Published in final edited form as:

J Clin Epidemiol. 2010 February ; 63(2): 154. doi:10.1016/j.jclinepi.2009.05.007.

Trial Sample Size, but not trial Quality is Associated with Positive Study Outcome

Jasvinder A Singh, MBBS, MPH^{1,2}, Stephen Murphy, BS³, and Mohit Bhandari, MD⁴

¹Dept of Medicine, Minneapolis VA Medical Center and University of Minnesota, Minneapolis, MN

² Departments of Health Sciences Research and Orthopedic Surgery, Mayo Clinic School of Medicine, Rochester, MN

³University of Notre Dame, Notre Dame, IN

⁴Department of Orthopaedic Surgery, McMaster University, Hamilton, Ontario

Abstract

Objective—To assess whether reported trial quality or trial characteristics are associated with trial outcome.

Study Design and Setting—We identified all eligible randomized controlled trials (RCTs) of arthroplasty from 1997 and 2006. Trials were classified based on whether the main trial outcome was reported to be positive (n=90) or negative (n=94). Multivariable logistic regression analyses studied the association of reporting of trial quality measures (blinding, placebo use, allocation procedure; overall quality) and trial characteristics (intervention type, number of patients/centers, funding) with positive trial outcome.

Results—RCTs that used placebo or blinded care providers, used pharmacological interventions, had higher Jadad quality scores or sample size >100 patients were significantly more likely to report positive result in univariate analyses. Multivariable regression did not identify methodological quality of RCTs, but rather that sample size, was associated with trial outcome. Studies with >100 patients were 2.2 times more likely to report a positive result than smaller studies (p=0.04).

Conclusions—Lack of association of reported trial quality with positive outcome in multivariable analyses suggests that previously observed association of reported study quality with study outcome in univariate analyses may be mediated by other study characteristics, such as study sample size.

What's New?

- Arthroplasty RCTs with sample size of 100 patients or more are significantly more likely to report a positive rather than a negative outcome.
- RCT quality was not associated with study outcome (positive vs. negative) in multivariable analyses, in contrast to previous studies that found an association of quality with outcomes in univariate analyses that did not adjust for sample size or type of intervention.

Corresponding Author: Jasvinder A Singh MBBS, MPH Minneapolis VA Medical Center, Rheumatology Office (111 R), One Veterans Drive, Minneapolis MN 55417, Email: jasvinder.md@gmail.com, Phone: 612-467-4190, Fax 612-467-2267.

Conflict Of Interest: None of the authors have any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

- Previously reported associations of study quality and outcomes may have been mediated by these characteristics.
- Future studies examining correlates of study outcomes should control for sample size, study quality and type of intervention.

Keywords

Assessment; Quality; Arthroplasty; Randomized Controlled Trial; Trial Outcome

Arthroplasty is one of the most significant advances in the treatment of end-stage arthritis, the most common cause of disability in adults in the U.S. [1]. Approximately one million arthroplasties are performed in the U.S. annually [2], a procedure associated with significant pain relief, improvement of function and quality of life [3]. Due to cost implications of arthroplasty and high volume, it is critical that clinical care should be evidence-based, relying on the most reliable research evidence [4,5]. Randomized Controlled Trials (RCTs) are widely accepted as the best method of assessing the treatment effects [1].

The extent to which RCT quality is associated with trial outcomes remains unclear, as previous research provides contradictory evidence. While some systematic reviews of RCTs from various therapeutic areas [6,7] and orthopedic surgery [8] found a larger treatment effect in studies of lower quality, Balk et al. found no such association in their study of RCTs of cardiovascular disease, infectious disease, pediatrics and surgery [9]. Evidence also suggests that trial size is not related to treatment effects (i.e. good quality large and good quality small RCTs are similar) [6]. A previous study found that multi-center studies were associated with stronger treatment effect in cardiovascular and pediatric trials but a weaker treatment effect in infectious diseases and surgery trials [9].

