Systematic Review

Methodologic Quality of Knee Articular Cartilage Studies

Joshua D. Harris, M.D., Brandon J. Erickson, M.D., Geoffrey D. Abrams, M.D., Gregory L. Cvetanovich, M.D., Frank M. McCormick, M.D., Anil K. Gupta, M.D., M.B.A., Bernard R. Bach Jr., M.D., and Brian J. Cole, M.D., M.B.A.

Purpose: (1) To evaluate the quality of knee articular cartilage surgery literature using established methodologic quality instruments, and (2) to assess whether study quality has improved with time. **Methods:** A systematic review was performed using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Studies of autologous chondrocyte implantation (ACI), osteochondral autograft and allograft transplant, and microfracture were analyzed. Study methodologic quality was assessed by the level of evidence and 9 different methodologic quality questionnaires. Comparisons were made between different surgical technique groups by use of Student's *t* tests. Assessment of study quality improvement with time was performed by comparison of the Coleman Methodology Score (CMS) from the included studies (2004 to present) and CMS from a prior study assessing quality of articular cartilage studies (1985 to 2004). Furthermore, assessment of study quality improvement with time was performed over the period of the included studies (2004 to present). **Results:** We included 194 studies (11,787 subjects). Most evidence was Level IV (76%) and nonrandomized (91%). ACI was the most commonly reported technique (62% of studies). Only 34% of studies denied the presence of a financial conflict of interest. The mean subject age was 33.5 ± 8.2 years, and the mean length of follow-up was 3.7 ± 2.3 years. By use of study quality questionnaires, the methodologic quality of articular cartilage studies was poor. However, study quality (after 2004) was significantly improved versus that reported from a prior study (before 2004) using the CMS ($P < .01$). The mean level of evidence, CMS, CONSORT (Consolidated Standards of Reporting Trials) score, and Jadad score showed no significant improvement over the period of the included studies ($P > .05$). The quality of randomized controlled trials (RCTs) was significantly higher than that of non-RCTs ($P < .05$). The most common study weaknesses included blinding, subject selection process, study type, sample size calculation, and outcome measures and assessment. **Conclusions:** The methodologic quality of knee articular cartilage surgery studies was poor overall and also for individual techniques (ACI, osteochondral autograft transplant, osteochondral allograft transplant, and microfracture). However, the overall quality of the investigations in this review (after June 2004) has significantly improved in comparison to those published before 2004. The quality of RCTs was significantly higher than that of non-RCTs. Level of evidence, CMS, CONSORT score, and Jadad score did not significantly improve with later publication date within the period of the studies analyzed. Methodologic quality deficiencies identified in this investigation may be used to guide future articular cartilage studies' design, conduct, and reporting. **Level of Evidence:** Level IV, systematic review of studies with Levels of Evidence I-IV.

From Midwest Orthopaedics at Rush, Rush University Medical Center, Chicago, Illinois, U.S.A.

The authors report that they have no conflicts of interest in the authorship and publication of this article.

Received January 10, 2013; accepted February 20, 2013.

Address correspondence to Joshua D. Harris, M.D., Midwest Orthopaedics at Rush, Rush University Medical Center, 1611 W Harrison St, Ste 300, Chicago, IL 60612, U.S.A. E-mail: joshuaharrismd@gmail.com

0749-8063/1335/$36.00

http://dx.doi.org/10.1016/j.arthro.2013.02.023

Note: To access the supplementary tables accompanying this report, visit the [Month] issue of Arthroscopy at www.arthroscopyjournal.org.

Observation of articular cartilage pathology is common during knee arthroscopy.[1] A wide spectrum of chondral disease exists, ranging from superficial to full-thickness lesions that may or may not involve the underlying subchondral bone.[2] Defect progression is multifactorial, and several concomitant patient-, knee-, and defect-specific factors influence progression of the defect to osteoarthritis.[3] Although indications vary slightly between patients and surgeons, most agree that suitable candidates for cartilage repair and restoration surgery are symptomatic, young or middle-aged, motivated individuals with either normal or correctable comorbidities (including malalignment and meniscoligamentous deficiency). However, patients who meet these criteria (those with isolated full-thickness chondral

defects) comprise only 5% of those with cartilage pathology in the knee.[4] Despite this, the true prevalence of isolated full-thickness chondral defects is likely unknown, because asymptomatic defects may go unrecognized and undiagnosed.

The difficulty in identification of symptomatic chondral pathology in the knee warrants caution in proceeding with surgical techniques used to treat them. Although the exact mechanism is not completely known, it is recognized that an isolated chondral defect may cause significant pain.[5] At this time, the natural history of an isolated cartilage defect in the knee is unknown. Despite this, there are many viable surgical techniques used to treat these defects. Whereas many procedures are simple and inexpensive (e.g., arthroscopy with debridement, drilling, and microfracture), others require considerable investments of time and money (e.g., cell-based therapies or allograft transplants [osteochondral and meniscal]). In the current era of patient satisfaction–driven outcome measures and metrics used to rate cost-efficient physician performance and reimbursement,[6] it is necessary to practice not only "evidence-based medicine" but also "high-quality evidence-based medicine."

The quality of evidence for articular cartilage surgery has been limited by several methodologic quality deficiencies.[7,8] However, recent improvements in the performance and reporting of studies examining cartilage surgery show recognition of these methodologic quality deficiencies and have made attempts to address them. Several recent systematic reviews in both cartilage and non-cartilage surgery have shown a temporal relation between later study publication date and significant improvements in study methodologic quality.[7,9-11]

