An Analysis of the Quality of Cartilage Repair Studies

Rune B. Jakobsen, Lars Engebretsen and James R. Slauterbeck


---

This information is current as of April 28, 2006

**Supplementary material**

Commentary and Perspective, data tables, additional images, video clips and/or translated abstracts are available for this article. This information can be accessed at [http://www.ejbjs.org/cgi/content/full/87/10/2232/DC1](http://www.ejbjs.org/cgi/content/full/87/10/2232/DC1)

**Subject Collections**

Articles on similar topics can be found in the following collections

- **Adult Disease** (1286 articles)
- **Cartilage Injury and Repair** (24 articles)

**Reprints and Permissions**

Click here to [order reprints or request permission](http://www.ejbjs.org) to use material from this article, or locate the article citation on *jbjs.org* and click on the [Reprints and Permissions] link.

**Publisher Information**

The Journal of Bone and Joint Surgery

20 Pickering Street, Needham, MA 02492-3157

[www.jbjs.org](http://www.jbjs.org)
An Analysis of the Quality of Cartilage Repair Studies

BY RUNE B. JAKOBSEN, LARS ENGBRETSSEN, MD, PHD, AND JAMES R. SLAUTERBECK, MD, PHD

Investigation performed at the Oslo Sports Trauma Research Center and the Orthopaedic Center, Ullevaal University Hospital, Oslo, Norway

Background: Most lesions of articular cartilage do not heal spontaneously and may lead to secondary osteoarthritis. It is not known whether the optimistic reports on the short and long-term results of several different cartilage repair techniques are based on sound methodological quality.

Methods: We performed a literature search in MEDLINE, CINAHL, the Cochrane Central Register, and EMBASE and included studies in which the primary aim of the investigation was to report the outcome after cartilage repair in the knee with use of microfracture, autologous osteochondral transplantation, autologous periosteal transplantation, or autologous chondrocyte implantation. We scored the quality of the studies using a modified Coleman Methodology Score with ten criteria, which results in a final score between 0 and 100. Studies were also assessed with use of the level-of-evidence rating used in the American Volume of The Journal of Bone and Joint Surgery. We collected data on the year of publication, the reported postoperative results, and the outcome measures used to assess the results.

Results: Sixty-one studies involving a total of 3987 surgical procedures were included. The average methodology score was 43.5 of 100. Methodological deficiencies were found with respect to five criteria: the type of study, description of the rehabilitation protocol, outcome criteria, outcome assessment, and subject selection process. Large variations in the reported outcome were seen within each treatment modality, and no significant differences were found between each kind of therapy (p = 0.11). The methodology score correlated positively with the level-of-evidence rating (r = 0.668, p < 0.0001), but there were large variations in the methodology score within each level. The linear regression analysis weighted by the number of patients demonstrated a negative yet not significant correlation between the methodology score and the results reported in nineteen studies with use of the Lysholm Scale (r = –0.29, p = 0.19). A total of twenty-seven different clinical outcome measurement scales were used to assess outcome.

Conclusions: The generally low methodological quality found in the studies included in this analysis indicates that caution is required when interpreting results after surgical cartilage repair. Firm recommendations on which procedure to choose cannot be given at this time on the basis of these studies. More attention should be paid to methodological quality when designing, performing, and reporting clinical studies.

Level of Evidence: Therapeutic Level III. See Instructions to Authors for a complete description of levels of evidence.

Surgical treatment for articular cartilage injury is of major interest to orthopaedic surgeons because most lesions of articular cartilage do not heal spontaneously and may predispose the joint to the subsequent development of secondary osteoarthritis. In a series of 993 knee arthroscopies performed because of pain, substantial cartilage lesions considered suitable for surgical treatment (those that were >2 cm² in size and grade 3 or 4, according to the system of the International Cartilage Repair Society) were detected in 6% of the patients. Treatment for articular cartilage injuries includes the microfracture technique, autologous periosteal transplantation, autologous osteochondral transplantation, and autologous chondrocyte implantation; however, much controversy surrounds the best treatment option. Numerous published articles, in which the above treatment options were used, have described good or excellent results for a majority of patients. However, several authors have pointed out methodological weaknesses in the published studies.

