global evaluation involves subjectivity, we cannot exclude a bias against nonpharmacological treatment studies.

Before undertaking the study, the 2 reviewers practiced evaluation of a distinct set of 10 articles. Then, during a meeting, they discussed the interpretation of the different scales to resolve any differences in scoring. During the study, each reviewer independently examined the selected articles in a different computer-generated random sequence. Reviewers assessed the title and the “Methods” and “Results” sections. Reviewers were not blinded to the journal name and authors, as evidence concerning the effect of masking on assessments of trial quality is inconsistent.

Discrepancies in the assessment of the selected articles between the 2 reviewers were resolved by consensus. For each inconsistent item, reviewers read the article again and came to an agreement. The data presented herein resulted from this consensus. The quality of the selected articles was determined by use of the mean of the 2 assessors’ global appreciation of the articles’ quality on a numerical rating scale ranging from 1 to 10, the Jadad scale (score, 0-5), and the Delphi list’s overall quality score, consisting of the number of items satisfied and ranging from 0 to 9. On all scales, a high score indicates high quality.

Statistical Analysis
Descriptive statistics (means, SDs, and minimum and maximum values) were used for continuous variables. Categorical variables were described with frequencies and percentages. The degree of agreement between the 2 reviewers was determined with use of the κ coefficient for categorical variables. Intrarater reliability was assessed by use of the intraclass correlation coefficient (ICC) for continuous variables. All data analyses were performed using SAS version 8.2 (SAS Institute Inc, Cary, NC).

RESULTS
Selected Articles
Of 198 articles identified, 119 were selected for assessment (FIGURE 1). The 79 excluded articles were abstracts only (n=5), were duplicate publications (n=3), were not RCTs (n=12), did not assess a therapeutic intervention (n=45), did not assess treatment of hip or knee osteoarthritis (n=13), or were phase 2 trials (n=1). Nine articles were secondarily excluded after obtaining the full text because they were not RCTs (n=3) or because they were subgroup analyses (n=1) or extended follow-ups of RCTs described in other articles (n=5). One hundred ten articles were included in the
analysis. Fifty articles (45.5%) assessed nonpharmacological treatment: surgery (n=17), arthroscopic lavage (n=3), joint lavage (n=3), rehabilitation (n=23), education (n=2), spa therapy (n=1), and acupuncture (n=1). Rehabilitation interventions were related to physiotherapy (n=14), technical devices (n=5), and transcutaneous electrical nerve stimulation or laser therapy (n=4). A total of 60 articles (54.5%) assessed pharmacological treatment: oral drug administration (n=41); topical (n=3), intramuscular (n=1), and intravenous (n=1) treatments; and intra-articular injection (n=14).

Interrater reliability was good for random sequence generation (agreement, 92.7%; k=0.86; 95% confidence interval [CI], 0.77-0.95), allocation concealment (agreement, 89.1%; k=0.79; 95% CI, 0.67-0.90), patient blinding (agreement, 96.4%; k=0.92; 95% CI, 0.85-1.00), care provider blinding (agreement, 96.4%; k=0.93; 95% CI, 0.86-1.00), and outcome assessor blinding (agreement, 97.3%; k=0.90; 95% CI, 0.78-1.00). Determining whether the study was performed according to an intention-to-treat analysis resulted in a lower k value (agreement, 70.9%; k=0.59; 95% CI, 0.48-0.71). Interrater reliability assessed by use of the ICC was good for the Jadad scale (ICC, 0.84; 95% CI, 0.78-0.89), the Delphi list (ICC, 0.88; 95% CI, 0.83-0.92), and the numerical rating scale (ICC, 0.62; 95% CI, 0.49-0.72).

Only 10 articles (9.1%) were published in a general medical journal (JAMA, The Lancet, BMJ, Annals of Internal Medicine, and Archives of Internal Medicine). The other articles were published mainly in the Journal of Rheumatology (20.9%), Osteoarthritis and Cartilage (16.4%), Arthritis & Rheumatism (11.8%), Journal of Bone & Joint Surgery, British Volume (11.8%), Rheumatology (10.0%), and Annals of Rheumatic Disease (9.1%).