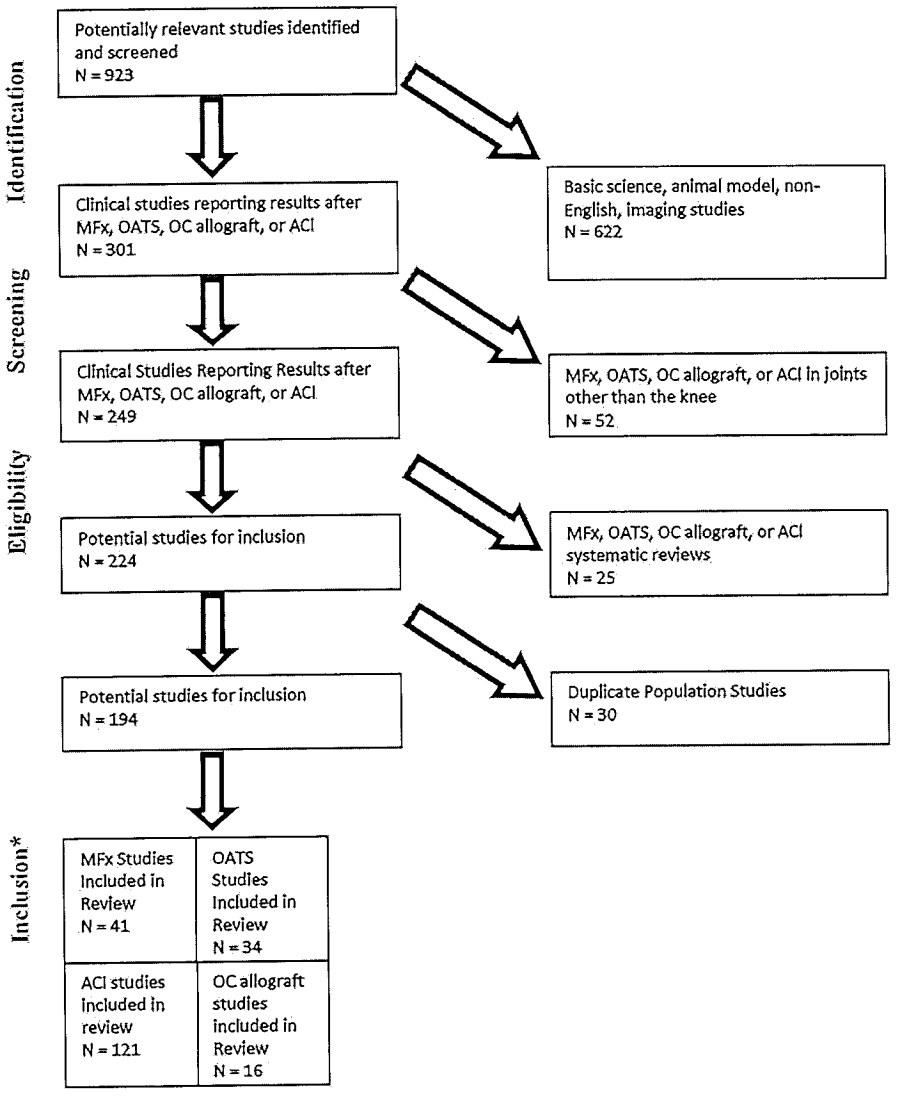

Fig 1. Systematic review search algorithm within Medline database according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. After application of all exclusion criteria, 194 studies were identified for inclusion and further analysis. The asterisk indicates that although 194 studies were identified for final analysis, the sum of the 4 different techniques is greater than 194 because of more than 1 technique being reported in individual studies. (ACI, autologous chrondrocyte implantation; MFx, microfracture; OATS, osteochondral autograft transplant; OC, osteochondral.)

The purposes of this study were to evaluate the quality of all knee articular cartilage surgery studies in the orthopaedic literature using previously established methodologic quality instruments, to identify their strengths and weaknesses, and to assess whether study quality has improved with time. We hypothesized that the quality of articular cartilage evidence has been improving with time but will be inherently limited in certain aspects of study design and reporting because of the use of inappropriate questionnaires.

Methods

We conducted a systematic review of multiple medical databases using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines with a PRISMA checklist without a formal protocol or registration number.[12] Two independent reviewers completed the search individually on August 10, 2012, using search databases PubMed (Medline), CINAHL (Cumulative Index to Nursing and Allied Health Literature), SportDiscus, and Cochrane Central Register of Controlled Trials for trials published from June 25, 2004, to the present. The former date represents the last date of study inclusion for a similar review on the methodologic quality of knee articular cartilage studies.[7] Our study extends on these methods to the ensuing 8 years. The search terms were as follows: articular cartilage, repair, restoration, knee, surgery, autologous chondrocyte implantation, transplantation, microfracture, subchondral drilling, marrow-stimulation, osteochondral autograft, mosaicplasty, and osteochondral allograft. Levels of Evidence I, II, III, and IV (as graded per the Oxford Centre for Evidence-Based Medicine used by the American version of the *Journal of Bone and Joint Surgery*)[13] were deemed inclusive. Both print journal articles and those published only electronically were eligible for inclusion and analysis. All references within included studies were cross-referenced for potential inclusion if missed by the initial search. Figure 1 shows the flowchart of study identification, screening, eligibility, and inclusion.

Inclusion Criteria

The inclusion criteria were as follows:

1. English language
2. Clinical outcome studies after articular cartilage surgery in the knee
   a. Chondroplasty
   b. Debridement
   c. Marrow-stimulation cartilage repair techniques
      i. Microfracture
      ii. Subchondral bone drilling
      iii. Abrasion arthroplasty
      iv. Autologous matrix-induced chondrogenesis
   d. Cartilage restoration techniques
      i. Autologous chondrocyte implantation (ACI)
      ii. Osteochondral autograft transplant (OAT)/ mosaicplasty
      iii. Osteochondral allograft transplant
      iv. One-stage ACI (Cartilage Autograft Implantation System; DePuy Mitek, Raynham, MA)
      v. Minced juvenile allograft cartilage (DeNovo NT or DeNovo ET; Zimmer, Warsaw, IN; ISTO Technologies, St. Louis, MI)
3. Levels of Evidence I through IV

Exclusion Criteria

The exclusion criteria were as follows:

1. Non-English language
2. Studies failing to report clinical outcomes after cartilage surgery in the knee
3. Studies of cartilage surgery in joints other than the knee
4. Basic science, biomechanical, anatomic, and surgical technique studies, as well as letters to the editor

Methodologic Quality Analysis

Study methodologic quality for all studies analyzed was evaluated (by 2 reviewers) with the Coleman Methodology Score (CMS)[14,15] and Quality Appraisal Tool (QAT).[16] The CMS is a 2-part (A and B), 10-item questionnaire (Supplementary Table 1, online only, available at www.arthroscopyjournal.org), scored from 0 to 100 (excellent, 85 to 100; good, 70 to 84; fair, 55 to 69; and poor, <55) that has been used in evaluation of both randomized controlled trials (RCTs) and non-RCTs. The QAT is a 12-item questionnaire, scored from 0 to 24, assigned a percentile "quality rating" used to assess non-RCTs (Supplementary Table 2, online only, available at www.arthroscopyjournal.org). For all RCTs analyzed in this review, the following methodologic quality scores were used (by 2 reviewers): CONSORT (Consolidated Standards of Reporting Trials) score (Supplementary Table 3, online only, available at www.arthroscopyjournal.org)[17]; Jadad scale (Supplementary Table 4, online only, available at www.arthroscopyjournal.org)[18]; Modified Coleman Methodology Score (MCMS)[14] (Supplementary Table 5, online only, available at www.arthroscopyjournal.org); Detsky Quality Assessment Scale[19] (Supplementary Table 6, online only, available at www.arthroscopyjournal.org); the Cochrane Bone, Joint and Muscle Trauma Group Quality Assessment Tool[20] (Supplementary Table 7, online only, available at www.arthroscopyjournal.org); Delphi List[21] (Supplementary Table 8, online only, available at www.arthroscopyjournal.org); and CLEAR-NPT (Checklist to Evaluate a Report of a Non-Pharmacologic Trial)[22] (Supplementary Table 9, online only, available at www.arthroscopyjournal.org). The CONSORT questionnaire (2010) is a 25-item questionnaire with 12

Table 1. Study, Subject, and Surgical Technique Demographic Data

	Data
No. of studies	194
Level I evidence [n (%)]	16 (8.2)
Level II evidence [n (%)]	14 (7.2)
Level III evidence [n (%)]	17 (8.8)
Level IV evidence [n (%)]	147 (76)
ACI*	121
OATS*	34
OC allograft transplant*	16
Microfracture*	41
Studies that self-reported level of evidence [n (%)]	94 (48)
Studies for which reader determined level of evidence [n (%)]	100 (52)
RCTs [n (%)]	16 (9.0)
Non-RCTs [n (%)]	178 (91)
Single-center studies [n (%)]	163 (84)
Multicenter studies [n (%)]	31 (16)
Financial COI reported as present [n (%)]	50 (26)
Financial COI reported as absent [n (%)]	65 (34)
Financial COI presence/absence not reported [n (%)]	79 (40)
Studies that reported presence/absence of complications [n (%)]	102 (53)
Studies that did not report presence/absence of complications [n (%)]	92 (47)
No. of subjects	11,787
ACI [n (%)]	7,086 (60)
OATS [n (%)]	1,994 (17)
OC allograft transplant [n (%)]	411 (3.5)
Microfracture [n (%)]	2,296 (19)
Mean age of subjects (± SD) (yr)	33.5 ± 8.2
Mean length of follow-up (± SD) (yr)	3.7 ± 2.3