The purpose of this analysis was to determine whether the optimistic reports in the literature are supported by sound methodological quality in the studies. Our main hypothesis was that the majority of the studies have methodological limitations that may limit the value of the reported results. Our second hypothesis was that studies of lesser methodological quality describe higher rates of success. We addressed methodological limitations by calculating a modified Coleman Methodology Score and level-of-evidence rating and correlated these to the reported results. We also correlated the Coleman Methodology Score with the year of publication to study trends over time. In addition, we collected the different outcome measurement scales used to assess outcome in order to determine the diversity in this area.
Materials and Methods

We modified a search strategy (see Appendix) on the basis of the “optimum trial search strategy” described in the Cochrane Reviewers’ Handbook. Selection criteria included studies in which the primary aim was to report the outcome after cartilage repair in the knee with use of microfracture, autologous osteochondral transplantation, autologous periosteal transplantation, or autologous chondrocyte implantation, including matrix-assisted chondrocyte transplantation. We performed the search in MEDLINE In-Process and other Non-Indexed Citations, EMBASE, and CINAHL using Ovid. We also searched the Cochrane Central Register of Controlled Trials. No language restrictions were applied initially. The searches were performed on June 24, 2004.

A total of 673 abstracts were reviewed with use of the selection criteria. Studies clearly failing to meet the criteria were excluded. We also excluded studies that were in languages other than English, French, and German, as we did not have interpreters available in Chinese (three studies), Spanish (one), Lithuanian (two), or Polish (one). Full-text articles were obtained if the decision to include or exclude could not be made from the abstract.

Full-text versions of the sixty-one studies that were included were reviewed for methodological quality with use of a modified version of the methodology score previously used by Coleman et al. and Tallon et al. (see Appendix). We modified the methodology score in the category describing the rehabilitation protocol because stratification of the score was based largely upon rehabilitation compliance and only one study mentioned compliance. The modification enabled us to assess whether the rehabilitation protocol was satisfactory with regard to (1) describing the immediate postoperative rehabilitation and the rehabilitation before discharge (including the use of continuous passive motion), (2) describing the physiotherapy protocol for twelve weeks postoperatively (the use of weight-bearing and the type of exercises), and (3) giving instructions on when patients could resume sports activities.

The Coleman Methodology Score, which was originally developed for and used to grade clinical studies on patellar and Achilles tendinopathy, assesses methodology with use of ten criteria, giving a total score between 0 and 100. A score of 100 indicates that the study largely avoids chance, various biases, and confounding factors. The subsections that make up the Coleman Methodology Score are based on the subsections of the CONSORT statement (for randomized controlled trials) but are modified to allow for other trial designs. The studies were also assessed with use of the level-of-evidence rating introduced in the American Volume of The Journal of Bone and Joint Surgery in 2003 and recently also in Arthroscopy. The clinical outcome scales used in the selected papers, as well as the reported clinical outcomes at the time of the follow-up when most patients were assessed, were collected. We also collected, if the data were reported, the average value for the different clinical scores and the percentage of patients with a score that equaled an excellent or good clinical outcome as assessed in the study. We correlated the outcome with the total Coleman Methodology Score and with the level-of-evidence rating to assess the impact of methodology on the reported outcomes. We also correlated the Coleman Methodology Score with the year of publication to investigate trends in methodology over time.

Statistical Methods

SPSS software (version 11.0.0; SPSS, Chicago, Illinois) was used to analyze the data. Both parametric (mean and standard deviation) and nonparametric (median and interquartile range) descriptive statistics were used, since some variables were not normally distributed (Shapiro-Wilk test). The Pearson corre-

| TABLE I Coleman Methodology Score for Studies on Surgical Repair of Cartilage |
|---------------------------------|--------|------|------|------|--------|
| **Section Score (Maximum Score)** | **Mean** | **Standard Deviation** | **Range** | **Median** | **25th to 75th Percentile** |
| Part A                           |        |      |      |      |        |
| Study size (10)                  | 4.6    | 3.7  | 0-10 | 4    | 0.7    |
| Mean duration of follow-up (5)   | 3.1    | 2.1  | 0-5  | 5    | 2-5    |
| No. of surgical procedures (10)  | 8.8    | 3.0  | 0-10 | 10   | 10-10  |
| Type of study (15)               | 4.6    | 5.5  | 0-15 | 0    | 0-10   |
| Diagnostic certainty (5)         | 5.0    | 0.0  | 5    | 5    | 5-5    |
| Description of surgical procedure (5) | 4.4 | 1.2  | 0-5  | 5    | 3-5    |
| Description of postoperative rehabilitation (10) | 5.3 | 2.7  | 0-10 | 5    | 5-5    |
| Part B                           |        |      |      |      |        |
| Outcome measures (10)            | 2.5    | 1.5  | 0-5  | 2    | 2-4    |
| Outcome assessment (15)          | 4.7    | 3.2  | 0-12 | 5    | 3-8    |
| Selection process (15)           | 0.7    | 1.7  | 0-5  | 0    | 0.0    |
| Total part A (60)                | 35.7   | 9.3  | 10-52| 37   | 29-43  |
| Total part B (40)                | 7.8    | 4.7  | 0-21 | 7    | 4.5-12 |
| Total score (100)                | 43.5   | 12.5 | 12-73| 44   | 34-53  |
A weighted analysis was used for normally distributed data. We performed tests for linear regression that were weighted and unweighted with respect to the number of patients included in each study. As weighted analysis did not markedly change the results, we report the unweighted estimates when the estimate is not explicitly described as weighted. The Kruskal-Wallis test was used to test whether the outcomes from different kinds of therapy differed significantly.