Financial support was totally or partially private in 57 articles (51.8%), public in 26 articles (23.6%), and not reported in 27 articles (24.6%). Pharmacological treatment funds were mainly private (75%) or not reported (23.3%), whereas in nonpharmacological treatment, funds were provided by public support in 25 articles (50.0%) and private support in 12 articles (24.0%) and were not reported in 13 articles (26.0%). Half of the articles concerned multicenter trials. Multicenter trials were reported more often in pharmacological treatment than in nonpharmacological treatment articles (68.3% vs 30.0%).

Items from the CONSORT diagram (flow of participants through each stage of the trial) were reported in the text or in a flowchart in 48 of the 75 articles published since the CONSORT statement was published in 1996.23 These data were reported in 71.4% (30/42) of the pharmacological treatment articles and 54.5% (18/33) of the nonpharmacological treatment articles and in only 9.1% (1/11) of the surgical articles.

Quality Assessment
Whatever tool was used for assessment, the quality scores were better for articles on pharmacological treatment than those on nonpharmacological treatment (FIGURE 2). On the Jadad scale, pharmacological treatment articles had a mean (SD) score of 7.5 (1.1) vs 4.9 (2.0) for nonpharmacological treatment articles. On the numerical rating scale, pharmacological treatment articles had a mean (SD) score of 7.0 (1.7) vs 4.9 (2.0) for nonpharmacological treatment articles. Moreover, as shown in Figure 2, reports of surgery/arthroscopy/lavage had the lowest quality scores, those of rehabilitation and intra-articular injection.

Figure 1. Study Screening Process

196 Articles Identified in MEDLINE and Cochrane Central Register of Controlled Trials
79 Articles Excluded on the Basis of the Title and Abstract
5 Abstracts Only
3 Duplicate Reports
12 Non-Randomized Controlled Trials
45 Reports Not Assessing a Therapeutic Intervention
13 Reports Not Assessing a Treatment of Hip or Knee Osteoarthritis
1 Phase 2 Clinical Trial
9 Articles Excluded After Obtaining the Full Text
5 Extended Follow-ups of Randomized Controlled Trials Described in Other Reports
1 Subgroup Analysis
50 Articles on Nonpharmacological Treatment
17 Surgery
3 Arthroscopic Lavage
3 Joint Lavage
23 Rehabilitation
14 Physiotherapy
5 Technical Device
4 Transcutaneous Electrical Nerve Stimulation or Laser Therapy
2 Education
1 Spa Therapy
1 Acupuncture
60 Articles on Pharmacological Treatment
41 Oral Drugs
3 Topical Treatments
1 Intramuscular Treatment
1 Intravenous Treatment
14 Intra-articular Injections

©2003 American Medical Association. All rights reserved.
had similarly low quality scores, and other pharmacological treatment articles had the highest scores. More information is available at [http://www.bichat.inserm.fr/fichierpdf/emi0357/dataonarticlesassessed.pdf](http://www.bichat.inserm.fr/fichierpdf/emi0357/dataonarticlesassessed.pdf).

**Trial Characteristics**

**Study Design.** The studies were all of parallel-group (n=104) or crossover (n=6) design. All of the selected articles involved randomization of patients; however, generation of randomization sequence was adequate in only 49.1% of the articles and was concealed from the investigators who enrolled the patients in only 20.9% (Table 1). There were no differences between nonpharmacological treatment and pharmacological treatment articles in the adequacy of random sequence generation. Adequate allocation concealment was similarly reported in nonpharmacological treatment and pharmacological treatment articles. However, allocation concealment was more often inadequate in nonpharmacological treatment articles than in pharmacological treatment articles, whereas allocation concealment was more often not reported in pharmacological treatment articles than in nonpharmacological treatment articles (Table 1).