COI, conflict of interest; OC, osteochondral.
*Because several studies reported on more than 1 surgical technique, the sum of the total number of studies reflects this, and not 194, the actual number of studies reviewed.

sub-items, for a total of 37 questions, scored 1, 2, or 3; the total score can range from 37 to 111. A percent score is calculated based on the responses to each item. The Jadad scale is a 3-question test evaluating study randomization, blinding, and withdrawals/dropouts, with a score ranging from 0 to 5. Each item on the CLEAR-NPT and Delphi List was scored +1, 0, or −1; each item on the Cochrane Bone, Joint and Muscle Trauma Group Quality Assessment Tool and the Detsky Quality Assessment Scale was scored 0, 1, or 2; and items on the MCMS were scored from 0 to 9. The Detsky Quality Assessment Scale denominator was either 20 or 21 depending on the study's reported outcomes being positive or negative. The measured score was then converted to a percent. These questionnaires were intended to quantify the methodologic quality of randomized controlled studies.

Descriptive statistics were calculated when applicable. All continuous variable data were reported as the mean ± standard deviation. All categorical variable data were reported as the frequency with the percentage. Comparison of groups' quality scores were made by use of the 2-tailed Student *t* test, with the assumption of equal variance. When applicable, group sample populations were compared by use of a 2-proportion Z test with an α of .05. Linear regression analysis was used to test for significance of improvement in study methodologic quality over the period of the included studies. Study methodologic quality with the CMS (inclusive studies published between June 25, 2004, and present) was compared with the CMS results from Jakobsen et al.[7] in 2005 (inclusive studies published between 1985 and June 24, 2004). Kolmogorov-Smirnov and Shapiro-Wilk analyses (Gaussian distribution of data) for study methodologic quality scores showed normality (>0.05). For all statistical analyses, $P < .05$ was statistically significant. SPSS software, version 18.0 (IBM, Armonk, NY), was used for statistical analysis.

Results

There were 194 studies identified overall (11,787 subjects). Most studies (76%) were Level IV evidence, with Level I, II, and III evidence being nearly equally represented (8%, 7%, and 9%, respectively) (Table 1). Nearly half (48%) of all studies self-reported the study's level of evidence. Study level of evidence showed no significant improvement with time ($P = .839$) (Fig 2A). There were 16 RCTs (9%). The methodologic quality of RCTs was significantly greater than that of non-RCTs, assessed by level of evidence, CMS, and QAT score ($P < .001$) (Table 2).

There were 31 multicenter studies (16%). ACI was the most represented technique in the literature (used in 62% of 194 studies), followed by microfracture (21%), OAT (18%), and osteochondral allograft transplant (8%). The mean length of study follow-up was 3.7 ± 2.3 years. The mean subject age was 33.5 ± 8.2 years. Forty percent of studies failed to report the presence or absence of a financial conflict of interest, whereas 26% reported its presence. With later publication dates, the number of studies that failed to report the presence or absence of a financial conflict of interest decreased ($P = .014$) (Fig 2B). During this time, the number of studies that reported the presence of a financial conflict of interest increased ($P = .017$).

Coleman Methodology Score

The overall CMS for the reviewed studies was 50.8 ± 10 (poor rating) (Table 2). The CMSs for all individual techniques received poor ratings (<55). Significant differences were identified in comparison of techniques: ACI and microfracture studies had significantly greater CMSs than osteochondral allograft transplant ($P = .0002$ and $P = .0034$, respectively). The overall CMS (after June 24, 2004) was significantly higher

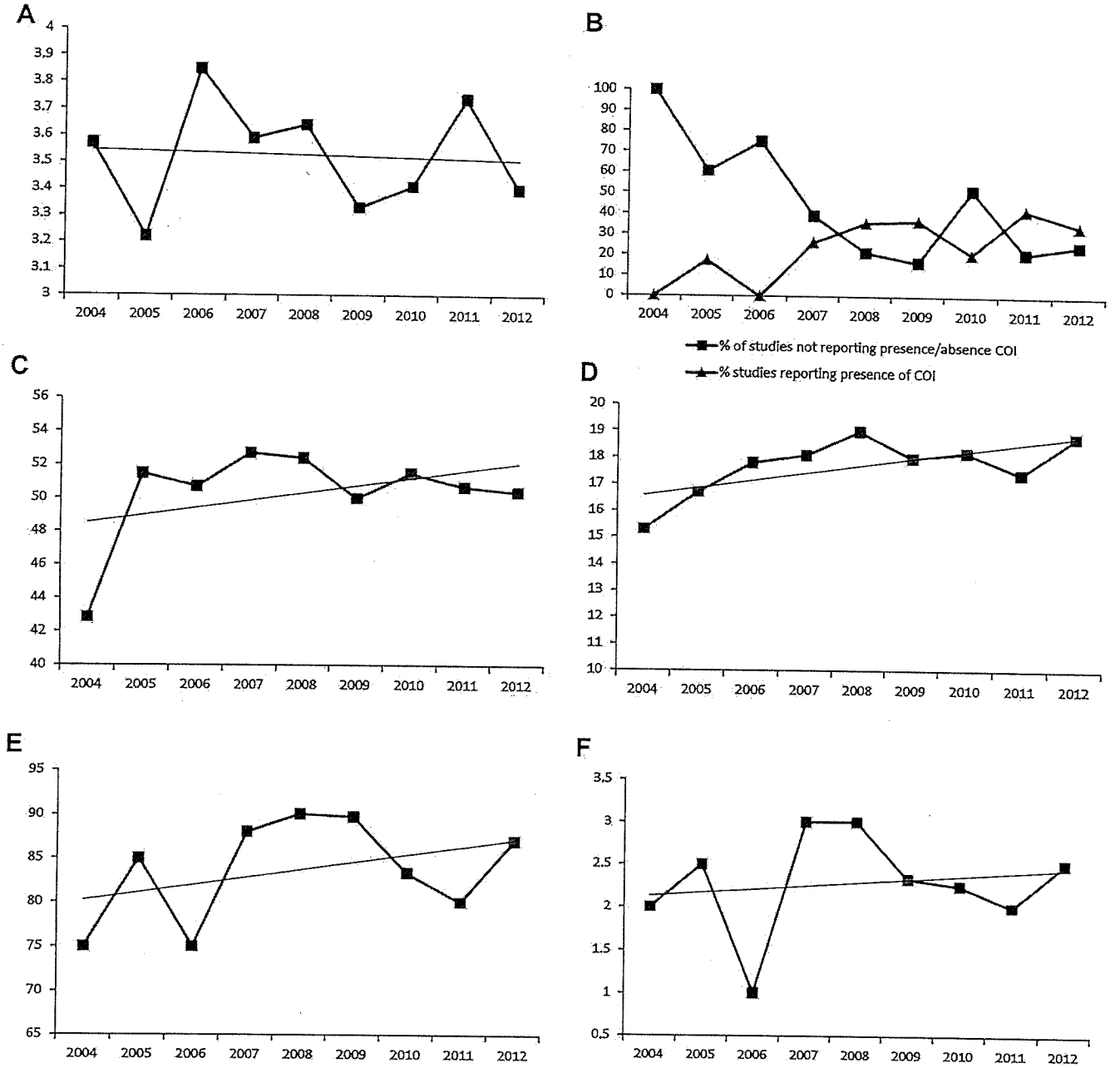

Fig 2. (A) Level of evidence of all cartilage surgery studies per year. The y-axis shows the level of evidence, with lower levels at the top and higher levels at the bottom. The change in level of evidence over time was not significantly different ($P = .839$). (B) With later publication dates, the number of studies that failed to report the presence or absence of a financial conflict of interest (COI) decreased ($P = .014$). The number of studies that reported the presence of a financial COI increased during the same period ($P = .017$). (C) CMS of all cartilage surgery studies per year. (D) QAT score of all cartilage surgery studies per year. (E) CONSORT score of all cartilage surgery studies per year. (F) Jadad scale of all cartilage surgery studies per year.