Results

The sixty-one studies that were included described a total of 3987 operations, of which only 260 were from randomized controlled trials. The median number of patients included in the studies was thirty, with the 25th to 75th percentile ranging from fourteen to fifty-eight patients. The median duration of follow-up, which was stated clearly in forty studies, was thirty-three months.

The majority of studies had methodological limitations, with an average Coleman Methodology Score of 43.5 (95% confidence interval, 40.3 to 46.7) (all studies are listed in the Appendix with the total Coleman Methodology Score, study size, and type of therapy). There were especially low scores in the following five categories: (1) type of study, (2) description of postoperative rehabilitation, (3) outcome criteria, (4) outcome assessment, and (5) subject selection process. The average Coleman Methodology Score for each criterion and the total Coleman Methodology Score are given in Table I. The distribution of the studies with regard to type of treatment, type of study, and level-of-evidence rating is given in Table II.

When outcome results were analyzed with respect to the type of therapy (excluding randomized controlled trials), no significant differences among the reported outcomes were detected either with respect to the reported Lysholm scores (seventeen studies; \( p = 0.30 \)) or the reported percentages of good or excellent outcome (forty-seven studies; \( p = 0.11 \)). Large variations in reported outcome were demonstrated within each treatment modality (Figs. 1 and 2).

We did not find a significant weighted correlation between methodological quality and reported results when analyzing the reported Lysholm scores (nineteen studies; \( r = -0.29, p = 0.19 \)) or the reported percentages of good or excellent results (fifty studies; \( p = 0.20 \)). The mean average outcome at the time of the last reported follow-up with the Lysholm scale was 86.6 (95% confidence interval, 81.1 to 92.2; nineteen studies), and the mean percentage of patients who had a rating of good or excellent with use of other scales was 86.4% (95% confidence interval, 83.1 to 89.7; fifty studies).

Twenty-seven different scales for clinical outcome mea-

![](https://example.com/figure1.png)

**Fig. 1**

Box plot of the outcome on the Lysholm scale for each type of therapy reported in seventeen studies. Randomized controlled trials are not shown. Each box with bars shows the median, quartiles, and minimum and maximum values. If the minimum or maximum values fall more than 1.5 box lengths from the upper or lower edge of the box, these are instead illustrated by an “O”. The difference in outcome between each type of therapy was not found to be significant (Kruskal-Wallis test, \( p = 0.30 \)). The 95% confidence interval was 78.1 to 96.6 for microfracture (mfx), 85.7 to 95.3 for autologous osteochondral transplantation (OATS), and 67.4 to 99.1 for autologous chondrocyte implantation (ACI). Periost = periosteal transplantation.

![](https://example.com/figure2.png)

**Fig. 2**

Box plot of the outcome percentages in the good and excellent categories for each type of therapy as reported in forty-seven studies. Randomized controlled trials are not shown. Each box with bars shows the median, quartiles, and minimum and maximum values. If the minimum or maximum values fall more than 1.5 box lengths from the upper or lower edge of the box, these are instead illustrated by an “O”. The difference in outcome between each type of therapy was not found to be significant (Kruskal-Wallis test, \( p = 0.11 \)). The 95% confidence interval was 55.9 to 98.0 for periosteal transplantation (periost), 67.3 to 98.5 for microfracture (mfx), 85.8 to 95.2 for autologous osteochondral transplantation (OATS), and 81.4 to 91.3 for autologous chondrocyte implantation (ACI).
measurement were used. None of them had been validated for use in patients with cartilage injuries at the time of the publication of each study. The most frequently reported scale was the one introduced by Lysholm and Gillquist\(^7\) (twenty-eight studies), but the average value at the time of the last follow-up was reported in only nineteen studies. The full and modified Cincinnati scale\(^7\) was used in fifteen studies. Frequently, several scales were reported in the study, but only one scale was reported in the results section.