**Description of Interventions. Reproducibility and Individualization.** The intervention was more often described with enough detail to be reproducible in pharmacological treatment than in nonpharmacological treatment articles (Table 1). Among nonpharmacological treatment articles, surgical treatments were less often considered reproducible (Table 2). The intervention was almost never individualized in pharmacological treat-

---

**Figure 2. Quality Assessment of Reports of Nonpharmacological and Pharmacological Treatment Trials**

Scores are based on a numerical rating scale, ranging from 1 to 10; the Delphi list,²² ranging from 0 to 9; and the Jadad scale,¹⁴ ranging from 0 to 5. High scores indicate high quality. Boxes represent median observations (horizontal rule) with 25th and 75th percentiles of observed data (top and bottom of box). In some instances, the median observation coincided with the 25th and 75th percentiles. Error bars represent the 10th and 90th percentiles.
ment articles, whereas one third of the nonpharmacological treatment articles described individualized interventions (Table 1). Finally, the technical quality of the nonpharmacological treatment was never evaluated in nonpharmacological treatment articles.

Care Providers. Care provider skill level and experience could influence the treatment effect, including all surgical interventions, in most of the nonpharmacological treatment articles but only in pharmacological treatment articles assessing intra-articular injection (Table 1 and Table 2). In contrast, care provider experience was reported in only 6 articles, hospital volume was reported in only 1 study, and the learning curve of care providers was never taken into account. Finally, care provider training before the beginning of the trial was mentioned in only 2 articles, one assessing intra-articular injection and the other assessing rehabilitation.

Compliance. We evaluated compliance with only repeated interventions (ie, treatment necessitating iterative interventions, such as drug treatment or physiotherapy), which comprised 57 (95.0%) of the pharmacological treatment articles and 26 (52.0%) of the nonpharmacological treatment articles. Among these articles, reporting of compliance was similar between pharmacological treatment articles (32 [56.1%] of 57) and nonpharmacological treatment articles (16 [61.5%] of 26). However, qualitative compliance was never evaluated in these articles. Compliance was assessed with at least 1 objective criterion in all pharmacological treatment articles (pill counts or drug dosage) vs in 7 (43.7%) of 16 nonpharmacological treatment articles (number of sessions attended for physiotherapy or a timer recording hours of use for transcutaneous electrical nerve stimulation).

Control Intervention. The control intervention was more often reported to be a placebo in at least 1 group in pharmacological treatment articles, whereas in nonpharmacological treatment articles, experimental treatments were more often compared with active control treatments or with usual care or waiting lists (Table 1). Surgical interventions were always compared with an active control intervention and never with a placebo (Table 2). The potential placebo effect of the different treatments being compared was considered to be similar more often in pharmacological treatment than in nonpharmacological treatment articles (Table 1).

Concomitant Treatments. The description of concomitant treatments (ie, additional care outside of the intervention provided to either comparison group) was reported more often in pharmacological treatment articles (58.3%) than in nonpharmacological treatment articles (24.0%). Contamination between groups was reported in only 3 articles.

Blinding. Patients were almost always reported to be blinded in pharmacological treatment articles, but only about one quarter of the nonpharmacological treatment articles described blinding (Tables 1). Care providers were reported to be blinded in 81.7% of the pharmacological treatment articles but were rarely blinded in nonpharmacological treatment articles (Table 1). When patients and care providers were reported to be blinded, care provider blinding was never tested, and patient blinding was tested in only 1 pharmacological treatment study and 1 nonpharmacological treatment study. When patients were not blinded, in only 2 nonpharmacological treatment articles were they instructed not to inform the outcome assessor about the treatment they received. Finally, outcome assessors were less often blinded

Table 1. Characteristics of Randomized Controlled Trials of Nonpharmacological and Pharmacological Treatments for Hip and Knee Osteoarthritis