than the CMS obtained from Jakobsen et al.[7] in 2005 (before June 24, 2004) (50.8 *v* 43.5, $P < .01$) (Fig 3). The overall CMS did not significantly improve over the period of the included studies ($P = .325$) (Fig 2C). The weakest items on the CMS for the studies analyzed were (1) subject selection process, (2) type of study, (3) outcome measures, and (4) outcome assessment. Strengths of the analyzed studies included the numbers and descriptions of surgical procedures.

Quality Appraisal Tool

The overall QAT score for the reviewed studies was 17.9 ± 2.7 (75% quality rating) (Table 2). In the comparison of different techniques, ACI had

Table 2. Overall Quality Scores and Scores Based on Individual Surgical Techniques

Mean Quality Score	Mean ± SD
Overall score	
CMS	50.8 ± 10
QAT	17.9 ± 2.7
CONSORT	75.2 ± 6.4
Jadad	2.06 ± 0.6
ACI	
Level of evidence	3.40 ± 1.02
CMS	52.9 ± 9.27
QAT	18.3 ± 2.27
CONSORT	85.6 ± 6.44
Jadad	2.45 ± 0.51
OAT	
Level of evidence	3.46 ± 1.11
CMS	49.6 ± 11.8
QAT	17.3 ± 2.87
CONSORT	81.8 ± 4.15
Jadad	2.0 ± 0.71
Osteochondral allograft transplant	
Level of evidence	3.88 ± 0.49
CMS	42.9 ± 12.5
QAT	15.7 ± 4.65
CONSORT	NA
Jadad	NA
Microfracture	
Level of evidence	3.0 ± 1.25
CMS	53.4 ± 11.0
QAT	18.4 ± 1.81
CONSORT	86.7 ± 7.37
Jadad	2.44 ± 0.50
RCTs	
Level of evidence	1.1 ± 0.24
CMS	65.8 ± 6.75
QAT	18.8 ± 2.11
CONSORT	84.6 ± 6.40
Jadad	2.2 ± 0.60
CLEAR-NPT	1.41 ± 2.43
Delphi List	0.47 ± 1.23
Cochrane BJMTG score	14.2 ± 2.84
Detsky Quality Assessment Scale	75.6% ± 8.01%
MCMS	54.0 ± 6.91
Non-RCTs	
Level of evidence	3.8 ± 0.58
CMS	49.4 ± 9.13
QAT	17.8 ± 2.75
CONSORT	NA
Jadad	NA

BJMTG, Bone, Joint, Muscle and Trauma Group; NA, not available.

a significantly greater QAT score than OAT ($P = .028$) and osteochondral allograft transplant ($P < .01$); microfracture had a significantly greater QAT score than OAT ($P = .0029$) and osteochondral allograft transplant ($P = .043$). The overall QAT score improved with time ($P = .05$) (Fig 2D). The weakest items on the QAT for the studies analyzed were (1) sample size calculation/justification, (2) valid conclusions and clinical recommendations, (3) specific hypotheses, and (4) data presented for each hypothesis. The strengths of the analyzed studies included thorough literature reviews to define the research questions and standardized measurement techniques.

CONSORT Score

The overall CONSORT score for the reviewed randomized controlled studies was 75.2 ± 6.4 (Table 2). In the comparison of different techniques, there were no significant differences in CONSORT scores. The overall CONSORT score showed no significant improvement with time ($P = .291$) (Fig 2E). The weakest items on the CONSORT score for the studies analyzed were (1) study/trial registration, (2) blinding, (3) sample size determination, (4) ancillary analysis (distinguishing prespecified *v* exploratory), (5) trial protocol, (6) why trial ended or stopped, and (7) presentation of outcome effect sizes. Strengths of the analyzed randomized studies included appropriate assignment of randomly selected subjects and analysis for the primary outcome and appropriate statistical methods to compare groups for primary and secondary outcomes.

Jadad Scale

The overall Jadad score for the reviewed randomized controlled studies was 2.06 ± 0.6 (Table 2). The overall Jadad score showed no significant improvement with time ($P = .641$) (Fig 2F). In the comparison of different techniques, there were no significant differences in overall Jadad scores or any of its individual questions. Given the nature of the surgical intervention (1-stage *v* 2-stage surgery) and the second item on the Jadad scale (double blinding), all "appropriate double blinding" responses were scored 0.

Other Methodologic Quality Scores Used for RCTs

Study weaknesses identified by the CLEAR-NPT were (1) care provider blinding, (2) participant blinding, and (3) adequate concealment of treatment allocation; those identified by the Delphi List were (1) care provider blinding, (2) patient blinding, and (3) adequate treatment allocation concealment; those identified by the Cochrane Bone, Joint and Muscle Trauma Group were (1) provider blinding to assignment status, (2) participant blinding to assignment status after allocation, and (3) inadequate treatment assignment concealment before allocation; those identified by the Detsky Quality Assessment Scale were (1) treatment group concealed to investigator, (2) assessors blinded to treatment, and (3) number of patients excluded and reasons why; and those identified by the MCMS were (1) blinding, (2) clinical effect measurement and number needed—to—treat analysis, and (3) description of inclusion criteria with enrollment percentages.

Discussion

The purposes of this systematic review were to characterize the methodologic quality of all knee articular

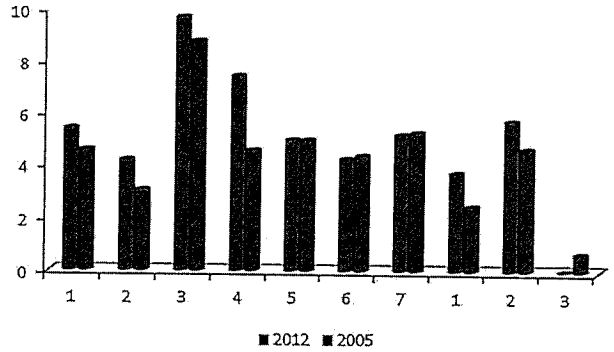

Fig 3. Comparison of individual items of CMS (7 items in part A and 3 items in part B) from publications dated before and after June 24, 2004.

cartilage studies in the literature, to identify strengths and weaknesses, and to evaluate for a temporal relation of study quality. We hypothesized that the quality of evidence has been improving with time but that it would be inherently limited in certain aspects of study design and reporting because of the use of inappropriate questionnaires.