We conducted a systematic review of all RCTs in patients with arthroplasty to examine the association of reporting of RCT quality with trial outcome. Specifically, we aimed (1) to examine if quality of trial reporting is associated with positive trial outcome; and (2) to test whether trial (funding source, number of centers, number of patients per trial, type of intervention) and publication (year of publication, type of journal, journal impact factor) characteristics are associated with positive trial outcome. We hypothesized that better quality of reporting will be associated with higher likelihood of positive trial outcome and that larger trials and those testing pharmacological interventions will be more likely to have positive results.

Methods

Search Strategy

A Cochrane Systematic review librarian (I.R.) searched Medline using the following search terms, with results limited to randomized controlled trials published in the two calendar years, 1997 and 2006: “exp arthroplasty, Replacement, Knee/ or exp Joint Prosthesis/ or exp Arthroplasty, Replacement/ or joint arthroplasty.mp. or exp Arthroplasty, Replacement, Hip.” We selected two years for review to have a large enough sample size; the reason to select the articles from two calendar years was to examine the effect of year of publication (9 years apart) on positive trial outcome. Upon review of the titles and abstracts by a senior author (J.S.), articles were excluded if they were letter/editorial, non-randomized, published in non-English language, not arthroplasty-related or did not include clinical outcomes (i.e., economic analyses etc.). There were no restrictions by the journal name or specialty.

Detailed evaluation of reporting of study quality

The data abstraction was performed by a single abstractor (S.M.) trained by the senior epidemiologist (J.S.). The training consisted of: (1) review of the key articles in the literature describing the RCT quality assessment; (2) detailed discussion of key assessment components, including allocation concealment, allocation sequence generation, blinding, use of placebo, intention to treat (ITT), modified ITT analyses and loss to follow-up; (3) three rounds of independent abstraction of articles by both the senior author (J.S.) and the trained abstractor, which led to >95% agreement on all abstracted data. The abstractor was blinded to the study hypotheses. Trial, publication and quality characteristics and trial outcome were abstracted from all included studies using a structured abstraction form, modified from that used by Burton et al [10] and entered into Microsoft Access 2003 database (Redmond, WA).

We obtained the following characteristics for each included study: (1) trial characteristics: funding source (public, private, neither, both or not clear), number of centers (<2, ≥2, not clear), number of patients per trial (trial sample size), body region (knee, hip, both or other) and type of intervention (pharmacologic vs. non-pharmacologic); as previously [10-14]); (2) publication characteristics: year of publication (1997, 2006), type of journal (orthopedics/surgery, anesthesia, internal medicine/medical subspecialties and rehabilitation/others), journal impact factor (≤2, >2); and (3) trial outcome- negative, positive or both.

Trial outcome was the outcome of interest and it was classified similar to a previously published study [15]. If primary outcome in an RCT (as specified in the methods section) was positive or negative, the RCT outcome was classified as positive or negative, respectively. If the study did not specify an outcome measure as primary, we reviewed the introduction, results, abstract and methods sections to select the most clinically relevant outcome, as described by the authors, as the primary outcome. Often, this was the outcome stated as the “main objective” or the “main purpose” of the study. When not clearly stated, it was the outcome most closely related to the main objective/purpose. If an RCT specified more than one outcome as primary and the results were both positive and negative, the outcome was classified as mixed, and excluded from the analyses.

