

Neglected external validity in reports of randomized trials: The example of hip and knee osteoarthritis.

Nizar Ahmad, Isabelle Boutron, David Moher, Isabelle Pitrou, Carine Roy,
Philippe Ravaud

► **To cite this version:**

Nizar Ahmad, Isabelle Boutron, David Moher, Isabelle Pitrou, Carine Roy, et al.. Neglected external validity in reports of randomized trials: The example of hip and knee osteoarthritis.. *Arthritis and Rheumatism*, Wiley, 2009, 61 (3), pp.361-9. 10.1002/art.24279 . inserm-00372070

HAL Id: inserm-00372070

<https://www.hal.inserm.fr/inserm-00372070>

Submitted on 15 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

External validity is neglected in reports of randomized trials

The example of hip and knee osteoarthritis

Nizar Ahmad MD¹, Isabelle Boutron MD, PhD¹, David Moher PhD², Isabelle Pitrou MD¹,
Carine Roy¹, Philippe Ravaud MD, PhD¹

¹ INSERM, U738, Paris, France; AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Bichat-Claude Bernard, Département d'Epidémiologie Biostatistique et Recherche Clinique, Paris, France; Université Paris 7, Faculté de Médecine, Paris, France.

² Chalmers Research Group, Ottawa Health Research institute, Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Canada.

Address correspondence and reprint requests to Isabelle Boutron, Département d'Epidémiologie Biostatistique et Recherche Clinique, INSERM U738, Groupe Hospitalier Bichat-Claude Bernard, 46 rue Henri Huchard, 75018 Paris, France.

Tel: +33 01 40 25 79 43

Fax: +33 01 40 25 67 73

E-mail: isabelle.boutron@bch.ap-hop-paris.fr

Word count: 3 793

ABSTRACT

Objective: To evaluate the reporting of data related to external validity (i.e., applicability of the results of a trial in routine practice) from randomized controlled trials (RCTs) assessing pharmacologic treatment and nonpharmacologic treatment for hip and knee osteoarthritis.

Methods: All RCTs assessing pharmacologic treatments (e.g., oral drugs, intra-articular injection, topical treatment) and nonpharmacologic treatments (e.g., surgery, rehabilitation, education, joint lavage, nonimplantable devices) for hip and knee osteoarthritis indexed between January 2002 and December 2006 were selected. A sample of 120 articles was randomly selected, 30 each of trials assessing pharmacologic treatments, surgery or technical interventions, rehabilitation and nonimplantable devices.

Results

The country where the trial took place was clearly reported in only 21% of reports (n = 25). The setting was described in one-third of reports (n = 40) and the number of centres in 45% (n = 54). Details of centres (volume of care) were given in 20% of reports (n = 24). Reporting rates were lower for surgical trials for the country (3%), the setting (3%), the number of centres (13%), and details of centres (volume of care) (7%). The intervention was adequately described in all pharmacologic reports and more than 80% of reports of trials assessing rehabilitation. In reports of surgical intervention trials, the technical procedure was given in all reports, but the type of anaesthesia was reported in 13% (n = 4), pre-operative care in 7% (n = 2), and post-operative care in 50% (n = 15). In reports of device trials, the device was described in 93%, but the manufacturer was reported in only one-third of the reports. The method of recruitment was reported in 36% of reports (n = 43). Eligibility criteria were described in most reports (98%), but 22.9% of the exclusion criteria were poorly justified.

Conclusion: This study highlights the low reporting lack of data related to external validity in reports of RCTs assessing pharmacologic and nonpharmacologic treatments for hip and knee osteoarthritis.

BACKGROUND

Well-conducted randomized controlled trials (RCTs) are adopted as the gold standard for evaluating medical interventions (1-4). For results to be clinically useful, RCTs must take into account the internal validity (i.e., the extent to which systematic errors or bias are avoided) and the external validity (sometimes called applicability - i.e., whether the results of a trial can be reasonably applied or generalized to a definable group of patients in a particular setting in routine practice) (5, 6).

Historically, internal validity has been considered a priority for research. Several publications identified methods to avoid bias(7, 8). The Consolidated Standards of Reporting Trials (CONSORT) Statements, endorsed by many major medical journals, improved on the reporting of data related to internal validity(1, 9). Tools(10-13) have been developed mainly to evaluate internal validity in reports of trial results included in systematic reviews(14).

Funders and journals have tended to be more concerned with the scientific rigor of interventions studied than with the applicability of the results. Consequently, external validity has been frequently neglected(6, 15-17). This neglect has probably contributed to the failure to translate research into clinical practice. Lack of external validity is frequently advocated as the reason why interventions found to be effective in clinical trials are underused in clinical practice(5). However, assessing the external validity of a trial to turn research into action supposes that information is adequately reported in published articles. Further, as highlighted by the extension of the CONSORT Statements to nonpharmacologic treatment, assessing external validity is probably more difficult for trials assessing nonpharmacologic treatments (e.g., surgery, technical interventions, rehabilitation, psychotherapy, devices) than pharmacologic treatments (e.g., oral drugs)(18, 19).

The aim of this study was to evaluate and compare the reporting of external validity from RCTs assessing pharmacologic and nonpharmacologic treatments for hip and knee osteoarthritis (OA). We chose these conditions because they are highly prevalent and can result in disability and reduced quality of life. Further, international guidelines require the use of a combination of pharmacologic and nonpharmacologic treatments for optimal management of patients with these conditions(20, 21).

METHODS

Search strategy and selection of reports

We identified all English-language reports of RCTs indexed between January 2002 and December 2006 in PubMed using the search terms “Osteoarthritis Hip” OR “Osteoarthritis Knee,” with a limitation to RCTs, in MEDLINE via PubMed and to articles published in English. A similar search strategy was used in a previous work on internal validity (22).