Our hypotheses were partially confirmed. Study quality was low overall and also for individual techniques (ACI, OAT, osteochondral allograft transplant, and microfracture). However, in comparison to Jakobsen et al.[7] (before June 24, 2004), study quality (as measured by the CMS) (after June 24, 2004) significantly improved with time. Study level of evidence, CMS, CONSORT score, and Jadad score, however, did not significantly improve over the period of included studies. QAT was the only methodologic quality score that showed a significant ($P = .05$) improvement over this period. The quality of RCTs was significantly higher than that of non-RCTs ($P < .001$). The same items that were deficient for Jakobsen et al. were also poorly performed or reported in this article (Fig 3). The instruments used to make the assessment of quality essentially create a pseudo-ceiling effect on study quality, limiting the validity and responsiveness of the instruments. Certain questions in the current methodologic scoring systems relevant to articular cartilage are insufficiently detailed to adequately assess true differences between studies. These instruments are generic and may be inappropriate for certain orthopaedic conditions, such as articular cartilage disease of the knee. This is because certain study design parameters, such as care provider and/or participant blinding (a common weakness among all questionnaires used), are impossible to be performed ethically (open *v* arthroscopic, 1-stage *v* 2-stage surgery) in surgical intervention studies of this type.

Future studies for articular cartilage surgery should strive for higher levels of evidence, which should be more convincing to clinicians attempting to solve a clinical problem.[23] Although the best evidence is a high-quality RCT, sometimes an RCT is impossible (perhaps because a true control group does not exist, as in articular cartilage surgery) and other levels of evidence are of significant value. The presence of a financial conflict of interest may present a bias to the reader that, even when disclosed, may affect outcome interpretation. However, industry funding is often needed to support orthopaedic research, given public and government constraints of the monetary amount available for distribution and utilization. Nevertheless, the level of evidence of industry-funded research is lower than that found in non–industry-funded research or investigations funded by government or public sources.[24] Presentations of authors with financial conflicts of interest have been found to more likely describe positive findings.[25] Publications with financial conflicts of interest present, however, have contributed to the increase in negative outcomes of studies reported in the literature with later publication dates.[26] Over a period of 12 years, a significant trend from positive to negative study outcomes was noted whereas, concurrently, a similar significantly negative trend was seen in studies written by authors with financial disclosures.[26] In this systematic review, the number of studies that adequately reported either the presence or absence of a financial conflict of interest significantly increased, reflecting journal editors' and study authors' recognition of the impact that these conflicts may have on study outcomes. It also must be recognized that the number of studies that reported the presence of a financial conflict of interest also increased, showing the necessary importance of external funding (either public or private) for trials of potentially costly medical treatments and devices with high levels of evidence. Therefore, in the future, the quality of the articular cartilage literature must continue to improve so that effective and efficient technologic innovations are economically justified, are readily available to the patient, and provide value to the patient.

Sample size calculation (power analysis) is necessary to determine the minimum sample size needed so that an investigation is powered to detect an effect of a given size. Of the 16 randomized trials analyzed in this review, only 5 (31%) performed an appropriate power analysis based on validated cartilage outcome scores. Study transparency on patient inclusion and exclusion criteria for enrollment and analysis is an important component of all studies, including cartilage research. However, this was a significant weakness in multiple questionnaires in this review. Length of clinical follow-up is a key component of the assessment of outcome of articular cartilage surgery, because it reflects the durability of the intervention. However, the mean clinical follow-up of included studies was 3.7 years. Many questionnaires (e.g., MCMS) report 2 years as "long-term follow-up." However, for ACI, it takes 2 years for

Table 3. Different Outcome Measures Used in All Studies

Outcome Score	n
IKDC	75
KOOS	36
Lysholm	79
Cincinnati/modified Cincinnati	49
KSS	8
HSS	6
Brittberg	5
WOMAC	5
Tegner	51
Marx	3
SF-36/SF-12	29
Patient satisfaction questionnaire	16
MRI postoperatively	72

NOTE. Of the studies, 136 (70%) used at least 1 validated, reliable, and responsive score (IKDC, KOOS, Lysholm) for articular cartilage of the knee.

HSS, Hospital for Special Surgery score; IKDC, International Knee Documentation Committee; KOOS, Knee Injury and Osteoarthritis Outcome Score; KSS, Knee Society Score; MRI, magnetic resonance imaging; SF, Short Form; WOMAC, Western Ontario and McMaster Universities Arthritis Index.

tissue maturation to occur based on previous investigations using second-look arthroscopy and biopsy.[27,28] Furthermore, it can take athletes up to 2 years to recover after ACI.[29] It is also known that fibrocartilage wear and durability can lead to declines in clinical outcome beginning as early as 18 to 24 months after microfracture.[30-33] Therefore this assessment needs to be adjusted in the future, because 2 years is clearly insufficient for defining "long-term" follow-up. This further underscores the "pseudo-ceiling" effect of articular cartilage studies analyzed in this review (173 of 194 studies had >2 years' follow-up) using the current scoring systems, thus receiving the maximum score possible. Prior knee surgery has the potential to affect current surgical outcomes. This has been shown for several procedures, including meniscectomy,[34,35] prior marrow-stimulation techniques,[36] cruciate ligament reconstruction,[37] and high tibial osteotomy.[38] Concomitant surgery also may significantly influence cartilage surgery outcomes.[9,39,40] Most studies included permitted both prior and concomitant surgeries with articular cartilage surgery, thus confounding the effect of the intervention performed.

Outcomes for tibiofemoral and patellofemoral chondral surgery are unique and should be separately reported in studies,[29,41,42] as are outcomes of chondral and osteochondral defects.[9,10,41,43] Using outcome measures with appropriate psychometric properties (reliability, validity, responsiveness) relative to knee articular cartilage allows for the best assessment of the true outcome caused by the intervention. These outcome measures are the International Knee Documentation Committee subjective score, Knee Injury and Osteoarthritis Outcome Score, and Lysholm knee score.[44] There were 136 studies (70%) analyzed in this review that used at least 1 of these scores (Table 3). Given that patient satisfaction rates are increasingly used to justify physician and hospital organization reimbursement, this item is important to consider.[6] However, this was only assessed in 16 studies (8.2%). An independent observer is necessary to reduce the likelihood of detection bias. However, this was only assessed in 15 studies (7.7%). The reporting of complications and reoperations is essential because of the high rate of reoperations after cartilage surgery[9,40] but was only reported in 53% of the studies analyzed in this review. The goal of articular cartilage surgery is to reduce pain, improve function, and attempt to prevent or slow down the progression to osteoarthritis. This assessment is easily made by radiography; however, fewer than 20% of the studies analyzed reported this outcome. Beyond radiographs, advanced magnetic resonance imaging techniques have the ability to detect cartilage macromolecular structure and physiology both before and after surgery, allowing for a reliable, noninvasive method of assessment of repair tissue that correlates with clinical outcome.[4,45-47] All of the aforementioned weaknesses identified can serve as a general guide to assist in study design, conduct, and reporting.

