Does ACL Reconstruction Alter Natural History?
A Systematic Literature Review of Long-Term Outcomes

Peter N. Chalmers, MD, Nathan A. Mall, MD, Mario Moric, MS, Seth L. Sherman, MD,
George P. Paletta, MD, Brian J. Cole, MD, MBA, and Bernard R. Bach Jr., MD

Investigation performed at the Departments of Orthopaedic Surgery and Anesthesiology, Rush University Medical Center, Chicago, Illinois, the St. Louis Center for Cartilage Restoration and Repair, Regeneration Orthopedics, St. Louis, and the Department of Orthopaedic Surgery, University of Missouri, Columbia, Missouri

Background: Anterior cruciate ligament (ACL) injury can lead to tibiofemoral instability, decreased functional outcomes, and degenerative joint disease. It is unknown whether ACL reconstruction alters this progression at long-term follow-up.

Methods: A systematic literature review of the long-term results (minimum follow-up, more than ten years) after operative intra-articular reconstruction of ACL injuries and after nonoperative management was performed to compare (1) knee stability on physical examination, (2) functional and patient-based outcomes, (3) the need for further surgical intervention, and (4) radiographic outcomes. After application of selection criteria, forty patient cohorts with a mean of 13.9 ± 3.1 years of postoperative follow-up were identified. Twenty-seven cohorts containing 1585 patients had undergone reconstruction, and thirteen containing 685 patients had been treated nonoperatively.

Results: Comparison of operative and nonoperative cohorts revealed no significant differences in age, sex, body mass index, or rate of initial meniscal injury (p > 0.05 for all). Operative cohorts had significantly less need for further surgery (12.4% compared with 24.9% for nonoperative, p = 0.0176), less need for subsequent meniscal surgery (13.9% compared with 29.4%, p = 0.0017), and less decline in the Tegner score (−1.9 compared with −3.1, p = 0.0215). A difference in pivot-shift test results was observed (25.5% pivot-positive compared with 46.6% for nonoperative) but did not reach significance (p = 0.09). No significant differences were seen in outcome scores (Lysholm, International Knee Documentation Committee [IKDC], or final Tegner scores) or the rate of radiographically evident degenerative joint disease (p > 0.05 for all).

Conclusions: At a mean of 13.9 ± 3.1 years after injury, the patients who underwent ACL reconstruction had fewer subsequent meniscal injuries, less need for further surgery, and significantly greater improvement in activity level as measured with the Tegner score. There were no significant differences in the Lysholm score, IKDC score, or development of radiographically evident osteoarthritis.

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

Peer Review: This article was reviewed by the Editor-in-Chief and one Deputy Editor, and it underwent blinded review by two or more outside experts. It was also reviewed by an expert in methodology and statistics. The Deputy Editor reviewed each revision of the article, and it underwent a final review by the Editor-in-Chief prior to publication. Final corrections and clarifications occurred during one or more exchanges between the author(s) and copyeditors.

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. Also, one or more of the authors has had another relationship, or has engaged in another activity, that could be perceived to influence or have the potential to influence what is written in this work. The complete Disclosures of Potential Conflicts of Interest submitted by authors are always provided with the online version of the article.

Anterior cruciate ligament (ACL) tears are among the most common injuries in young athletes. ACL injury has been associated with tibiofemoral instability, decreased functional outcomes, and meniscal injury. The damage that leads to these sequelae may occur at the time of the initial event because of associated meniscal injury, chondral injury, subchondral bone impaction, hemorrhage, and associated ligamentous injuries. Alternately, continued ligamentous deficiency may lead to loss of neuromuscular feedback, altered knee kinematics, increased shear and contact stress, an increased rate of meniscal injury, and an increased rate of chondral injury, each of which could accelerate degenerative changes.
Ligamentous reconstruction may protect against continued instability, improve functional outcomes, and decrease rates of degenerative joint disease.