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Nonpharmacological Treatment Trials (n = 50)</th>
<th>Pharmacological Treatment Trials (n = 60)</th>
<th>All Trials (n = 110)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation of allocation sequences</td>
<td>Adequate 27 (54.0)</td>
<td>27 (45.0)</td>
<td>54 (49.1)</td>
</tr>
<tr>
<td></td>
<td>Inadequate 2 (4.0)</td>
<td>0</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td></td>
<td>Not reported 21 (42.0)</td>
<td>33 (55.0)</td>
<td>54 (49.1)</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>Adequate 9 (18.0)</td>
<td>14 (23.3)</td>
<td>23 (20.9)</td>
</tr>
<tr>
<td></td>
<td>Inadequate 15 (30.0)</td>
<td>1 (1.7)</td>
<td>16 (14.5)</td>
</tr>
<tr>
<td></td>
<td>Not reported 26 (52.0)</td>
<td>45 (75.0)</td>
<td>71 (64.6)</td>
</tr>
<tr>
<td>Intervention reproducible</td>
<td>41 (82.0)</td>
<td>59 (98.3)</td>
<td>100 (90.9)</td>
</tr>
<tr>
<td>Individualization</td>
<td>15 (30.0)</td>
<td>2 (3.3)</td>
<td>17 (15.4)</td>
</tr>
<tr>
<td>Influence of care provider skill</td>
<td>42 (84.0)</td>
<td>14 (23.3)</td>
<td>56 (50.9)</td>
</tr>
<tr>
<td>Control intervention</td>
<td>Placebo 14 (28.0)</td>
<td>43 (71.7)</td>
<td>57 (51.8)</td>
</tr>
<tr>
<td></td>
<td>Active control treatment 28 (56.0)</td>
<td>16 (26.7)</td>
<td>44 (40.0)</td>
</tr>
<tr>
<td></td>
<td>Usual care or waiting list 8 (16.0)</td>
<td>1 (1.7)</td>
<td>9 (8.2)</td>
</tr>
<tr>
<td>Potential similar placebo effect</td>
<td>35 (70.0)</td>
<td>59 (98.3)</td>
<td>94 (85.4)</td>
</tr>
<tr>
<td>Blinding</td>
<td>Patients 13 (26.0)</td>
<td>58 (96.7)</td>
<td>71 (64.5)</td>
</tr>
<tr>
<td></td>
<td>Care providers 3 (6.0)</td>
<td>49 (81.7)</td>
<td>52 (47.3)</td>
</tr>
<tr>
<td></td>
<td>Outcome assessors 34 (68.0)</td>
<td>59 (98.3)</td>
<td>93 (84.5)</td>
</tr>
<tr>
<td>Analysis</td>
<td>Full ITT performed 12 (24.0)</td>
<td>21 (35.0)</td>
<td>33 (30.0)</td>
</tr>
<tr>
<td></td>
<td>Modified ITT performed 10 (20.0)</td>
<td>19 (31.7)</td>
<td>29 (26.4)</td>
</tr>
<tr>
<td></td>
<td>Per-protocol analysis performed 24 (48.0)</td>
<td>17 (28.3)</td>
<td>41 (37.3)</td>
</tr>
<tr>
<td></td>
<td>Unclear 4 (8.0)</td>
<td>3 (5.0)</td>
<td>7 (6.3)</td>
</tr>
<tr>
<td>Sample size justification or study power reported</td>
<td>15 (30.0)</td>
<td>42 (70.0)</td>
<td>57 (51.8)</td>
</tr>
</tbody>
</table>

Abbreviation: ITT, intention-to-treat analysis.
*All data are expressed as No. (%).
†All randomized participants were included in the analysis according to their original group assignment.
‡Analysis excluded participants who never received treatment or who were never evaluated while receiving treatment.
in nonpharmacological treatment articles than in pharmacological treatment articles (Table 1).