Two other correlations were significant. The Coleman Methodology Score correlated positively with the publication year \((r = 0.43, p < 0.01)\) (Fig. 3), but this correlation was no longer significant when the correlation was weighted by the number of patients in each study \((p = 0.197)\). The Coleman

\[
\begin{array}{|c|c|c|}
\hline
\text{No. of Studies} & \text{Mean Coleman Methodology Score (Range)} \\
\hline
\text{Autologous chondrocyte implantation} & 19 & 48.0 (29-67) \\
\text{Autologous osteochondral transplantation} & 24 & 39.3 (12-65) \\
\text{Microfracture} & 6 & 41.8 (26-55) \\
\text{Periosteal transplantation} & 8 & 37.0 (25-49) \\
\text{Retrospective} & 35 & 36.4 (12-55) \\
\text{Prospective} & 22 & 51.1 (38-67) \\
\text{Randomized controlled trial} & 4 & 63.5 (56-73) \\
\hline
\text{I} & 4 & 63.5 (56-73) \\
\text{II} & 21 & 51.6 (38-67) \\
\text{III} & 21 & 40.0 (29-52) \\
\text{IV} & 15 & 31.7 (12-55) \\
\hline
\end{array}
\]

*Randomized controlled trials are not included.
Methodology Score correlated positively (weighted analysis) with the level-of-evidence rating (r = 0.668, p < 0.0001), but the variations within each level were large (Fig. 4). The variation was progressively smaller with higher levels of evidence.

We unexpectedly identified several double publications. In addition, one article described a group of patients that may have been a subgroup of the patients included in a randomized trial. We also noticed that the pilot studies on microfracture referred to in one article had been published in a journal that was not indexed by the National Library of Medicine at that time and, therefore, were not readily available in PubMed or MEDLINE.

Discussion
Research on the surgical treatment of cartilage injury has been extensive over the last two decades. The results have varied from good to excellent in most studies with use of different surgical techniques. Although numerous articles have been published, the methodology of the studies in general has been questioned.

In the present study, we explored the hypothesis that studies on the surgical treatment of cartilage injuries have methodological limitations and that studies of lesser quality describe higher success rates. The low methodological quality of the studies in the present review may not necessarily apply to all orthopaedic studies in general, but in this particular area it should be a call for caution when recommending any treatment to patients. However, literature reviews on the surgical intervention in patellar and Achilles tendinopathy that used the same methodology score also found generally low methodological quality (a mean Coleman Methodology Score of 37 in both reviews). The five areas of methodological deficiency identified in those reviews were also found to be deficient in the articles evaluated in the present study.

We found no significant correlation between outcome and methodological quality. This is contrary to the findings in other reviews. The difference may be due to the heterogeneity of the studies and the large diversity of outcome measurement scales used, which, when combined, could conceal a possible correlation.

We found no significant difference among the outcomes of the four treatment modalities. This is in agreement with two of the published randomized controlled trials that did not find a significant difference between autologous chondrocyte implantation and autologous osteochondral transplantation and between autologous chondrocyte implantation and microfracture. Another randomized controlled trial did find a significant difference between autologous chondrocyte implantation and autologous osteochondral transplantation, but it scored lower on the Coleman Methodology Score than did the first two studies (56 compared with 65 and 73). Besides the differences in methodological quality, other reasons for the disagreement could be the use of unvalidated clinical scores; differences in inclusion criteria; the difference between active, scheduled follow-up and passive, usual clinical follow-up; and differences in randomization procedures. The findings of the present review agreed with those in another review published in 1996, in which no difference was found in the outcome among the studied treatment options. However, it is also important to be aware that this systematic review is not a meta-analysis of well-done randomized controlled trials. The summary results of the treatment efficacy are provided mainly to point out the large variation in all four treatment modalities, not to substantiate an argument as to which treatment is superior. Our finding that there was no difference among the treatments should be used as an incentive to perform more randomized controlled trials comparing the treatment modalities in use today to determine whether this is truly the case.