The largest and highest-quality randomized clinical trial comparing operative and nonoperative treatment suggests that there are "no significant differences in self-reported outcomes at 2 years among the subjects treated with rehabilitation plus early ACL reconstruction, those treated with rehabilitation plus delayed ACL reconstruction, and those treated with rehabilitation alone." However, considerable deterioration can occur between short and long-term follow-up. The sequelae of continued instability are unlikely to be noted without extended clinical and radiographic monitoring. In addition, the decision not to perform an ACL reconstruction may be guided by a dominant treatment strategy based on identification of the patient as a "coper" with the ability to function normally or nearly normally with an inherently unstable knee. To our knowledge, no long-term randomized clinical trial evidence exists, although long-term outcomes have been reported for numerous non-comparative case series.

The purpose of the present study was to conduct a contemporary systematic review of all published clinical studies with a minimum follow-up of more than ten years after reconstruction and/or nonoperative treatment of ACL-deficient knees, evaluating (1) knee stability on physical examination, (2) functional and patient-based outcomes, (3) the need for further surgical intervention (meniscal surgery and all-cause reoperation), and (4) radiographic outcomes. We hypothesized that reconstruction would lead to greater knee stability on physical examination, improved functional outcomes with decreased rates of later meniscal injury, and a decrease in radiographically evident degenerative joint disease compared with nonoperative therapy.

**Materials and Methods**

**Search Strategy**

A systematic literature review was performed in March 2012 with use of the PubMed, Cochrane, BMJ Clinical Evidence, and Embase databases. Search terms used included anterior cruciate ligament, ACL, graft, reconstruct, autograft, patella hamstring, gracilis, semi-, tendon, and long-term. Exclusion criteria were a follow-up of less than the ten-year minimum, lack of either physical examination findings or clinical data at the time of final follow-up, a sample size of less than ten, surgery on patients with open physes, outcomes that were not segregated by operative and nonoperative treatment, and a language other than English. In addition, studies that employed ligamentous repair, ligamentous augmentation with a synthetic device, or extra-articular reconstruction or augmentation were excluded. The references of each included article and the table of contents of the last two years of *The Journal of Bone and Joint Surgery* (American and British Volumes); *The American Journal of Sports Medicine; Clinical Orthopaedics and Related Research; Arthroscopy; and Knee Surgery, Sports Traumatology, Arthroscopy* were searched manually for any additional studies. Additionally, the librarian at our institution performed an independent search to ensure that no manuscripts were missed. Study authorship and data were cross-checked to prevent patient data duplication, with longer-term data preferentially included and shorter-term data excluded in such a situation. Exceptions were made for studies reporting on different outcomes in the same patient population. We adhered to the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines. Studies with Level-I through Level-IV evidence were included.

**Data Collection**

General demographic data were collected for each study (see Appendix). The following information at the time of final follow-up was collected: need for knee surgery for any cause between the initial treatment and final follow-up, need for further meniscal surgery, number of patients with a side-to-side difference of >5° in range of knee motion, number of patients with an abnormal Lachman test, number of patients with an abnormal pivot-shift test, number of patients with a difference of ≥3 mm (the most commonly used threshold in the literature) on KT-1000 arthrometer testing (MEDmetric, San Diego, California), mean side-to-side difference on maximum manual testing, mean Lysholm score, mean pre-injury and final follow-up Tegner activity scores as well as the change between these scores, percentage of patients returning to their pre-injury level of athletic activity, mean Cincinnati Knee Score, mean Knee injury and Osteoarthritis Outcome Score (KOOS), mean Short Form (SF)-36 score, mean International Knee Documentation Committee (IKDC) score, and mean overall IKDC grade. The Lysholm, Tegner, and IKDC scores have been validated in this patient population. Study quality was graded with use of the modified Coleman Methodology Score (CMS).

Osteoarthritis of the knee at the time of final follow-up was assessed with use of the IKDC system, the Kellgren and Lawrence system, the Ahlbäck system, the Fairbank system, and the Osteoarthritis Radiographic Severity Index (OARSI). Radiographic outcomes were then divided (a priori) into those with osteoarthritis as defined previously by Øiestad et al. (any patient with IKDC grade C or greater) and those without; similar divisions were...
made on the basis of a Kellgren and Lawrence grade of ≥2, Ahlbäck grade of ≥1, and Fairbank grade of ≥2.