Eligibility criteria and screening process

We collected the electronic records in an Endnote data file. One of us (AN) assessed each reports by screening the titles and abstracts to identify relevant studies. A second person (IB) checked for adequate selection of the abstracts. Articles were included if the study was identified as an RCT assessing pharmacologic or nonpharmacologic treatment for hip or knee OA in a parallel-group or cross-over design. We excluded reports of cluster RCTs, nonrandomised trials, observational studies (cohort and case-control studies), extended follow-up trials (i.e., extended follow-up of patients included in an RCT beyond the last outcome assessment), nontherapeutic trials (metrological studies, epidemiological studies), pathophysiological studies, letters, an ancillary study of an RCT such as a subgroup analysis, cost-effectiveness evaluations, systematic reviews and/or meta-analyses. We also excluded reports of trials assessing the organization of the health care system or interventions provided to care providers. We excluded reports with these designs because we wanted to have a relatively homogeneous sample.

The selected abstracts were classified according to the category of treatment assessed: pharmacologic treatments, surgery or technical interventions (e.g., joint lavage), rehabilitation, or nonimplantable devices.

For each category of treatment, we used a computer-generated list to randomly select 30 articles and retrieved the full text articles. Articles not fulfilling the inclusion criteria were replaced by a random selection of articles in the corresponding category. We chose a total of 120 articles for practical reasons, mainly to provide enough articles describing each category of treatment, and randomly selected articles to avoid selection bias.

Data collection

To assess external validity, as well as internal validity of the selected reports, we reviewed the literature and generated a standardized data extraction form (available upon request); we used items related to external validity proposed by the CONSORT Statement for RCTs(1), the extension of the CONSORT Statement for nonpharmacologic trials(18, 19) and Rothwell et al.(5) Before data extraction, as a calibration exercise, the standardized form was tested independently by two members of the team (AN, IB) on a separate set of 20 reports. A meeting followed in which the ratings were reviewed and any disagreements were resolved by consensus. One reviewer (AN) independently completed all the data extraction. A random sample of 20 articles was reviewed for quality assurance.

The data extraction form covered the following data:

1) Characteristics of the selected studies: year of publication, journal, medical area of the study (i.e., hip OA, knee OA, hip and knee OA), type of treatment (i.e., pharmacologic treatment, surgical intervention, rehabilitation or education, nonimplantable device), type of control intervention (i.e., active intervention, placebo, usual care), funding sources (i.e.,

public, private, both, no funding, not reported or unclear), study design (i.e., parallel-group, cross-over) and sample size.

Internal validity of the selected reports was assessed with use of the following specific criteria recommended by the Cochrane collaboration and by most quality tools assessing results of pharmacologic and nonpharmacologic trials(10, 12): allocation sequence generation; allocation concealment; blinding of patients, care providers and outcome assessors; and intent-to-treat (ITT) analysis.

2) External validity

The reporting of the following data related to external validity was evaluated:

Recruitment: data on the method of recruitment (i.e., referral from a rheumatologist or general physician, self-selection of patients through advertisement) and duration of recruitment.

Patients: data on patients' eligibility criteria as defined in a previous work (23), inclusion criteria (i.e., criteria governing entry or recruitment of individuals into the trial and describing the medical conditions of interest) and exclusion criteria (all other criteria limiting the eligibility of individuals)(23). The exclusion criteria were classified as strongly justified, potentially justified and poorly justified reasons for excluding individuals from an RCT according to the classification proposed by Van Spall et al.(23); exclusion criteria were considered strongly justified if an individual or substitute decision-maker was unable to grant informed consent, the intervention or placebo would likely be harmful, the intervention would likely be ineffective, or the effect of the intervention would be difficult to interpret.

Data on the number of eligible patients, number of patients not meeting inclusion criteria and number of patients refusing to participate were collected. We also checked whether the article

reported baseline characteristics of excluded patients, as well as essential data on baseline characteristics of randomized patients (i.e., age, sex, weight/body mass index, ethnicity, coexisting diseases or co-morbidities, duration of the disease, measure of function status, level of pain, description of radiographic evidence of damage, and use of nonsteroidal anti-inflammatory drugs).

Centre and care provider: data on the number of centers/care providers, expertise of centers/care providers and details of the centers (name, sources, organization, and expertise). The reporting of the number of patients recruited in each center or by each care provider was recorded.

Intervention (i.e., whether and how details on the interventions were reported): for pharmacologic treatments, the route of administration, dosage, duration, frequency of treatment, and patient compliance; for rehabilitation, the number, timing, duration and content of each session, mode of delivery, supervision or not, and patient compliance; for surgical interventions, the type of anaesthesia, pre-operative care, post-operative care, description of the technical procedure and surgeons' compliance with the planned procedure; and for nonimplantable devices, the reporting of the manufacturer, description of the devices and patient compliance.

Abstract and discussion sections: information related to external validity reported in the abstract (i.e., country where the trial took place, setting, number of centers, number of eligible patients, number of patients randomized, length of recruitment, length of follow-up, and data on care providers) and whether the external validity was discussed in the discussion section of the study as recommended by the CONSORT Statement (1).

Global assessment of external validity: quantitative assessment of external validity reporting may offer complementary information. Although it is difficult to specify which aspect of external validity is the most important, we decided to focus on 3 important components that are probably indispensable to assess the external validity of a trial: 1) the participants, 2) the description of the experimental treatment, and 3) the context of care (centres, setting, care providers' expertise). For each component, we identified items that were considered essential to allow an adequate assessment of the external validity of a published trial. These items are described in the box. The quantitative assessment of external validity was evaluated by the percentage of the selected items that were adequately reported, for each component.

Statistical analysis

All data analyses were completed using SAS for Windows, Release 9.1 (SAS Institute, Cary, NC).