Limitations

Limitations of this systematic review are relegated to the intrinsic parameters of the studies that it analyzed. There was a selection bias present in that we selected only studies published after June 24, 2004. This, however, was intentional so as to draw a comparison between the findings of Jakobsen et al.[7] and our article. The current studies show a significant improvement in methodologic quality versus that of Jakobsen et al. However, study quality over the current period analyzed (2004 to 2012) did not significantly improve. The same methodologic weaknesses present before 2004 (blinding, type of study, outcome criteria, outcome assessment, and subject selection process) were still observed in the current set of investigations analyzed in this study. The nature of articular cartilage surgery does not allow for a true control group, presents an inability to be blinded (2-stage *v* 1-stage surgery and surgeon who must perform the procedure), and has significant heterogeneity in subjects, defects, knees, surgical techniques (performance bias), outcome measures (detection bias), and outcome follow-up (transfer bias). These limitations are reported in the current investigation as inherent limitations of the studies that were analyzed. Another limitation is that there was no attempt to correlate study methodologic quality with clinical outcome. Although we recognize that no gold standard methodologic quality score exists, the CONSORT criteria are well recognized as an expert panel–based checklist of relevant items to guide design

and reporting of RCTs. However, the CONSORT score is not intended to measure and calculate a numerical score for "grading" the quality of a study. Instead, we chose several other commonly used study methodologic quality scores to achieve this purpose.

Conclusions

The methodologic quality of knee articular cartilage surgery studies was poor overall and also for individual techniques (ACI, OAT, osteochondral allograft transplant, and microfracture). However, the overall quality of the investigations in this review (after June 2004) has significantly improved in comparison to those published before 2004. The quality of RCTs was significantly higher than that of non-RCTs. Level of evidence, CMS, CONSORT score, and Jadad score did not significantly improve with later publication date within the period of the studies analyzed. Methodologic quality deficiencies identified in this investigation may be used to guide future articular cartilage studies' design, conduct, and reporting.

References

1. Curl W, Krome J, Gordon E, Rushing J, Smith B, Poehling G. Cartilage injuries: A review of 31,516 knee arthroscopies. *Arthroscopy* 1997;13:456-460.
2. Widuchowski W, Widuchowski J, Trzaska T. Articular cartilage defects: Study of 25,124 knee arthroscopies. *Knee* 2007;14:177-182.
3. Behery O, Karnes J, Siston R, Harris J, Flanigan D. Factors influencing the outcome of ACI: A systematic review. *J Knee Surg.* 12 November, 2012. [Epub ahead of print.]
4. Harris JD, Brophy RH, Jia G, et al. Sensitivity of magnetic resonance imaging for detection of patellofemoral articular cartilage defects. *Arthroscopy* 2012;28:1728-1737.
5. Heir S, Nerhus T, Rotterud J, et al. Focal cartilage defects in the knee impair quality of life as much as severe osteoarthritis. *Am J Sports Med* 2010;38:231-237.
6. CfMMS (CMS [Center for Medicare & Medicaid Services]). Hospital care quality information from the consumer perspective. Available from: www.hcahpsonline.org. Accessed May 6, 2012.
7. Jakobsen R, Engebretsen L, Slauterbeck J. An analysis of the quality of cartilage repair studies. *J Bone Joint Surg Am* 2005;87:2232-2239.
8. Worthen J, Waterman CP, Davidson PA, Lubowitz JH. Limitations and sources of bias in clinical knee cartilage research. *Arthroscopy* 2012;28:1315-1325.
9. Harris JD, Siston RA, Brophy RH, Lattermann C, Carey JL, Flanigan DC. Failures, re-operations, and complications after autologous chondrocyte implantation—A systematic review. *Osteoarthritis Cartilage* 2011;19:779-791.
10. Harris JD, Siston RA, Pan X, Flanigan DC. Autologous chondrocyte implantation: A systematic review. *J Bone Joint Surg Am* 2010;92:2220-2233.
11. Griesser MJ, Harris JD, Campbell JE, Jones GL. Adhesive capsulitis of the shoulder: A systematic review of the effectiveness of intra-articular corticosteroid injections. *J Bone Joint Surg Am* 2011;93:1727-1733.
12. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *J Clin Epidemiol* 2009;62:1006-1012.
13. Obremskey W, Pappas N, Attallah-Wasif E, Tornetta P, Bhandari M. Levels of evidence in orthopaedic journals. *J Bone Joint Surg Am* 2005;87:2632-2638.
14. Coleman B, Khan K, Maffulli N, Cook J, Wark J. Studies of surgical outcome after patellar tendinopathy: Clinical significance of methodological deficiencies and guidelines for future studies. *Scand J Med Sci Sports* 2000;10:2-11.
15. Tallon C, Coleman B, Khan K, Maffulli N. Outcome of surgery for chronic Achilles tendinopathy. A critical review. *Am J Sports Med* 2001;29:315-320.
16. Roy J, MacDermid J, Woodhouse L. Measuring shoulder function: A systematic review of four questionnaires. *Arthritis Rheum* 2009;61:623-632.
17. Trials TCG-TRo (The CONSORT Group - Transparent Reporting of Trials). The CONSORT statement. January 20, 2012. Available from: http://www.consort-statement.org/consort-statement/overview0/#checklist. Accessed November 3, 2012.
18. Jadad A, Enkin M, eds. *Randomized controlled trials: Questions, answers, and musings.* Ed 2. Hoboken, New Jersey: Blackwell, 2007.
19. Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA. Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992;45:255-265.
20. Cochrane Collaboration. Cochrane Bone, Joint and Muscle Trauma Group. Resources for developing a review. Available from: http://bjmtg.cochrane.org/sites/bjmtg.cochrane.org/files/uploads/Quality%20assessment%20tool.pdf. Accessed January 8, 2013.
21. Verhagen A, deVet H, deBie R, 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;51:1235-1241.
22. Boutron I, Moher D, Tugwell P, et al. A checklist to evaluate a report of a nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol* 2005;58:1233-1240.
23. Wright J, Swiontkowski M, Heckman J. Introducing levels of evidence to the journal. *J Bone Joint Surg Am* 2003;85:1-3.
24. Noordin S, Wright JG, Howard A. Relationship between declared funding support and level of evidence. *J Bone Joint Surg Am* 2010;92:1647-1651.
25. Okike K, Kocher MS, Mehlman CT, Bhandari M. Conflict of interest in orthopaedic research. An association between findings and funding in scientific presentations. *J Bone Joint Surg Am* 2007;89:608-613.
26. Hsu J, Liu S, Lee G. Can we trust studies published by authors with financial conflicts? Using the decline of metal-on-metal total hip arthroplasty to investigate. *J Arthroplasty* 2012;27(suppl):41-45.e1.
27. Gikas P, Morris T, Carrington R, Skinner J, Bentley G, Briggs T. A correlation between the timing of biopsy after autologous chondrocyte implantation and the histological appearance. *J Bone Joint Surg Br* 2009;91:1172-1177.