**Statistical Analysis**

To avoid excessive bias from a single cohort, analysis was attempted only for those outcomes reported in at least three operative and three nonoperative studies. Studies with both nonoperative and operative arms or with multiple graft type subgroups were included in the systematic review as separate cohorts to reduce the effects of correlated data. If the variance or standard deviation was not given, it was calculated from the standard error, 95% confidence interval (CI), range, or p value as available. Dichotomous data such as odds ratios and proportions were re-expressed in the form of the standardized mean difference.

Analyses were performed to pool primary outcome measures among cohorts in the same treatment group. Because of the variability inherent in the use of observational data from a diverse set of studies, the conservative strategy of using random-effects modeling was used for each outcome to estimate and adjust for heterogeneity. This involved use of a mixed-effects meta-regression model (one involving a combination of both random and fixed effects).

For each pooled outcome measure, the outcome in each cohort was weighted according to the inverse of the sample variance of the outcome in the cohort. After weighting and adjusting for sample size, variability, and sampling error, the overall size of the effect of operative intervention on each outcome variable was estimated with use of a mixed-effects meta-regression model. The treatment effect was estimated with use of a maximum-likelihood estimation of the mean difference between the operative and nonoperative groups yielded an estimate of the treatment effect. The treatment effect was standardized according to the pooled variance estimate to calculate the estimated effect size and 95% CI.

If necessary, the sign of the estimated effect size was transposed so that a positive value would indicate a better clinical outcome for the operative group compared with the nonoperative group. All analyses were performed with use of SAS software (version 9.2; SAS Institute, Cary, North Carolina); the mixed-effects meta-regression model analyses were performed with use of the MIXED procedure.

**Source of Funding**

This study did not receive any external funding.

**Results**

The initial search revealed 952 potentially relevant abstracts, and twenty-nine remained after application of our study selection algorithm. Thirteen of these reported outcomes of nonoperative therapy and twenty-four (twenty-seven cohorts) reported outcomes of operative therapy (Fig. 1). Two of the studies were randomized clinical trials, two were matched cohort series, seven were prospective cohort series, and eighteen were retrospective cohort series. The twenty-seven operative cohorts had a mean sample size of 58.7 (range, twenty-two to 181), with a total of 1585 included subjects and a mean follow-up of 12.9 years (range, ten to sixteen years) (see Appendix). The ACL reconstruction utilized the patellar tendon in twenty-three (85%) of the operative cohorts, the hamstring tendon in three (11%) and the tibialis anterior tendon in one (4%). Although most studies used autograft, allograft was used in some cohorts. Reconstruction took place at a mean

Fig. 2

Summary forest plot showing the mean effect size (with 95% CI) for the comparisons between the operative and nonoperative groups with regard to the Lysholm score, percentage of patients with a positive pivot-shift test, IKDC score, percentage of patients who subsequently developed radiographically evident degenerative joint disease (DJD), percentage of patients requiring further surgery, percentage of patients requiring subsequent meniscal surgery, percentage of patients with preoperative meniscal injury, preoperative Tegner score, Tegner score at the final postoperative follow-up, change in Tegner score between the preoperative and final follow-up time points, age, sex, and BMI.
(and standard deviation) of 20.8 ± 11.3 months after injury. Although most studies used an arthroscopic or mini-open technique, several either involved reconstruction through a medial parapatellar arthrotomy or did not indicate the method of reconstruction. The thirteen nonoperative cohorts had a mean sample size of 52.7 (range, eighteen to ninety-four), with a total of 685 included subjects and a mean follow-up of 16.2 years (range, ten to twenty-seven years). Reporting of each of the outcomes of interest varied among the studies (see Appendix). Preoperatively, the two treatment groups were equivalent with respect to age (p = 0.7958), sex (p = 0.5837), body mass index (BMI) (p = 0.8732), Tegner score (p = 0.4173), and the percentage of patients with concomitant meniscal damage (p = 0.3547) (Fig. 2).