We used descriptive statistics for continuous variables: means, standard deviation (SD), median (lower quartile; upper quartile) and minimum and maximum values. Categorical variables are described with frequencies and percentages. The results were adjusted for the clustering effect by journals as recommended (24). The reporting of data related to external validity according to category of treatment was compared by a linear mixed-effects model, with the percentage of items with external validity as the dependent variable, fixed effects for treatment category, and journal as a random effect.

RESULTS

Articles selected

We identified 388 citations from our electronic search, of which 123 were excluded. Among the 265 included reports, we randomly chose 120 reports, 30 for each category of treatment. After obtaining and reviewing the full text, 11 articles were replaced. The flow of articles through the study is presented in appendix 1.

Characteristics of the selected studies

Characteristics of the included studies are reported in table 1. The 120 articles were indexed in 53 journals. Among them, 13 (11%) were published in a general medical journal with a high impact factor and 107 (89%) in a general medical journal with low impact factor or in a special medical journal. Most trials, 118 (98%), had a parallel-group design. Three-quarters of the reports assessed knee OA (n = 90). The source of funding was described as public in 45 articles (38%), completely or partially private in 25 (21%). A funding source was not reported in 50 reports (42%).

The median sample size (interquartile range) was 100 (60-216) and was twice as high for reports of pharmacologic trials as nonpharmacologic trials.

The control group was described as receiving active treatment in 63 reports (52%), a placebo intervention in 43 (36%) and usual care in 14 (12%). Pharmacologic treatments and nonimplantable devices were mainly compared to placebo or active treatments, whereas rehabilitation interventions were mainly compared to usual care or active treatments, and surgical procedures were compared to active treatment in most reports.

The generation of allocation sequences was adequate in 51% of the reports. The treatment allocation was adequately concealed in 41% of reports (n = 49). Blinding was

reported and was adequate in 43% of reports for patients, 32% for care providers and 59% for outcome assessors. An ITT analysis was described in only one-third of the reports.

External validity

The results for assessing external validity are reported in tables 2 and 3 and figure 1.

Recruitment (table 2)

The method of recruitment was described in about one-third of the reports; when described, this method relied on referral in 67% of reports (n = 29/43) and self-selection in 33% (n = 14/43). The duration of recruitment was described in 47% of reports (n = 56); reporting was better in articles about rehabilitation. The median (interquartile range [IQR]) duration of recruitment for 10 patients per month described was 0.4 [0.2-0.8] months for pharmacologic trials; 0.8 [0.3-1.9] for device trials; 1.2 [0.9-2.7] for rehabilitation trials; and 2.5 [1.1-4.4] for surgical trials.

Participants (table 2)

Inclusion criteria were described in almost all reports (98%, n = 118) and exclusion criteria in 88% (n = 106). Exclusion criteria focused on age in 53% of reports (n = 64), medical co-morbidities in 66% (n = 79), sex in 14% (n = 17), medication in 48% (n = 57), socioeconomic status in 2% (n=3), and patients participating in another trial in 5% (n=6).

Twenty-three percent of reports poorly justified exclusion criteria. These rates did not differ by category of treatment.

A flow diagram of participants through the trial was given in 40% of reports (n = 48). Data related to the number of eligible participants, number of participants not meeting inclusion criteria and those refusing participation were reported in less than half of the reports, but reporting was better for rehabilitation trials. When given, the mean rate of participants not

meeting inclusion criteria and refusing to participate was 22.5 (30%) and 19.2 (16%), respectively.

Baseline data of excluded participants were given in only one report. The baseline clinical characteristics of randomized participants were described in 91% of reports (n = 109); characteristics concerned age and sex in more than 80% of reports, weight or body mass index in 62%, and severity of disease (i.e., duration of the disease, pain, function, radiographic evidence of damage) in less than half. Patients' comorbidities were provided in only 12% of reports.

Interventions (table 3)

The treatments were described according to the CONSORT recommendations in all reports of pharmacologic trials and in most reports of rehabilitation trials but were missing in reports of devices and surgery trials. In the reports of medical device trials, a description of the device was given in 93% of reports (n = 28), but the manufacturer was stated in only 30% (n = 9). In the reports of surgical intervention trials, the technical procedure was given in all reports, but the type of anaesthesia was reported in only 13% (n = 4), pre-operative care in 7% (n = 2), and post-operative care in 50% (n = 15). Control treatment was described in most reports (98%, n = 117). Descriptions of co-interventions were lacking in 23% of reports (n = 28), mainly reports of pharmacologic trials.

Centres and care providers (table 2)

The setting was described in 33% of reports (n = 40) and the number of centres in 45% (n = 54). The country where the trial took place was clearly reported in only 21% of reports (n = 25). Details of centres were given in 20% of reports (n = 24). Other details such as centre sources, organization and expertise were never reported. The number of participants recruited

in each centre was never reported. Details on care providers were given in 29% of reports (n = 35).

Abstract and discussion sections

Information related to external validity was provided in the abstract of reports as follows: 4% of articles (n = 5) described the country where the trial took place, 15% (n = 18) the setting, 12% (n = 14) the number of centres, 2% (n = 2) the number of eligible patients, 92% (n = 110) the number of patients randomized, 5% (n = 6) the length of recruitment, 82% (n = 98) the length of follow-up, and 2% (n = 2) data on care providers. External validity was discussed in the discussion section of 11 articles (9%).

Global assessment of external validity (figure 1)

Figure 1 highlights the reporting of each component of external validity by category of treatment. Reporting of essential baseline characteristics items was lower in reports of surgical trials, with a median [IQR] of 30% [30-40] of the essential items reported, than for those of trials of pharmacologic treatments, nonimplantable devices and rehabilitation, with a median [IQR] of 50% [40-60], 50% [30-60] and 45% [30-60], respectively, of the essential items reported (p= 0.006).