28. Brun P, Dickinson SC, Zavan B, Cortivo R, Hollander AP, Abatangelo G. Characteristics of repair tissue in second-look and third-look biopsies from patients treated with engineered cartilage: Relationship to symptomatology and time after implantation. *Arthritis Res Ther* 2008;10:R132.
29. Harris J, Brophy R, Siston R, Flanigan D. Treatment of chondral defects in the athlete's knee. *Arthroscopy* 2010;26:841-852.
30. Gudas R, Stankevicius E, Monastyreckiene E, Pranys D, Kalesinkskas R. Osteochondral autologous transplantation versus microfracture for the treatment of articular cartilage defects in the knee joint in athletes. *Knee Surg Sports Traumatol Arthrosc* 2006;14:834-842.
31. Mithoefer K, Williams R, Warren R, Wickiewicz T, Marx R. High-impact athletics after knee articular cartilage repair: A prospective evaluation of the microfracture technique. *Am J Sports Med* 2006;34:1413-1418.
32. Gobbi A, Nunag P, Malinowski K. Treatment of full thickness chondral lesions of the knee with microfracture in a group of athletes. *Knee Surg Sports Traumatol Arthrosc* 2005;13:213-221.
33. Blevins F, Steadman J, Rodrigo J, Silliman J. Treatment of articular cartilage defects in athletes: An analysis of functional outcome and lesions appearance. *Orthopedics* 1998;21:761-767.
34. Fauno P, Nielsen AB. Arthroscopic partial meniscectomy: A long-term follow-up. *Arthroscopy* 1992;8:345-349.
35. Fairbank T. Knee joint changes after meniscectomy. *J Bone Joint Surg Br* 1948;30:664-670.
36. Minas T, Gomoll A, Rosenberger R, Royce R, Bryant T. Increased failure rate of autologous chondrocyte implantation after previous treatment with marrow stimulation techniques. *Am J Sports Med* 2009;37:902-908.
37. Li RT, Lorenz S, Xu Y, Harner CD, Fu FH, Irrgang JJ. Predictors of radiographic knee osteoarthritis after anterior cruciate ligament reconstruction. *Am J Sports Med* 2011;39:2595-2603.
38. Gomoll AH. High tibial osteotomy for the treatment of unicompartmental knee osteoarthritis: A review of the literature, indications, and technique. *Phys Sportsmed* 2011;39:45-54.
39. Gomoll AH, Kang RW, Chen AL, Cole BJ. Triad of cartilage restoration for unicompartmental arthritis treatment in young patients: Meniscus allograft transplantation, cartilage repair and osteotomy. *J Knee Surg* 2009;22:137-141.
40. Harris JD, Cavo M, Brophy R, Siston R, Flanigan D. Biological knee reconstruction: A systematic review of combined meniscal allograft transplantation and cartilage repair or restoration. *Arthroscopy* 2011;27:409-418.
41. Harris J, Flanigan D. Management of knee articular cartilage injuries. In: Dragoo J, ed. *Modern arthroscopy*. New York: InTech, 2011;103-128.
42. Brittberg M, Lindahl A, Nilsson A, Ohlsson C, Isaksson O, Peterson L. Treatment of deep cartilage defects in the knee with autologous chondrocyte transplantation. *N Engl J Med* 1994;331:889-895.
43. Trinh TQ, Harris JD, Flanigan DC. Surgical management of juvenile osteochondritis dissecans of the knee. *Knee Surg Sports Traumatol Arthrosc* 2012;20:2419-2429.
44. Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J Knee Surg* 2011;23:137-151.
45. Braun HJ, Gold GE. Advanced MRI of articular cartilage. *Imaging Med* 2011;3:541-555.
46. Quatman CE, Quatman-Yates CC, Schmitt LC, Paterno MV. The clinical utility and diagnostic performance of MRI for identification and classification of knee osteochondritis dissecans. *J Bone Joint Surg Am* 2012;94:1036-1044.
47. Quatman CE, Hettrich CM, Schmitt LC, Spindler KP. The clinical utility and diagnostic performance of magnetic resonance imaging for identification of early and advanced knee osteoarthritis: A systematic review. *Am J Sports Med* 2011;39:1557-1568.

Supplementary Table 1. Coleman Methodology Score

Part A

1. Study size (10)
2. Mean duration of follow-up (5)
3. Number of surgical procedures (10)
4. Type of study (15)
5. Diagnostic certainty (5)
6. Description of surgical procedure (5)
7. Description of postoperative rehabilitation (10)

Part B

1. Outcome measures (10)
2. Outcome assessment (15)
3. Selection process (15)

NOTE. Part A (60) plus part B (40) equals the total score (100).

Supplementary Table 2. Quality Appraisal Tool

1. Thorough literature review to define research questions
2. Specific inclusion and exclusion criteria
3. Specific hypotheses
4. Appropriate scope of psychometric properties
5. Sample size calculation/justification
6. Appropriate retention/follow-up
7. Authors referenced specific procedures for administration, scoring, and interpretation of procedures
8. Measurement techniques were standardized
9. Data presented for each hypothesis
10. Appropriate statistical point estimates
11. Appropriate statistical error estimates
12. Valid conclusions and clinical recommendations

NOTE. Each item is scored as 0, 1, or 2. A score of 2 is assigned for adequate and complete reporting of item. A score of 0 is assigned for not reporting the item or for providing an inadequate description. A score of 1 is assigned for anything in between. The score range was from 0 to 24, with a percent score assigned, giving a "quality rating."

Supplementary Table 3. CONSORT 2010

Title and abstract

1. Identification as a randomized trial in the title
2. Structured summary of trial design, methods, results, and conclusions

Introduction

3. Scientific background and explanation of rationale
4. Specific objectives or hypotheses

Methods

5. Description of trial design (such as parallel, factorial) including allocation ratio
6. Important changes to methods after trial commencement (such as eligibility criteria), with reasons
7. Eligibility criteria for participants
8. Settings and locations where the data were collected
9. The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
10. Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed
11. Any changes to trial outcomes after the trial commenced, with reasons
12. How sample size was determined
13. When applicable, explanation of any interim analyses and stopping guidelines
14. Method used to generate the random allocation sequence
15. Type of randomization; details of any restriction (such as blocking and block size)
16. Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
17. Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
18. If performed, who was blinded after assignment to interventions (e.g., participants, care providers, or those assessing outcomes) and how
19. If relevant, description of the similarity of interventions
20. Statistical methods used to compare groups for primary and secondary outcomes
21. Methods for additional analyses, such as subgroup analyses and adjusted analyses

Results

22. For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome
23. For each group, losses and exclusions after randomization, together with reasons
24. Dates defining the periods of recruitment and follow-up
25. Why the trial ended or was stopped
26. A table showing baseline demographic and clinical characteristics for each group
27. For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
28. For each primary and secondary outcome, results for each group, as well as the estimated effect size and its precision (such as 95% confidence interval)
29. For binary outcomes, presentation of both absolute and relative effect sizes is recommended
30. Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory
31. All important harms or unintended effects in each group

Discussion

32. Trial limitations, addressing sources of potential bias, imprecision, and if relevant, multiplicity of analyses
33. Generalizability (external validity, applicability) of the trial findings
34. Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence

Other information

35. Registration number and name of trial registry
36. Where the full trial protocol can be accessed, if available
37. Sources of funding and other support (such as supply of drugs), role of funders

NOTE. A score of 1, 2, or 3 is assigned. A score of 3 is assigned for adequate and complete reporting of item. A score of 1 is assigned for not reporting the item or for providing an inadequate description. A score of 2 is assigned for anything in between. The score range was 37 to 111, with a percent score assigned.

Supplementary Table 4. Jadad Scale

1. Was the study described as randomized?
   a. If yes, 1 point given.
   b. Additional point if method of randomization described and that method was appropriate.
      i. If method of randomization inappropriate, 1 point deducted.
2. Was the study described as double blind?
   a. If yes, 1 point given.
   b. Additional point if method of blind described and that method was appropriate.
      i. If method of blinding inappropriate, 1 point deducted.
3. Was there a description of withdrawals and dropouts?
   a. If yes, 1 point given.