Assessment with the pivot-shift test revealed a trend toward a difference between the treatment groups (p = 0.0941), with 46.6% of nonoperatively treated knees and 25.5% of operatively treated knees being unstable to the pivot (Fig. 2). Because fewer than three nonoperative cohorts had reported results for Lachman testing or KT-1000 testing, no analysis could be performed for these outcomes.

Neither the Lysholm score (mean, 84.4 for nonoperative compared with 88.7 for operative, p = 0.0837) nor the IKDC score (mean, 79.2 for nonoperative compared with 84.5 for operative, p = 0.2944) were significantly higher in the operative cohorts (Fig. 2). The operative and nonoperative cohorts showed no significant difference in the Tegner score at the time of final follow-up (4.3 compared with 4.8, p = 0.1351), although the operative cohorts did show significantly less decline in the Tegner score relative to the preoperative level (mean, 2.19 compared with 2.31 for nonoperative, p = 0.0215) (Figs. 2 and 3). Because fewer than three nonoperative cohorts had reported results for the Cincinnati Knee Score, KOOS, or SF-36 score, no analysis could be performed for these tests.

Further knee surgery was required approximately twice as frequently in the nonoperative cohorts than in the operative cohorts (24.9% compared with 12.4%) (Fig. 4), with the difference in reoperation rate being significant (p = 0.0176). The nonoperative cohorts also required subsequent meniscal surgery more than twice as frequently (29.4% compared with 13.9% in the operative cohorts) (Fig. 5), and this difference was also significant (p = 0.0017).

Radiographically evident degenerative joint disease was observed in a similar number of patients in the operative and nonoperative cohorts (35.3% compared with 32.8%, p = 0.7678) (Fig. 2).
Fig. 4
Forest plot showing the mean percentage of patients requiring further surgery in each operative (blue) and nonoperative (green) cohort for which this outcome was reported as well as the two grand means. See the Figure 3 legend for further details.

Fig. 5
Forest plot showing the percentage of patients requiring subsequent meniscal surgery in each operative (blue) and nonoperative (green) cohort for which this outcome was reported as well as the two grand means. See the Figure 3 legend for further details.
Discussion

ACL deficiency can lead to symptomatic instability, further intra-articular damage, and accelerated degenerative joint disease\(^{21}\). ACL reconstruction is often recommended to protect against continued instability\(^{22,23}\), to improve knee function, to reduce the likelihood of meniscal tears, and possibly to decrease the rate of degenerative joint disease\(^{25}\). However, it is currently unknown whether reconstruction actually confers these benefits. The purpose of this study was to conduct a systematic review of published clinical studies with more than a minimum of ten years of follow-up of reconstructed and non-operatively treated ACL-deficient knees, evaluating (1) knee stability on physical examination, (2) functional and patient-based outcomes, (3) the need for further surgical intervention (meniscal surgery and all-cause reoperation), and (4) radiographic outcomes.

This study revealed that, compared with nonoperatively treated patients, patients who had undergone ACL reconstruction using modern techniques had fewer subsequent meniscal tears, less need for further surgery, and a greater change in activity from the preoperative level as measured with the Tegner score. No significant differences were observed in the Lysholm score, IKDC score, or radiographic progression of arthritis. Given that the minimum detectable change in the Lysholm score for this population is 8.9\(^{44}\) and that the minimum clinically important difference for the IKDC score in a similar population is 16.7\(^{64}\), the observed differences in the Lysholm and IKDC scores were also not clinically relevant.