**Outcome Assessment.** In only 3 nonpharmacological treatment and 4 pharmacological treatment articles were outcome assessors trained, and outcomes were never reported to be assessed by an end-point review committee (an independent committee that ultimately decides whether a participant meets the criteria for a study’s outcome). Occurrence of adverse effects was reported more often in pharmacological treatment articles than in nonpharmacological treatment articles (83.3% vs 46.0%). Adverse effects were assessed by an independent committee in only 2 articles.

**Intention-to-Treat Analysis and Sample Size Justification.** Statistical analysis was performed according to an intention-to-treat principle (ie, all randomized participants were included in the analysis and kept in their original group) in only 30% of all articles (Table 1). A sample size justification or an estimation of the power of the study posteriori was reported more often in pharmacological treatment articles than in nonpharmacological treatment articles and was especially rare in surgical articles.

**Effect of Journals**
We analyzed the “journal effect” and found that articles published in general and internal medicine journals (n=10) and rheumatologic journals (n=36) had a higher quality score than articles published in orthopedic journals (n=36) and rehabilitation journals (n=7) according to the Jadad scale (mean [SD] score of 2.8 [1.3] vs 1.5 [1.4]), the Delphi list (7.0 [1.2] vs 5.5 [1.9]), and the numerical rating scale (6.7 [1.8] vs 5.1 [2.1]).

### Table 2. Characteristics of Selected Trials of Nonpharmacological and Pharmacological Treatments

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Nonpharmacological Treatment Trials</th>
<th>Pharmacological Treatment Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Surgery (n = 17)</td>
<td>Rehabilitation (n = 23)</td>
</tr>
<tr>
<td></td>
<td>Oral Drugs (n = 41)</td>
<td>Intra-articular Injection (n = 14)</td>
</tr>
<tr>
<td>Randomization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generation of allocation sequences</td>
<td>Adequate (5 [29.4])</td>
<td>Adequate (29 [46.1])</td>
</tr>
<tr>
<td></td>
<td>Inadequate (1 [5.9])</td>
<td>Inadequate (1 [4.4])</td>
</tr>
<tr>
<td></td>
<td>Not reported (11 [64.7])</td>
<td>Not reported (7 [30.4])</td>
</tr>
<tr>
<td>Allocation concealment</td>
<td>Adequate (0)</td>
<td>Adequate (6 [26.1])</td>
</tr>
<tr>
<td></td>
<td>Inadequate (8 [47.1])</td>
<td>Inadequate (5 [21.7])</td>
</tr>
<tr>
<td></td>
<td>Not reported (9 [52.9])</td>
<td>Not reported (12 [52.2])</td>
</tr>
<tr>
<td>Intervention reproducible</td>
<td>10 [58.8]</td>
<td>22 [95.6]</td>
</tr>
<tr>
<td>Individualization</td>
<td>3 [17.6]</td>
<td>11 [47.8]</td>
</tr>
<tr>
<td>Influence of care provider skill</td>
<td>17 [100]</td>
<td>16 [69.6]</td>
</tr>
<tr>
<td></td>
<td>0 [0]</td>
<td>0 [0]</td>
</tr>
<tr>
<td>Control intervention</td>
<td>Placebo (0)</td>
<td>10 [43.5]</td>
</tr>
<tr>
<td></td>
<td>Active control treatment (17 [100])</td>
<td>9 [39.1]</td>
</tr>
<tr>
<td></td>
<td>Usual care or waiting list (0)</td>
<td>4 [17.4]</td>
</tr>
<tr>
<td>Potential similar placebo effect</td>
<td>17 [100]</td>
<td>14 [60.9]</td>
</tr>
<tr>
<td>Blinding</td>
<td>Patients (4 [23.5])</td>
<td>6 [26.1]</td>
</tr>
<tr>
<td></td>
<td>Care providers (0)</td>
<td>2 [8.7]</td>
</tr>
<tr>
<td></td>
<td>Outcome assessors (7 [41.2])</td>
<td>18 [75.3]</td>
</tr>
<tr>
<td>Analysis</td>
<td>Full ITT performed†</td>
<td>6 [26.1]</td>
</tr>
<tr>
<td></td>
<td>Modified ITT performed‡</td>
<td>15 [36.6]</td>
</tr>
<tr>
<td></td>
<td>Per-protocol analysis performed</td>
<td>6 [34.1]</td>
</tr>
<tr>
<td></td>
<td>Unclear (2 [11.8])</td>
<td>12 [52.2]</td>
</tr>
<tr>
<td></td>
<td>Sample size justification or study</td>
<td>10 [43.5]</td>
</tr>
<tr>
<td></td>
<td>power reported (1 [5.9])</td>
<td>29 [70.7]</td>
</tr>
</tbody>
</table>