We extracted the impact factor of the journals from the Journal Citation Reports (Institute for Scientific Information) for the respective years of publication, 1997 and 2006. We also examined the reporting of the following quality characteristics, namely, whether:

1. generation of randomization sequence was- (a) appropriate - if selection bias was prevented by use of random numbers, computerized random number generation, pharmacy controlled, opaque sealed envelopes, numbered or coded bottles, (b) inappropriate - if patients were allocated alternately, according to date of birth, date of admission, hospital number etc., or (c) not described;
2. allocation concealment was- (a) appropriate - if both patients and investigators enrolling patients in the study could not foresee the assignments due to centralized randomization/pharmacy control/opaque envelopes etc., (b) inadequate or (c) not described;
3. blinding of patients, care providers and outcome assessors was- (a) appropriate – stated that respective individuals could not identify the intervention being tested or use of active placebos, identical placebos or dummies [16], (b) inappropriate – comparison of tablet vs. injection with no double-dummy, or (c) not described;
4. placebo was used as the control intervention- (a) yes, or (b) no;
5. trial reported a justification for sample size- (a) yes; (b) no; or not described;

6. whether the analyses were described as intention to treat analysis (ITT), i.e., all participants randomized were included in the analysis and kept in the original groups [17] or modified ITT, i.e., analysis excluded those who never received treatment or who were never evaluated while receiving treatment- (a) yes; (b) no; (c) or not described
7. the loss to follow-up was <20%- (a) yes; (b) no; (c) or not described.

Overall trial quality reporting was assessed using three validated measures, the Jadad score [18,19], the Delphi list's overall score [20] and overall subjective assessment of validity of the study as described in the *Users' Guides to the Medical Literature* [21]. Jadad scale is the most widely used validated study quality measure that assesses the appropriateness of randomization, blinding and loss to follow-up and ranges from 0-5. The Delphi list was chosen since it includes many items not included in the Jadad score that assess trial characteristics on a 0-9 score, including randomization, similarity at baseline, eligibility criteria, allocation concealment, blinding of outcome assessor, patient and care provider, inclusion of ITT and report of point estimates and variability. The overall subjective evaluation of the study's quality was assessed with is another single-item overall quality score 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?" For all three measures, a higher score indicates higher quality.

Statistical Analysis

We used chi-square tests to examine the univariate association of trial and publication characteristics. Student's t-tests and chi-square tests were used to examine the association of reporting of overall trial quality and each quality characteristic, respectively, with trial outcome (positive vs. negative). Multivariable logistic regression analyses included positive trial result as the outcome and the following potential predictors: all trial and publication characteristics associated with positive trial outcome with $p \leq 0.10$ in univariate analyses and the Jadad score as the overall assessment of trial quality reporting ($p=0.02$). Since regression analyses are sensitive to outliers, we categorized the two variables with a right-skewed distribution into categories of analyzable sizes: number of patients into ≤ 100 and >100 patients and journal impact factor into ≤ 2 and >2 .

Sensitivity analyses were performed by replacing Jadad score with Delphi list or the NRS (both overall quality scores) or the seven quality characteristics (as described above, including randomization technique, allocation concealment, blinding, use of placebo, sample size justification, ITT and loss to follow-up) in multivariable regression models. Additional sensitivity analyses were performed by replacing Jadad score with all individual trial quality characteristics, since Jadad scale includes blinding and use of placebo, which may not be possible and/or ethical in RCTs of surgical interventions [22]. We were unable to examine the effect of various quality characteristics on treatment effect sizes, since a large proportion of RCTs reported non-continuous outcomes and some did not report measures of variation. All data analyses were performed using SPSS version 15 (Chicago, IL). All tests were two-sided and we considered $p < 0.05$ statistically significant.

Results

From a total of 224 RCT reports, 196 met the initial eligibility criteria of which 184 comprised the analytic set - positive ($n=90$) or negative ($n=94$) trial results (total $n=184$), after excluding the 12 trials that had some outcomes with positive and some with negative results (Figure 1).

Of these, 69 had ≤ 50 patients, 66, 51-100 patients and 59, more than 100 patients. Private funds supported 31%, public funds 5%, both 2%, no funding in 11% and no information in 50%.

Journal impact factor was <2 for 66% RCTs, and >2 for 34%. Other trial and publication characteristics are summarized in Figure 1 & Table 1.