The reporting of the intervention was better in reports of trials of pharmacologic treatments and rehabilitation (median [IQR] 80% [80-100] and 86% [71-100], respectively) than for those of trials of nonimplantable devices and surgery (median [IQR] 33% [33-67] and 40% [20-40]), respectively; p <0.001).

The items dedicated to the context of the trial were poorly reported for trials of all treatments, especially pharmacologic treatments and surgery (median [IQR] 12% [12-25] and 25% [12-25]), respectively; p = 0.016).

DISCUSSION

This study assessed the reporting of external validity in a sample of 120 RCTs assessing pharmacologic and nonpharmacologic treatments for hip or knee OA during a 5-year period. Our results highlight the lack of data related to external validity in published reports of RCTs. Methods for recruiting patients were described in one-third of the reports; 22.9% of the exclusion criteria were poorly justified; important baseline data of patients were lacking; and setting, centers and care providers were described in one-third and less of articles. Further, the reporting of external validity differed depending on the category of treatment. Reports of trials assessing rehabilitation provided more adequate data related to recruitment, participants, setting and centers, and intervention. On the contrary, reports of trials assessing surgical procedures lacked such data, even though the reporting of some items, such as the setting, the number of centers and center volume, is particularly important in this field. In reports of pharmacologic trials and trials assessing nonimplantable devices, the reporting was of varying quality. In reports of pharmacologic trials, the reporting of the method of recruitment and of data related to centers and care providers was poor, but the reporting of the intervention was good.

To our knowledge, this is the first study that systematically appraised the reporting of data related to external validity from trials assessing pharmacologic and nonpharmacologic treatments. Most recent efforts of researchers and editors to improve the reporting of results of RCTs, such as the CONSORT initiative, have mainly focused on internal validity (1, 9). Nevertheless, external validity is also essential and needs to be emphasized (25, 26). The results of RCTs and systematic reviews cannot be relevant to all patients and all settings.

Consequently, reporting the results of RCTs should allow clinicians to judge to whom and in which context these results could reasonably be applied.

The setting, care providers and centers have obvious implications for external validity(5, 27). In fact, the applicability of results of trials performed in secondary or tertiary settings applied to primary settings is often a concern (5). Further, differences between health care systems can affect the applicability of results, especially regarding organization of care or reimbursement for the cost of care(5). These issues are crucial in trials assessing nonpharmacologic treatments such as surgery or technical interventions. In fact, hospital and care providers' volume and outcome are related (28-33). A surgical procedure might be found to be safe and effective in an RCT performed in high-volume centres by high-volume care providers, but applying these results to low-volume centres might result in very different results (27, 34, 35). Surprisingly the reporting of data on care providers and centers was far less than optimal in our study, especially for trials assessing surgical procedures.

The representativeness of patients included in a RCT is also a major issue for external validity. The inclusion and exclusion criteria are among the greatest challenges in achieving representativeness of participants. Highly selective eligibility criteria can considerably reduce the applicability of the trial results. Our results highlight the lack of reporting of exclusion criteria in 12% of the trial reports; 23% of reported exclusion criteria were poorly justified. These results are consistent with those of a systematic review of RCTs published in high-impact-factor journals between 1994 and 2006(23). Exclusion criteria reported in our articles concerned mainly elderly patients, those with medical comorbidities or those treated with specific categories of treatments. The exclusion of these specific categories of participants is problematic because it limits the representativeness of the patients.

The representativeness of the participants is also problematic because those agreeing to participate in RCTs often differ from those who do not participate. (36-39). Consequently, the number of eligible nonrandomized patients, as well as the number of participants who were invited to participate but declined, is important to adequately appraise the external validity of a trial(22); however, these data were reported in only one-third and one-quarter, respectively, of our reports, which is consistent with previous results (40).

Reporting the baseline clinical characteristics of participants included in RCTs should allow clinicians and others to assess external validity by comparison with their patients. Although baseline characteristics were described in almost all of our reports, some important data were missing: weight or body mass index, while essential, was given in only 62% of the selected articles. The reporting of ethnicity, comorbidities and severity and activity of the disease (pain, function, radiographic evidence of damage), which also predict response to and influence the generalisability of treatment was also inadequately reported (41-44).

External validity could also be affected if trials have treatment protocols that differ from usual clinical practice or have too stringent limitations on the use of cointerventions. Further, to be able to adequately apply the results of the trial in clinical practice, the treatments should be described in detail to allow for adequate reproducibility. Our results highlight the lack of descriptions of nontrial treatments in two-thirds of the reports of pharmacologic trials and that descriptions of all the components of nonpharmacologic trials were lacking, especially in reports of surgery (45).

Finally, despite a specific item of the CONSORT Statement dedicated to external validity, very few articles considered this issue in the discussion section.

Our study has several limitations. First, we focused on the reporting of the trial, not its conduct. Consequently, these results highlight the lack of adequate reporting of external validity criteria and do not provide information on the applicability of the results of the trial. Second, the results related to the rate of poorly justified exclusion criteria might be underestimated. In fact, some researchers have highlighted the inadequate reporting of eligibility criteria when comparing the published article with the protocol(46); among an average of 31 eligibility criteria, only 63% were described in the main trial reports. Third, we focused on RCTs assessing hip and knee OA, and these results should be confirmed in other medical areas. However, we chose this disease because it is frequent and involves a wide range of pharmacologic and nonpharmacologic treatments. Further, the authors had some expertise in rheumatology and orthopedics and could therefore adequately evaluate the context of the trials.