Abbreviation: ITT, intention-to-treat analysis.

*All data are expressed as No. (%).
†All randomized participants were included in the analysis according to their original group assignment.
‡All randomized participants were included in the analysis according to their original group assignment.

### COMMENT
This study assessed the methodological quality of all RCTs published on the topic of hip and knee osteoarthritis during a 10-year period in high-impact general medical and specialty journals. Several studies have assessed the methodological quality of a broad range of reports of randomized trials in several areas of health care. Whatever the domain assessed, the overall quality of published RCTs is poor. Methodological problems concerning randomization and intention-to-treat analysis are common. Moreover, results of these studies showed that such methodological deficiencies could influence the effect size. Inadequate random sequence generation and lack of allocation concealment and double-blinding yielded larger treatment effects. However, to our knowledge, no study has compared the methodological quality of nonpharmacological treatment and pharmacological treatment articles. We focused on the methodological quality of articles of RCTs assessing hip and knee osteoarthritis treatments because these treatments cover a wide range of nonpharmacological treatment (eg, surgery, arthroscopy, joint lavage, exercise therapy, physiotherapy, orthosis, spa therapy, acupuncture, and education) and pharmacological treatment (eg, oral drug administration, intra-articular injection).

Our analysis showed that the methodological quality of reports of nonpharmacological treatment trials was lower than that of reports of pharmacological treatment trials. Methodological deficiencies such as lack of reporting adequate random sequence generation, adequate allocation concealment, and...
intention-to-treat analysis were common in both nonpharmacological treatment and pharmacological treatment articles. These deficiencies could be reduced easily in the trials themselves (through changes in conduct when necessary) and in the reporting of the trials. However, some specific issues, such as choice of control intervention, lack of double-blinding, care provider effect, and standardization problems, could be difficult to resolve when assessing nonpharmacological treatment.

Control Intervention
When assessing nonpharmacological treatment, a placebo or sham intervention can be difficult or impossible to perform for ethical or technical reasons. Ethical concerns are substantial in trials assessing surgical procedures when sham interventions are prohibitive. In the trials investigated herein, surgery was always compared with another surgical procedure and never with a placebo. Recently, Moseley et al pointed out the difficulties of performing placebo-controlled trials of surgery when evaluating arthroscopic intervention for knee osteoarthritis. In that trial, the control group underwent simulated arthroscopic surgery, where small incisions were made but no instruments were inserted. Practical issues are also important, and implementing a sham intervention often requires creative solutions. For example, to perform double-blind placebo-controlled trials assessing acupuncture, Streitberger and Kleinhenz developed a placebo acupuncture needle that gave patients the feeling of penetration but did not penetrate the skin. Thus, it appears important to develop research in this domain.

Blinding
Blinding patients and care providers is usually possible in drug trials with a matching placebo but is often impossible to perform in nonpharmacological treatment trials. Although some investigators have proposed solutions such as use of standardized wound dressings when assessing laparoscopic cholecystectomy, surgeons usually know which intervention has been done and patients usually know which rehabilitation program they have undergone.