Univariate Correlates of Positive Trial Outcome

RCTs with >100 patients, pharmacological intervention, that used placebo or blinded care providers were significantly more likely to report positive result in univariate analyses (Appendix 1). Higher Jadad, Delphi list and NRS scores were significantly associated with higher likelihood of a positive result (Figure 2). Non-significant trend was noted for unfunded, single-center or 1997-published RCTs for a positive trial outcome (p -value, 0.055-0.10).

Multivariable Correlates of Positive Trial results

Multivariable models showed that Jadad score, a measure of reported trial quality, was not significantly associated with positive trial outcome (Table 2). RCTs without a funding source, published in 1997 or those using a pharmacological intervention had a non-significant trend of higher odds for a positive trial result ($p \leq 0.24$) (Table 2).

RCTs with more than 100 patients were 2.2 times more likely to report a positive result than trials with ≤ 100 patients ($p=0.04$). Further exploration of study sample size revealed significant associations with a variety of quality characteristics in univariate analyses, including: blinding of patients ($p=0.02$) and outcome assessors ($p=0.01$) and use of ITT/modified ITT ($p<0.001$) (Appendix 2).

Sensitivity Analyses

Sensitivity analyses were performed by replacing the Jadad score in multivariable regression analyses with other statistically significant quality correlates of positive outcome in univariate analyses i.e. (a) Delphi list, (b) NRS scores or (c) placebo intervention and blinding of care providers; or performing a multivariable regression analysis with (d) all univariate significant predictors of trial outcome except measures of RCT quality i.e. Jadad score (since it may not be possible/ethical to blind or use placebo in some surgical trials) and (e) allocation concealment, allocation sequence generation, blinding and study sample size, similar to recent study [6]. This led to minimal change in the odds ratio (95% confidence interval) estimates for the association of larger trials (≥ 100 patients) with positive trial outcome, changing from of 2.2 (1.02, 4.58) to 2.1 (0.97, 4.41) with Delphi list (analysis “a”), 2.1 (0.99, 4.45) with NRS (analysis “b”), 1.9 (0.86, 4.0) with individual quality characteristics (analysis “c”), 2.2 (1.04, 4.63) (analysis “d”) and 2.5 (1.28, 4.83) (analysis “e”) respectively.

Discussion

What does this report add to the literature?

In this first systematic review of a large number of arthroplasty RCTs, we found reported in multivariable analyses, trial quality as measured by an overall score, i.e. Jadad, Delphi or Numeric Rating Scale, was not associated with primary trial outcome being positive or negative. In univariate analyses, trials with higher overall quality scores, those studying pharmacological interventions or using placebo or blinding of care providers were more likely to report positive results, but these associations were no longer significant in multivariable analyses. Another significant finding of this study was lack of association of allocation procedure, blinding of patients/outcome assessors, use of ITT/modified ITT, loss of follow-up $<20\%$ or power calculations with the trial outcome. Many findings including the lack of association of trial characteristics and positive trial result in multivariable analyses, after accounting for trial sample size, are novel and add to the current literature.

Most previous studies studying the association of reporting of RCT quality with trial outcome did not adjust for sample size [7,23,24] [9], with the exception of one study [6]. Most studies reported association between reported trial quality and trial outcomes [6,7,23,24], while some didn't [9]. Study sample size is known to be associated with methodological quality [25], a finding we confirmed in our dataset. It seems quite likely that the observed association of reported trial quality with trial outcome in these previous studies was due to association between sample size and trial outcome, rather than reported study quality and trial outcome. This is evident by presence of association between reported quality and outcomes in univariate analyses in previous systematic reviews [6,7,23,26] and in our study, which was no longer significant in multivariable-adjusted analyses that controlled for important confounders such as sample size, type of intervention etc. We recommend that future systematic reviews examining the effect of RCT reporting quality on outcomes should include the study sample size and other important confounders in multivariable analyses. While reporting of study quality is variable across studies due to various reasons, sample sizes are usually reported accurately in most RCTs.