In conclusion, this study highlights the lack of consideration of external validity in published reports of RCTs. Much attention is paid to the internal validity of clinical trials; however, even results of well-designed clinical trials are of limited use to clinicians if they have poor external validity and are not applicable to the patients for whom the intervention is designed. Recently the CONSORT group developed an extension of the CONSORT Statements for pragmatic trials. This extension increases the focus on data related to external validity. This initiative should help improve the consideration of external validity.

REFERENCES

1. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134(8):663-94.
2. Schulz KF, Grimes DA. Blinding in randomised trials: hiding who got what. *Lancet* 2002;359(9307):696-700.
3. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet* 2002;359(9306):614-8.
4. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet* 2002;359(9305):515-9.
5. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet* 2005;365(9453):82-93.
6. Glasgow RE, Green LW, Klesges LM, Abrams DB, Fisher EB, Goldstein MG, et al. External validity: we need to do more. *Ann Behav Med* 2006;31(2):105-8.
7. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352(9128):609-13.
8. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Jama* 1995;273(5):408-12.
9. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185(5):263-7.
10. Boutron I, Moher D, Tugwell P, Giraudeau B, Poiraudau S, Nizard R, et al. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58(12):1233-40.
11. Higgins JPT., Altman DG. Assessing risk of bias in included studies. In: Higgins JPT., Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.0* (updated February 2008). The Cochrane Collaboration. Available from www.cochrane-handbook.org; 2008.
12. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12.
13. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998;51(12):1235-41.
14. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet* 1999;354(9193):1896-900.
15. Bath FJ, Owen VE, Bath PM. Quality of full and final publications reporting acute stroke trials: a systematic review. *Stroke* 1998;29(10):2203-10.
16. Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the *Journal of Bone and Joint Surgery* from 1988 through 2000. *J Bone Joint Surg Am* 2002;84-A(3):388-96.
17. Dziewaltowski DA, Estabrooks PA, Klesges LM, Bull S, Glasgow RE. Behavior change intervention research in community settings: how generalizable are the results? *Health Promot Int* 2004;19(2):235-45.

18. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Methods and Processes of the CONSORT Group: Example of an Extension for Trials Assessing Nonpharmacologic Treatments. *Ann Intern Med.* 2008(148):W60-W67.
19. Boutron I, Moher D, Altman DG, Schulz K, Ravaud P, for the CONSORT group. Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med.* 2008(148):295-309.
20. Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2005;64(5):669-81.
21. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2003;62(12):1145-55.
22. Boutron I, Tubach F, Giraudeau B, Ravaud P. Methodological differences in clinical trials evaluating nonpharmacological and pharmacological treatments of hip and knee osteoarthritis. *Jama* 2003;290(8):1062-70.
23. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *Jama* 2007;297(11):1233-40.
24. Hewitt C, Hahn S, Torgerson DJ, Watson J, Bland JM. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. *Bmj* 2005;330(7499):1057-8.
25. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. *Bmj* 2006;333(7563):346-9.
26. Glasgow RE, Bull SS, Gillette C, Klesges LM, Dziewaltowski DA. Behavior change intervention research in healthcare settings: a review of recent reports with emphasis on external validity. *Am J Prev Med* 2002;23(1):62-9.
27. Moore WS, Young B, Baker WH, Robertson JT, Toole JF, Vescera CL, et al. Surgical results: a justification of the surgeon selection process for the ACAS trial. The ACAS Investigators. *J Vasc Surg* 1996;23(2):323-8.
28. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002;137(6):511-20.
29. Hodgson DC, Zhang W, Zaslavsky AM, Fuchs CS, Wright WE, Ayanian JZ. Relation of hospital volume to colostomy rates and survival for patients with rectal cancer. *J Natl Cancer Inst* 2003;95(10):708-16.
30. Khuri SF, Daley J, Henderson W, Hur K, Hossain M, Soybel D, et al. Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. *Ann Surg* 1999;230(3):414-29; discussion 429-32.
31. Lavernia CJ, Guzman JF. Relationship of surgical volume to short-term mortality, morbidity, and hospital charges in arthroplasty. *J Arthroplasty* 1995;10(2):133-40.
32. McGrath PD, Wennberg DE, Dickens JD, Jr., Siewers AE, Lucas FL, Malenka DJ, et al. Relation between operator and hospital volume and outcomes following percutaneous coronary interventions in the era of the coronary stent. *Jama* 2000;284(24):3139-44.
33. Urbach DR, Baxter NN. Does it matter what a hospital is "high volume" for? Specificity of hospital volume-outcome associations for surgical procedures: analysis of administrative data. *Qual Saf Health Care* 2004;13(5):379-83.

34. Endarterectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. *Jama* 1995;273(18):1421-8.
35. Bond R, Rerkasem K, Rothwell PM. Routine or selective carotid artery shunting for carotid endarterectomy (and different methods of monitoring in selective shunting). *Stroke* 2003;34(3):824-5.
36. Steg PG, Lopez-Sendon J, Lopez de Sa E, Goodman SG, Gore JM, Anderson FA, Jr., et al. External validity of clinical trials in acute myocardial infarction. *Arch Intern Med* 2007;167(1):68-73.
37. Fortin M, Dionne J, Pinho G, Gignac J, Almirall J, Lapointe L. Randomized controlled trials: do they have external validity for patients with multiple comorbidities? *Ann Fam Med* 2006;4(2):104-8.
38. Coca SG, Krumholz HM, Garg AX, Parikh CR. Underrepresentation of renal disease in randomized controlled trials of cardiovascular disease. *Jama* 2006;296(11):1377-84.
39. Petersen MK, Andersen KV, Andersen NT, Soballe K. "To whom do the results of this trial apply?" External validity of a randomized controlled trial involving 130 patients scheduled for primary total hip replacement. *Acta Orthop* 2007;78(1):12-8.
40. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? *Ann Intern Med* 2002;137(1):10-6.
41. Ettinger WH, Davis MA, Neuhaus JM, Mallon KP. Long-term physical functioning in persons with knee osteoarthritis from NHANES. I: Effects of comorbid medical conditions. *J Clin Epidemiol* 1994;47(7):809-15.
42. Imamura K, Black N. Does comorbidity affect the outcome of surgery? Total hip replacement in the UK and Japan. *Int J Qual Health Care* 1998;10(2):113-23.
43. Kadam UT, Jordan K, Croft PR. Clinical comorbidity was specific to disease pathology, psychologic distress, and somatic symptom amplification. *J Clin Epidemiol* 2005;58(9):909-17.
44. Kadam UT, Croft PR. Clinical comorbidity in osteoarthritis: Association with physical function in older patients in family practice. *J Rheumatol* 2007.
45. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann Surg* 2006;244(5):677-83.
46. Shapiro SH, Weijer C, Freedman B. Reporting the study populations of clinical trials. Clear transmission or static on the line? *J Clin Epidemiol* 2000;53(10):973-9.
47. Better reporting, better research: guidelines and guidance in PLoS Medicine. *PLoS Med* 2008;5(4):e99.