NOTE. The score ranges from 0 to 5. Even though the Jadad scale is a 3-question scale, a maximum of 5 points is possible.

Supplementary Table 5. Modified Coleman Methodology Score

	Score
1. Inclusion criteria	
Not described	0
Described without percents given	3
Enrollment rate <80%	6
Enrollment rate >80%	9
2. Power	
Not reported	0
>80%, methods not described	3
>80%, methods described	6
3. α Error	
Not reported	0
<.05	3
<.01	6
4. Sample size	
Not stated or <20	0
20-40	3
41-60	6
>60	9
5. Randomization	
Not randomized	0
Modified/partial	
Not blinded	2
Blinded	4
Complete	
Not blinded	6
Blinded	8
6. Follow-up	
Short-term (<6 mo)	
Patient retention <80%	0
Patient retention 80%-90%	2
Patient retention >90%	4
Medium-term (6-24 mo)	
Patient retention <80%	2
Patient retention 80%-90%	4
Patient retention >90%	6
Long-term (>24 mo)	
Patient retention <80%	4
Patient retention 80%-90%	6
Patient retention >90%	8
7. Patient analysis	
Incomplete	0
Complete	3
Complete and intention-to-treat based	6
8. Blinding	
None	0
Single	2
Double	4
Triple	6
9. Similarity in treatment	
No	0
Similar co-interventions	3
No co-interventions	6
10. Treatment description	
None	0
Fair	3
Adequate	6
11. Group comparability	
Not comparable	0
Partially comparable	3
Comparable	6
12. Outcome assessment	
Written assessment by patient with assistance	0
Written assessment by patient without assistance	2
Independent investigator	4
Recruited patients	6
13. Description of rehabilitation protocol	
Not reported	0
Not adequately described	2
Well described	4
14. Clinical effect measurement	
Effect size	
Not reported	0
<50%	2
50%-75%	4
>75%	6
or Relative risk reduction	
Not reported	0
<25%	3
>25%	6
or Absolute risk reduction	
Not reported	0
<10%	3
>10%	6
15. Number of patients to treat	
Not reported	0
Reported	4

NOTE. A scaled score of 0 to 100 is possible: excellent, 85 to 100; good, 70 to 84; fair, 55 to 69; and poor, less than 55.

Supplementary Table 6. Detsky Quality Assessment Scale

	Score
1. Were patients randomly assigned?	
Yes	1
No	0
2. Was randomization adequate?	
Yes	2
Partly	1
No	0
3. Was the treatment group concealed to the investigator?	
Yes	1
No	0
4. Was the description of outcome measures adequate?	
Yes	1
No	0
5. Were the outcome measures objective?	
Yes	2
Partly	1
No	0
6. Were the assessors blind to the treatment?	
Yes	1
No	0
7. Were the inclusion and exclusion criteria well defined?	
Yes	1
No	0
8. Was the number of patients excluded described and the reason why?	
Yes	2
Partly	1
No	0
9. Was the treatment fully described for the study participants receiving the intervention being investigated?	
Yes	2
Partly	1
No	0
10. Was the treatment fully described for the controls?	
Yes	2
Partly	1
No	0
11. Was a statistical test stated and *P* value reported?	
Yes	1
No	0
12. Was statistical analysis appropriate?	
Yes	2
Partly	1
No	0
13. If trial was negative, were 95% confidence intervals or post hoc power calculations performed?	
Yes	1
No	0
14. Was a sample size calculation performed before the study?	
Yes	1
No	0

NOTE. The total score denominator is out of 20 (if study reports a positive finding) or 21 (if study reports a negative finding [question 13]). The numerator is then summed and a percentage score calculated.

Supplementary Table 7. Cochrane Bone, Joint and Muscle Trauma Group Methodologic Quality Score

	Score
A. Was assigned treatment adequately concealed before allocation?	
Method did not allow disclosure of assignment	2
Small, but possible, chance of disclosure of assignment	1
Quasi-randomized or open tables	0
B. Were outcomes of patients who withdrew described and included in the analysis (intent to treat)?	
Withdrawals well described and accounted for	2
Withdrawals described and analysis not possible	1
No mention, inadequate mention, or obvious differences and no adjustment	0
C. Were outcome assessors blinded to treatment status?	
Effective action taken to blind assessors	2
Small chance of unblinding of assessors	1
Not mentioned, not possible, or not performed	0
D. Were treatment and control groups comparable at entry?	
Good comparability of groups, or confounding adjusted for in analysis	2
Confounding small, mentioned, but not adjusted for	1
Large potential for confounding or not discussed	0
E. Were participants blind to assignment status after allocation?	
Effective action taken to blind participants	2
Small chance of unblinding of participants	1
Not mentioned, not possible, or not performed	0
F. Were providers blind to assignment status after allocation?	
Effective action taken to blind providers	2
Small chance of unblinding of providers	1
Not mentioned, not possible, or not performed	0
G. Were care programs other than trial options identical?	
Clearly identical	2
Clear, but trivial differences	1
Not mentioned or clear and important differences	0
H. Were inclusion and exclusion criteria clearly defined?	
Clearly	2
Inadequately	1
Not defined	0
I. Were interventions clearly defined?	
Clearly, with standard protocol	2
Clearly applied, but application protocol not standardized	1
Poorly defined or not defined	0
J. Were outcome measures clearly defined?	
Clearly	2
Inadequately	1
Not defined	0
K. Were the diagnostic tests used in clinical outcome assessment clinically useful?	
Optimal	2
Adequate	1
Not defined or not adequate	0

(continued)

	Score
L. Was surveillance active and of appropriate clinical duration?	
Active surveillance and appropriate time	2
Inappropriate time	1
Not active surveillance or not defined	0

NOTE. The total score for the 12 questions ranges from 0 to 24.

Supplementary Table 8. Delphi List

1a. Was method of randomization performed?
1b. Was treatment allocation concealed?
2. Were groups similar at baseline regarding most important prognostic indicators?
3. Were eligibility criteria specified?
4. Was outcome assessor blinded?
5. Was care provider blinded?
6. Was patient blinded?
7. Were point estimates and measures of variability provided for primary outcome measures?
8. Did analysis include intent to treat?

NOTE. One point is given for a yes response, 1 point is deducted for a no response, and 0 points are received for a "don't know" response.

Supplementary Table 9. CLEAR-NPT (Checklist to Evaluate a Report of a Non-Pharmacologic Trial)

1. Was generation of allocation sequences adequate?
2. Was treatment allocation concealed?
3. Were details of intervention administered to each group made available?
4. Was care providers' skill or experience in each arm appropriate?
5. Was participant adherence assessed quantitatively?
6. Were participants adequately blinded?
   6.1. If not adequately blinded, were all other co-interventions the same in each randomized group?
   6.2. If not adequately blinded, were withdrawals and patients lost to follow-up the same in each group?
7. Were care providers adequately blinded?
   7.1. If not adequately blinded, were all other co-interventions the same in each randomized group?
   7.2. If not adequately blinded, were withdrawals and patients lost to follow-up the same in each group?
8. Were outcome assessors adequately blinded?
   8.1. If not adequately blinded, were specific methods used to avoid sampling/ascertainment bias?
9. Was follow-up schedule the same in each group?
10. Were main outcomes analyzed according to intent-to-treat analysis?

NOTE. One point is given for a yes response, 1 point is deducted for a no response, and 0 points are received for a "don't know" response.