We found that the only independent predictor of positive trial result in multivariable analysis was the higher number of patients in RCTs. Arthroplasty RCTs with 100 or more patients were more likely to have positive results in both univariate and multivariable analyses. While to many methodologists and clinicians this may be intuitive, to our knowledge, none of the previous reviews of RCTs have performed multivariable analyses to examine this association and this is a novel finding. The only study that examined this question in some detail included >200 RCTs from multiple areas stratified by sample size. They found that trial size (≥ 1000 vs. < 1000) was not independently associated with treatment effect, after controlling for reporting of RCT quality- namely that only small trials with inadequate allocation sequence generation, inadequate allocation concealment and no double-blinding reported exaggerated intervention effects; small and large trials of high quality were similar [6]. Our findings differ in that we did not find allocation concealment or allocation sequence generation as significant correlates of trial outcome in univariate analyses. Sensitivity analyses that replicated the multivariable regression model in the previous study (adjusting for allocation concealment, sequence generation and blinding) did not diminish the strength of association of trial sample size to trial outcomes. This result was in contrast to the earlier study [6]. Our study differs from the previous study in the definition of large sample (≥ 1000 vs. ≥ 100), outcome used (ratio of events/participant in treatment and control group vs. positive/negative trial result in our study) and type of RCTs assessed (multiple therapeutic areas vs. arthroplasty RCTs). We postulate that arthroplasty studies with larger sample size are more likely to have adequate power and thus more likely to detect a difference between intervention and control (i.e., positive outcome) as compared to studies of smaller sample size.

We found a trend towards greater likelihood of positive trial result with multi-center trials compared to single-center trials (62% vs. 52%, $p=0.10$) in univariate, but not in multivariable-adjusted analyses. In a previous study, that used univariate analyses, contradictory results were found, i.e., multi-center RCTs were associated with smaller treatment effect size for surgery, but a larger effect size for pediatrics trials [9]. These analyses were not adjusted for sample size or quality of trial reporting. Since most multi-center trials also happen to be of larger size, we postulate that this association is at least partially due to sample size, as confirmed in our study. We found lack of association of allocation sequence generation with treatment effects, confirming findings from previous studies [7,23,24,27], but in contrast to one study [6]. Clearly, due to conflicting evidence in these areas, more studies are needed.

Our study has several limitations. We examined the reporting of RCTs, not the RCTs themselves; it is possible that reporting of methods was inadequate for some RCTs that were conducted with more rigor, as reported previously [28]. Secondly, we used Jadad score and

Delphi list to assess overall quality of trial reporting, which focus primarily on blinding and randomization, which may not fairly evaluate the quality of non-pharmacological interventions [22]. However, even in the multivariable analyses that did not include Jadad score, the association of trial sample size and positive outcome were unaltered, signifying that the differences were not due to quality standards that may not be applicable to surgical trials. Use of a different trial outcome such as ratio of odds ratio, intervention treatment effect size or RCTs from non-surgical fields may provide different results than noted in our study. We did not have adequate power to examine whether the association between sample size and likelihood of positive result would hold true for each RCT outcome, such as pain, function, length of stay or implant survival. This was due to the fact that there were a number of outcomes examined in these trials. Due to reporting of continuous outcomes for a small proportion of studies, we were unable to examine treatment effect size as an outcome. Inclusion of more number of years and databases other than Medline would have allowed us to identify more articles and likely made our analyses more robust. We did not restrict our analyses to a particular type of RCTs (one surgical technique vs. another; surgical only etc.), since our objective was not to study a subgroup of arthroplasty RCTs. This likely led to heterogeneity.

In summary, we found a larger sample size was an independent predictor of a positive trial outcome. Better RCT quality reporting or use of pharmacological intervention was associated with positive trial outcome in univariate analyses, but was not an independent predictor of positive trial outcome. These findings indicate that in future systematic reviews examining the association of quality or trial characteristics with trial outcomes should control for trial sample size and other important study characteristics.