Legend

Box: Components of external validity evaluated in selected reports of trials of pharmaceutical and nonpharmaceutical treatments for hip and knee osteoarthritis

Figure 1: Median percentage (interquartile range) of items of components of external validity that were reported in selected reports of pharmaceutical and nonpharmaceutical treatments for hip and knee osteoarthritis

Scores are based on the percentage of items of the following components (baseline data, intervention, context) that were reported. 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. The cross represents the mean value.

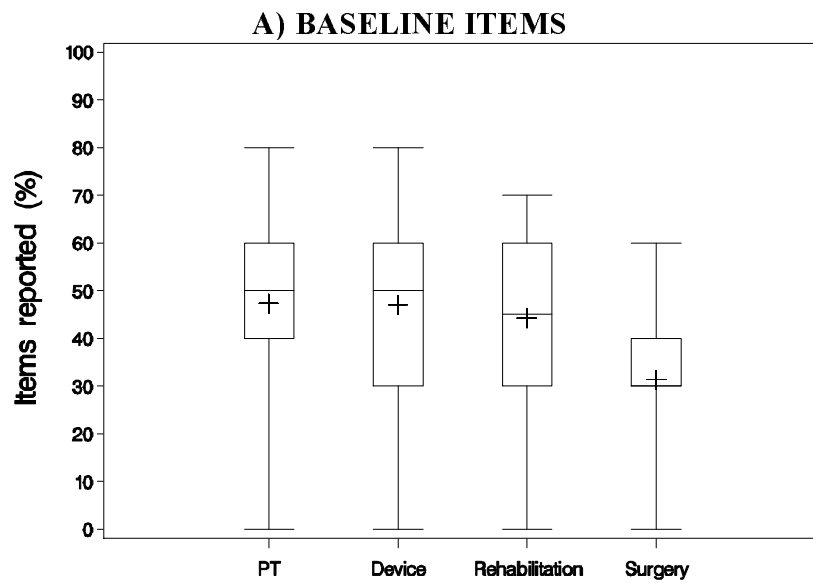
- A) Baseline items: percentage of essential data for the following baseline characteristics of randomized patients that were reported: age, sex, weight/body mass index, ethnicity, duration of disease, measure of function status, level of pain, description of radiographic evidence of damage, use of NSAIDs/other drugs, coexisting diseases.
- B) Interventions: percentage of the following essential items related to the intervention for each category of treatment that were reported:
 - a. Pharmacologic treatments (PT): route of administration, dosage, duration of treatment, frequency of treatment, compliance of patients
 - b. Devices: manufacturer, description of the devices, compliance of patients
 - c. Surgery: type of anesthesia, pre-operative care, post-operative care, technical procedure, compliance of care providers
 - d. Rehabilitation: number of sessions, timing of sessions, duration of each session, content of each session, mode of delivery, supervision or not of the session, compliance of patients

C) Context items: percentage of the essential items related to the context that were reported: location of recruitment, setting of recruitment, country where the trial took place, number of centers, details of the centers (e.g., name, center resources, center expertise, center organization), number of patients recruited in each center, details of care provider (qualification, name, experience, years of practice), number of care providers participating

Box: Components of external validity evaluated in selected reports of trials of pharmacologic and nonpharmacologic treatments for hip and knee osteoarthritis

<p>Description of participants</p>	<p>Percentage of the following baseline data reported</p> <ol style="list-style-type: none"> 1. Age 2. Sex 3. Weight/body mass index 4. Ethnicity 5. Duration of disease 6. Measure of function status 7. Level of pain 8. Description of radiographic evidence of damage 9. NSAIDs/other drugs 10. Coexisting diseases
<p>Intervention</p>	<p>Percentage of the following data describing the intervention reported</p> <p>Pharmacological Treatment</p> <ol style="list-style-type: none"> 1. Route of administration 2. Dosage 3. Duration of treatment 4. Frequency of treatment 5. Compliance of patients <p>Devices</p> <ol style="list-style-type: none"> 1. Manufacturer 2. Description of the devices 3. Compliance of patients <p>Surgery</p> <ol style="list-style-type: none"> 1. Type of anesthesia 2. Pre-operative care 3. Post-operative care 4. Technical procedure 5. Compliance of care providers <p>Rehabilitation</p> <ol style="list-style-type: none"> 1. Number of sessions 2. Timing of sessions 3. Duration of each session 4. Content of each session 5. Mode of delivery 6. Supervision or not of the session 7. Compliance of patients
<p>Context</p>	<p>Percentage of the following data related to the context reported</p> <ol style="list-style-type: none"> 1. Location of recruitment 2. Setting of recruitment 3. Country where the trial took place 4. Number of centers 5. Details of the centers (e.g., name, center resources, center expertise, center organization) 6. Number of patients recruited in each center 7. Details of care provider (qualification, name, experience, year of practice) 8. Number of care providers participating