The effort by major orthopaedic journals to more clearly mark articles on the ground of methodological quality is highly appreciated. The finding of a strong correlation between the Coleman Methodology Score and the level-of-evidence rating with a progressively smaller variance with higher levels of evidence supports this method. However, not all randomized controlled trials are created equally (for example, in this review only one of the randomized controlled trials noted that a power analysis had been carried out), and they have limitations based upon the inherent design. The instructions in The Journal of Bone and Joint Surgery do not contain any rules on how to grade randomized controlled trials with regard to randomization procedure, the use of independent reviewers, or power analysis. A randomized controlled trial with more than 80% follow-up is rated as level-I evidence regardless of other serious design flaws. Therefore, the reader can be fairly confident that if a study receives a high level-of-evidence rating, the methodological quality of the study was good. On the other hand, the reader should still be aware that the level-of-evidence rating does not take into account all areas of sound study design. One suggestion to improve the rating would be to include a detailed methodology score in the submission process with the scoring of each subsection published online. This would not only enable readers to
evaluate the methodology more thoroughly, but it would also serve as a guideline for authors when designing and reporting a study, hopefully increasing the overall awareness regarding methodological quality.

The finding of double publications in this review is in agreement with recent findings in the orthopaedic literature\textsuperscript{19} and in the surgical literature in general\textsuperscript{19}. Duplicate publications are strongly discouraged by editorial boards and virtually no journal, regardless of specialty, allows publication of duplicate articles without explicit permission. If an article is double published, this should be clearly stated in the paper with a full reference to the first publication\textsuperscript{19}. There are several reasons for not allowing duplicate publications, including the extra cost of editing and peer-reviewing a second time and the potentially confounding effect on meta-analyses. In addition, many scientists also consider double publications to be unethical.

The limitations of this study are in the assumptions made by the Coleman Methodology Score and the high number of scales used by different authors. The Coleman Methodology Score actually assesses the quality of reporting, not the quality of the study, i.e., a high-quality study that is reported poorly would receive a low score. Unless the individual authors are contacted directly, this is an inherent weakness of all methodology scores as they do not necessarily reflect the true validity of the study but are biased by the quality of the reporting\textsuperscript{80}. The assumption in this review is that existing guidelines on how to report a clinical study have been followed in all articles\textsuperscript{18}, and the Coleman Methodology Score assessed from the article therefore reflects the quality of the study. To a certain extent, the quality of the study can also be said to reflect the validity of the treatment effect reported in the study, but, in the present review, the generally very low Coleman Methodology Score highlights the fact that the treatment effects in all of the papers must be interpreted with caution. The twenty-seven unvalidated scoring systems used by the authors make comparison difficult. Although the Lysholm scale was the most frequently used scoring system, it was not validated for use in patients with cartilage injury until 2004\textsuperscript{82}.

In conclusion, the generally low methodological quality of all of the studies included in this review shows that caution is required when interpreting results after surgical cartilage repair. Firm recommendations on which cartilage repair procedure to choose cannot be given at this time on the basis of these studies. Surgeons should pay more attention to established guidelines\textsuperscript{80} when designing, conducting, and reporting trials, to improve the methodological quality. Journals could include a detailed methodology score in their submission process to further encourage surgeons to focus on sound methodology. In cartilage repair, several new techniques are emerging and sound methodological quality in all types of studies is therefore important.

We propose the following guidelines for future studies on the basis of the findings in the present review:

1. Studies should be prospective with a clearly defined hypothesis and one clearly defined primary end point. They should be randomized controlled trials with an adequate randomization procedure and power analysis for the primary end point. Secondary end points should only be used as supportive evidence to the primary hypothesis.

2. Patient inclusion and exclusion criteria should be clearly established and reported. The recruitment rate should be reported, and attempts should be made to account for eligible patients who are not included and those who are lost to follow-up.

3. The outcome measure should be validated for use on patients with cartilage injuries.

4. Outcome assessment should be made by an independent investigator. The assessment should be in a written form and ideally be completed by the patient without investigator assistance.

5. The timing of the outcome assessment should be clearly stated. Results from various time-points after surgery should not be reported as one outcome. Assessments should be both clinical and functional. The minimum duration of follow-up should be more than twenty-four months.

6. Detailed rehabilitation protocols should be established and reported. Attempts should be made to monitor compliance. The protocols should be applied in a standardized manner to both patient cohorts.

Appendix

The online search strategy, details of the Coleman Methodology Score, and a list of all identified articles are available with the electronic versions of this article, on our website at jbjs.org (go to the article citation and click on “Supplementary Material”) and on our quarterly CD-ROM (call our subscription department, at 781-449-9780, to order the CD-ROM).
References


20. Micheli LJ, Browne JE, Erggeleto P, Fu F, Mandelbaum B, Moseley JB, Zura-


44. Hangody L, Kish G, Karpati Z, Szerb I, Udvarhegyi I. Arthroscopic autogenous osteochondral mosaicplasty for the treatment of femoral condylar articular de-