Supplemental Materials

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Mr. Indy Rutks from the Cochrane Group for performing the literature search, Ms. Ruth Brady (research associate) and Ms. Perlita Ochoa (administrative assistant) for their administrative help.

Grant Support: Dr. Singh was supported by the NIH CTSA Award 1 KL2 RR024151-01 (Mayo Clinic Center for Clinical and Translational Research); and Minneapolis VA Medical Center, Minneapolis, MN; Dr. Mohit Bhandari was supported in part by Canada Research, McMaster University.

References

1. Prescott RJ, Counsell CE, Gillespie WJ, et al. Factors that limit the quality, number and progress of randomised controlled trials. *Health Technol Assess* 1999;3:1–143. [PubMed: 10683591]
2. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the united states from 2005 to 2030. *J Bone Joint Surg Am* 2007;89:780–5. [PubMed: 17403800]
3. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am* 2004;86-A:963–74. [PubMed: 15118039]
4. Guyatt GH, Haynes RB, Jaeschke RZ, et al. Users' guides to the medical literature: Xxv. Evidence-based medicine: Principles for applying the users' guides to patient care. Evidence-based medicine working group. *JAMA* 2000;284:1290–6. [PubMed: 10979117]
5. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. *BMJ* 1996;312:71–2. [PubMed: 8555924]
6. Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med* 2001;135:982–9. [PubMed: 11730399]

7. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609–13. [PubMed: 9746022]
8. Poolman RW, Struijs PA, Krips R, et al. Reporting of outcomes in orthopaedic randomized trials: Does blinding of outcome assessors matter? *J Bone Joint Surg Am* 2007;89:550–8. [PubMed: 17332104]
9. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973–82. [PubMed: 12052127]
10. 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;290:1062–70. [PubMed: 12941679]
11. Boutron I, Estellat C, Guittet L, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: A systematic review. *PLoS Med* 2006;3:e425. [PubMed: 17076559]
12. Boutron I, Guittet L, Estellat C, Moher D, Hrobjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Med* 2007;4:e61. [PubMed: 17311468]
13. Boutron I, Tubach F, Giraudeau B, Ravaud P. Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *J Clin Epidemiol* 2004;57:543–50. [PubMed: 15246122]
14. Foley NC, Bhogal SK, Teasell RW, Bureau Y, Speechley MR. Estimates of quality and reliability with the physiotherapy evidence-based database scale to assess the methodology of randomized controlled trials of pharmacological and nonpharmacological interventions. *Phys Ther* 2006;86:817–24. [PubMed: 16737407]
15. Gluud C, Gluud LL. Evidence based diagnostics. *BMJ* 2005;330:724–6. [PubMed: 15790646]
16. Schulz KF, Grimes DA. Blinding in randomised trials: Hiding who got what. *Lancet* 2002;359:696–700. [PubMed: 11879884]
17. Hill CL, Buchbinder R, Osborne R. Quality of reporting of randomized clinical trials in abstracts of the 2005 annual meeting of the american college of rheumatology. *J Rheumatol* 2007;34:2476–80. [PubMed: 17985398]
18. Jadad AR, Cook DJ, Jones A, et al. Methodology and reports of systematic reviews and meta-analyses: A comparison of cochrane reviews with articles published in paper-based journals. *JAMA* 1998;280:278–80. [PubMed: 9676681]
19. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Control Clin Trials* 1996;17:1–12. [PubMed: 8721797]
20. Verhagen AP, de Vet HC, de Bie RA, 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;51:1235–41. [PubMed: 10086815]
21. Guyatt GH, Veldhuyzen Van Zanten SJ, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: A taxonomy and review. *CMAJ* 1989;140:1441–8. [PubMed: 2655856]
22. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P. Extending the consort statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Ann Intern Med* 2008;148:295–309. [PubMed: 18283207]
23. Moher D, Cook DJ, Jadad AR, et al. Assessing the quality of reports of randomised trials: Implications for the conduct of meta-analyses. *Health Technol Assess* 1999;3:i–iv. 1–98. [PubMed: 10374081]
24. 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;273:408–12. [PubMed: 7823387]
25. Kjaergard LL, Nikolova D, Gluud C. Randomized clinical trials in hepatology: Predictors of quality. *Hepatology* 1999;30:1134–8. [PubMed: 10534332]
26. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60. [PubMed: 10493204]
27. Emerson JD, Burdick E, Hoaglin DC, Mosteller F, Chalmers TC. An empirical study of the possible relation of treatment differences to quality scores in controlled randomized clinical trials. *Control Clin Trials* 1990;11:339–52. [PubMed: 1963128]

28. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55:783–6. [PubMed: 12384192]

Figure 1. Flow chart of articles selected for review

^aThe totals add up to more than these numbers, since many trials had both pharmacological and non-pharmacological interventions. ^b Many RCTs had more than one type of intervention, therefore the total adds up to a greater number. Positive and negative RCTs were those that had the primary outcome reported as positive or negative, respectively. We excluded the 12 RCTs that reported the co-primary outcomes as one being positive and other being negative.

Figure 2. Association of overall trial quality with positive and negative results

Table 1
Characteristics of Eligible Studies*

	1997 (n=67) N (%)	2006 (n=129) N (%)	All (n=196) N (%)
Body part			
Knee	22 (32%)	58 (29%)	80 (40%)
Hip	33 (48%)	48 (37%)	81 (41%)
Both Hip and Knee	10 (15%)	9 (7%)	19 (9%)
Others	4 (6%)	16 (12%)	20 (10%)
Number of Centers			
Single-center	27 (40%)	43 (33%)	70 (36%)
Multi-center	15 (22%)	22 (17%)	37 (19%)
Unclear	25 (37%)	63 (49%)	89 (45%)
Number of Patients			
≤100	43 (64%)	92 (71%)	135 (69%)
>100	23 (36%)	36 (29%)	59 (31%)
Type of Intervention			
Pharmacological	32 (48%)	47 (36%)	79 (40%)
Non-pharmacological	37 (19%)	93 (47%)	130 (66%)
Funding Support			
No information provided	46 (69%)	54 (42%)	100 (50%)
None	3 (5%)	18 (14%)	21 (11%)
Private	18 (27%)	44 (34%)	62 (31%)
Private & public	0 (0%)	4 (3%)	4 (2%)
Public	0 (0%)	9 (7%)	9 (5%)
Type of Journal			
Ortho/Surgery	38 (57%)	87 (67%)	125 (64%)
Other/Rehab	14 (20%)	20 (16%)	10 (6%)
Anesthesia	13 (19%)	14 (11%)	34 (17%)
Internal Medicine	2 (3%)	9 (6%)	27 (14%)

* Total may add up to >100 since numbers are rounded to the nearest digit and many trials had >1 type of intervention

Table 2
Multivariable Predictors of Positive Trial Outcome

	Odds Ratio	95% CI	p-value
Number of patients			0.044
≤100	1.0		
>100	2.16	1.02, 4.56	
Type of Intervention			0.24
Non-pharmacological	1.0		
Pharmacological	1.54	0.75, 3.20	
Year			0.20
2006	1.0		
1997	1.55	0.79, 3.03	
Funding			0.15
Yes	1.0		
None or no information	1.66	0.84, 3.23	
Number of Centers			
Single-center	1.0		0.67
Multi-center	1.0	0.37, 2.70	
No information provided	0.74	0.36, 1.52	
Jadad Score *	n/a		0.97

* Jadad score was treated as a continuous variable, therefore no odds ratios were calculated