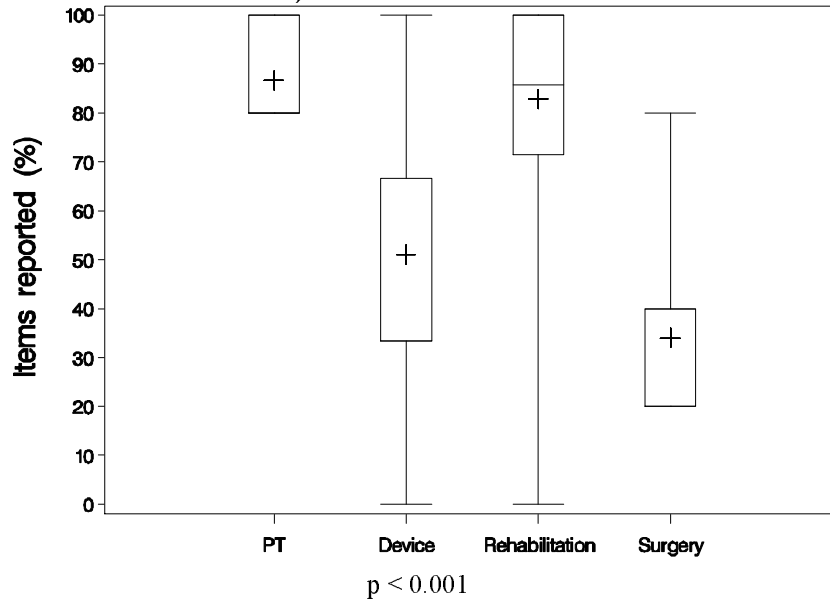
Figure 1: Median percentage (interquartile range) of items of components of external validity that were reported in selected reports of pharmaceutical and nonpharmaceutical treatments for hip and knee osteoarthritis



p= 0.006

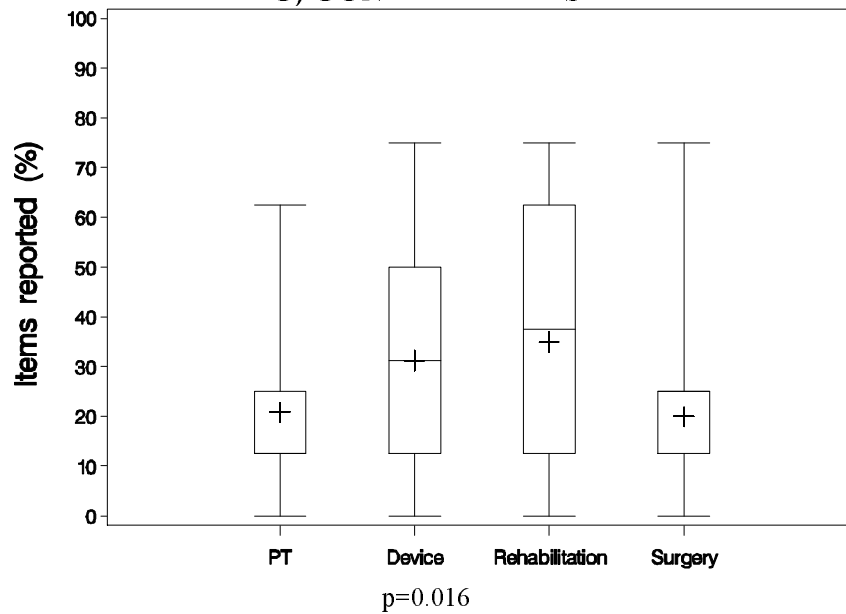
Pharmacologic treatment median, [IQR]	Device median, [IQR]	Rehabilitation median, [IQR]	Surgery median, [IQR]
50 [40-60]	50 [30-60]	45 [30-60]	30 [30-40]

B) INTERVENTION



Pharmacologic treatment median, [IQR]	Device median, [IQR]	Rehabilitation median, [IQR]	Surgery median, [IQR]
80 [80-100]	33 [33-67]	86 [71-100]	40 [20-40]

C) CONTEXT ITEMS



Pharmacologic treatment median, [IQR]	Device median, [IQR]	Rehabilitation median, [IQR]	Surgery median, [IQR]
12 [12-25]	31 [12-50]	37 [12-62]	25 [12-25]

PT: pharmacologic treatment;

Table 1: Characteristics of selected reports of trials of pharmacologic and nonpharmacologic treatments for hip and knee osteoarthritis.

	All Treatment N (%) N=120	Pharmacologic treatment N (%) N=30	Non implantable devices N (%) N=30	Rehabilitation N (%) N=30	Surgery N (%) N=30
Type of journal					
– General medical journal with high impact factor	13 (11)	4 (13)	4 (13)	4(13)	1 (3)
– Special medical journal and general medical journal with low impact factor	107 (89)	26 (87)	26 (87)	26 (87)	29 (97)
Medical area					
– Hip OA	18 (15)	1 (3)	1 (3)	5 (17)	11 (37)
– Knee OA	90 (75)	23 (77)	27 (90)	21 (70)	19 (63)
– Hip and knee OA	12 (10)	6 (20)	2 (7)	4 (13)	0
Funding					
– Public	45 (38)	6 (20)	15 (50)	17 (57)	7 (23)
– Manufacturer	18 (15)	8 (27)	4 (13)	0	6 (20)
– Public and manufacturer	7 (6)	2 (7)	0	2 (7)	3 (10)
– No funding	8 (7)	2 (7)	0	0	6 (20)
– Not reported	42 (35)	12 (40)	11 (37)	11 (37)	8 (27)
Sample size median (IQR¹)	100.0 (60-216)	213.5 (85-431)	66 (38-128)	107 (77-140)	95.5 (52-180)
Control group					
– Placebo intervention	43 (36)	18 (60)	17 (57)	6 (20)	2 (7)
– Active treatment	63 (52)	12 (40)	12 (40)	11 (37)	28 (93)
– Usual care	14 (12)	0	1 (3)	13 (43)	0
Internal validity: adequate					
– Generation of allocation sequences	61 (51)	19 (63)	11 (37)	18 (60)	13 (43)
– Allocation concealment	49 (41)	16 (53)	12 (40)	14 (47)	7 (23)
– Blinding of patients	52 (43)	26 (87)	16 (53)	2 (7)	8 (27)