To avoid these biases, efforts could be made to blind outcome assessors. Use of the Prospective Randomized Open Blinded End-point (PROBE) study design, which is based on blinded end points, could be an alternative. However, treatments of hip or knee osteoarthritis are usually assessed from patient-reported symptoms, and if patients are not blinded, outcome assessors cannot be blinded. Results of several studies that did not involve double-blinding yielded exaggerated estimates of treatment effects, while results of other studies showed no effect. Heterogeneity in who was blinded and in the outcomes assessed may be responsible for these discrepancies. Blinding is particularly important when outcome measures involve patient-reported symptoms such as pain but is less important for objective criteria such as death because of little detection bias.

Standardization and Care Providers’ Effect
Nonpharmacological treatments are usually complex; that is, they include several components and/or several health care professionals and are often individualized. The active component of such interventions is therefore difficult to identify, and the treatments are difficult to replicate. A detailed standardization of the intervention is necessary, and the technical quality of the intervention should be evaluated. Contrary to pharmacological treatment, in which the effect of health care professionals can generally be regarded as secondary, in nonpharmacological treatment the health care professional is an integral part of the intervention. The success of the intervention depends on care provider skill, experience, and training. Variation in care provider skills in each arm of a trial can be confounded with the treatment effect. Most nonpharmacological treatment, especially surgery, involves complex procedures, and quality in performance requires frequent repetitions. No trial evaluated in our study assessed the learning curve, and only 6 articles described care provider experience.

Surgery
Among nonpharmacological treatment articles, surgical articles had the lowest-quality scores. Some methodological deficiencies, such as low reporting of data in the CONSORT diagram, adequate allocation concealment, and detailed description of the intervention, should not be more difficult to resolve than in other nonpharmacological treatment trials. However, surgical trials also present specific issues. In these trials, care providers can always influence the treatment effect and are never blinded. In the articles analyzed, surgical interventions were never compared with a placebo, probably for ethical reasons. For example, patients included in trials assessing surgical interventions that are irreversible will not have an opportunity to benefit from results of the trial.

Limitations of This Study
Our study was limited because we assessed only the reports of RCTs, not the trials themselves. However, failure to report the methods of a trial does not necessarily mean that investigators did not carry out these methods. Some methodological deficiencies may lie in the reporting of trials rather than in their performance.

This study allowed us to point out the difficulties in assessing nonpharmacological treatment. Some improvement could occur with use of adequate random allocation generation and concealment and intention-to-treat analyses. However, methodological issues concerning choice of a control intervention; blinding of patients, care providers, and outcome assessors; standardization of interventions; and taking into account care provider skill are more difficult to resolve. Finally, tools used to measure quality, such as the Jadad scale and the Delphi list, are probably not appropriate for assessing
DIFFERENCES IN TRIALS OF OSTEOARTHRITIS TREATMENTS

the quality of nonpharmacological treatment articles because they focus on randomization, blinding, and intention-to-treat analysis. Nonpharmacological treatment represents a wide range of treatments available to patients, it is important to enhance research in this domain and to develop specific tools appropriate but no less rigorous for assessing nonpharmacological treatment trials. Moreover, despite the challenges posed by nonpharmacological treatment studies, the same expectations of quality should be applied to nonpharmacological treatment trials as are applied to pharmacological treatment trials.

REFERENCES


Author Contributions: Study concept and design: Boutron, Tubach, Giraudobe, Ravaud. Acquisition of data: Boutron, Tubach. Analysis and interpretation of data: Boutron, Tubach, Giraudobe, Ravaud. Drafting of the manuscript: Boutron. Critical revision of the manuscript for important intellectual content: Tubach, Giraudobe, Ravaud. Statistical expertise: Boutron, Tubach, Giraudobe. Obtained funding: Boutron.

Study supervision: Tubach, Giraudobe, Ravaud.

Funding/Support: This work was supported by a grant from the Caisse Nationale d’Assurance Maladie des Professions Indépendantes (CANAM), St Denis, France, and the Assistance Publique des Hôpitaux de Paris.