¹ IQR: interquartile range

– Blinding of care providers	38 (32)	24 (80)	11 (37)	2 (7)	1 (3)
– Blinding of outcome assessors	71 (59)	25 (83)	22 (73)	11 (37)	13 (43)
– Intent-to-treat analyses	38 (32)	10 (33)	10 (33)	12 (40)	6 (20)

Table 2: Number (%) of selected reports of trials of pharmaceutical and nonpharmaceutical treatments for hip and knee osteoarthritis that described items related to external validity

Reporting of	All treatment N (%) N=120	Pharmacologic treatment N (%) N=30	Devices N (%) N=30	Rehabilitation N (%) N=30	Surgery N (%) N=30
Recruitment					
– Method of recruitment	43 (36)	8 (27)	13 (43)	18 (60)	4 (13)
– Specific method to enrich patient’s recruitment	23 (19)	18 (60)	5 (17)	0	0
– Duration of recruitment (10 patients/month)	56 (47)	10 (33)	11 (37)	15 (50)	20 (67)
Patients					
– Inclusion criteria	118 (98)	30 (100)	30 (100)	30 (100)	28 (93)
– Exclusion criteria	106 (88)	30 (100)	27 (90)	28 (93)	21 (70)
– Rate of strongly justified exclusion criteria in each article mean (SD)	75.5 (23.6)	77.7 (20.7)	79.0 (21.3)	66.9 (26.1)	79.3 (25.6)
– Rate of potentially justified in each article mean (SD)	1.5 (4.7)	1.1 (3.4)	1.5 (4.7)	2.8 (6.8)	0.5 (2.3)
– Rate of poorly justified in each article mean (SD)	22.9 (23.0)	21.3 (20.5)	19.5 (19.4)	30.3 (26.3)	20.2 (25.3)
– Flow diagram	48 (40)	18 (60)	11 (37)	17 (57)	2 (7)
– Number of eligible patients	50 (42)	12 (40)	14 (47)	21 (70)	3 (10)
– Number of patients not meeting inclusion criteria	39 (33)	9 (30)	8 (27)	19 (63)	3 (10)
– Number of patients refusing participation	31 (26)	6 (20)	8 (27)	14 (47)	3 (10)
– Baseline characteristics of randomized patients	109 (91)	28 (93)	28 (93)	27 (90)	26 (87)
– Age	108 (90)	28 (93)	28 (93)	27 (90)	25 (83)
– Sex	101 (84)	27 (90)	24 (80)	25 (83)	25 (83)
– Weight/body mass index	74 (62)	22 (73)	18 (60)	17 (57)	17 (57)
– Ethnicity	18 (15)	8 (27)	3 (10)	5 (17)	2 (7)
– Duration of disease	47 (39)	13 (43)	20 (67)	9 (30)	5 (17)
– Measure of function status	55 (46)	17 (57)	15 (50)	16 (53)	7 (23)
– Level of pain	47 (39)	14 (47)	15 (50)	15 (50)	3 (10)
– Description of radiographic damage	27 (23)	5 (17)	11 (37)	4 (13)	7 (23)
– NSAIDs/other drugs	19 (16)	6 (20)	6 (20)	6 (20)	1 (3)
– Coexisting diseases	14 (12)	2 (7)	1 (3)	9 (30)	2 (7)
Setting/center/care provider					
– Location of recruitment	46 (38)	9 (30)	17 (57)	17 (57)	3 (10)
– Setting of recruitment	40 (33)	7 (23)	16 (53)	16 (53)	1 (3)
– Country where the trial took place	25 (21)	8 (27)	6 (20)	10 (33)	1 (3)
– Number of centers	54 (45)	19 (63)	16 (53)	15 (50)	4 (13)
– Details of the centers	24 (20)	3 (10)	10 (33)	9 (30)	2 (7)
– Number of patients recruited in each center	0	0	0	0	0
– Details of care provider	35 (29)	2 (7)	6 (20)	10 (33)	17 (57)

– Number of care providers

33 (28)

2 (7)

4 (13)

7 (23)

20 (67)

Table 3: Number (%) of selected reports of trials of pharmaceutical and nonpharmaceutical treatments for hip and knee osteoarthritis that described the intervention

Reports	Reporting of	N (%)
Pharmacologic Treatment (N=30)	Mode of administration	30 (100)
	Dosage	30 (100)
	Duration of treatment	30 (100)
	Frequency of treatment	30 (100)
	Compliance of patients	10 (33)
Surgery (N=30)	Type of anesthesia	4 (13)
	Pre-operative care	2 (7)
	Post-operative care	15 (50)
	Technical procedure	30 (100)
	Compliance of care providers	0
Rehabilitation (N=30)	Number of sessions	29 (97)
	Timing of sessions	26 (87)
	Duration of each session	24 (80)
	Content of each session	28 (93)
	Mode of delivery	27 (90)
	Supervision or not	25 (83)
	Compliance of patients	15 (50)
Devices (N=30)	Manufacturer	9 (30)
	Description of the device	28 (93)
	Compliance of patients	9 (30)