The natural history association between ACL injury, subsequent meniscal injury\(^{65,66}\), subsequent knee surgery, and eventual degenerative joint disease has been described with longitudinal studies\(^{15-17}\). Our review suggests that patients who undergo ACL reconstruction are less likely to undergo subsequent meniscal surgery compared with patients with ACL deficiency treated nonoperatively. The mean time between ACL injury and ACL reconstruction in this review was 20.8 months, during which time substantial meniscal or articular damage could occur, possibly exacerbating differences in meniscal injury between operative and nonoperative management. This review confirmed that ACL reconstruction protects the knee from the need for further operative interventions. However, the connection between meniscal injury occurring as a result of continued ligamentous instability and the subsequent development of radiographically evident degenerative joint disease remains unclear\(^{5,6,67}\). The lack of any difference in the rate of degenerative joint disease between reconstructed and non-reconstructed knees in our analysis suggests that the continued ligamentous instability and meniscal injury may not be the sole causes of joint degeneration. Further research regarding the role of meniscal injury in the development of degenerative joint disease is necessary\(^{4,11}\). Changes in patient activity level may also influence the subsequent development of degenerative joint disease, and a greater increase in activity was noted in operatively treated patients. The development of radiographic and clinical signs of osteoarthritis is likely due to a combination of the effects of the initial trauma\(^{8-12}\), meniscal pathology, subtle rotational instability, patient factors, and biochemical and genetic factors that are incompletely understood\(^{86}\). Our study was limited to radiographic evidence of degenerative joint disease, which serves as an imperfect marker of symptomatic degenerative joint disease.

The authors of previous series have also noted that the subjective stability conferred by ACL reconstruction allows patients to resume their activities\(^{16,37,61,64}\), whereas those patients with unreconstructed knees and continued instability often self-limit their activities\(^{63,59}\). Our data confirmed this difference, with the reconstructed cohorts experiencing a significantly smaller decrease in the Tegner score compared with the nonreconstructed cohorts. Patients with high expectations with regard to athletic activity should be counseled regarding the change in activity level associated with nonoperative management. Although no significant differences were observed in the Lysholm or IKDC scores, these outcome instruments may be insufficiently sensitive to demonstrate the subtle differences in the quality and stability of an individual knee that prevent a return to higher activity levels\(^{1,69}\).

Previous studies have noted resolution of the pivot-shift phenomenon with reconstruction\(^{27,70,71}\). The present review revealed a 21% difference in the percentage of patients with a positive pivot-shift test in the operative and nonoperative cohorts (26% compared with 47%), but this difference did not reach significance (\(p = 0.0941\)). The percentage of pivot-positive knees in the operative cohorts was very high compared with the 5% to 9% rates in recent series involving modern ACL reconstruction techniques. However, the recent techniques do not yet have ten-year follow-up data to allow a true comparison, and grafts may stretch between short and long-term follow-up. Alternatively, the high rate of persistence of the pivot phenomenon in the operative cohorts in the present review may reflect nonanatomic tunnel placement in older series\(^{27,70,71}\). Considerable change has occurred in our understanding of the effect of femoral tunnel positioning on rotational tibiofemoral stability during the thirty-year period encompassed by the studies in the present review\(^{72-74}\). Several trials have demonstrated significantly better rotational stability with anatomic reconstruction\(^{77-80}\); the anatomic femoral tunnel placement may be achievable with transtibial, anteromedial, and two-incision techniques\(^{81-83}\). The present review may thus underestimate the differences between reconstructed and non-reconstructed cohorts. In addition, the pivot maneuver is difficult to perform correctly; subtle variations can alter the results, and thus heterogeneity could have obscured relevant differences among cohorts\(^{84,85}\).

Several previous systematic reviews have been conducted to compare operative and nonoperative treatment of ACL injury\(^{86-89}\). The analysis reported by Linko et al. in 2005 was limited to randomized trials, and the authors did not find sufficient evidence to recommend either reconstruction or rehabilitation alone\(^{86}\). In 2009, Øiestad et al. reported the only previous systematic review of outcome studies that was limited to long-term outcomes (at more than ten years)\(^{89}\). The authors focused their analysis on radiographically evident degenerative joint disease...
and found no difference between the operative and nonoperative cohorts. Of note, over one-half of the studies included in the present review were released since the literature search by Øiestad et al. was performed. Several randomized clinical trials have been conducted to compare operative and nonoperative treatment in the short term; the largest and highest-quality of these suggests that there are “no significant differences in self-reported outcomes at 2 years among the subjects treated with rehabilitation plus early ACL reconstruction, those treated with rehabilitation plus delayed ACL reconstruction, and those treated with rehabilitation alone.” That same study described thirty-two episodes of clinical instability and meniscal signs and symptoms in the delayed reconstruction group compared with three in the early reconstruction group, confirming our findings.

The present study has a number of limitations. First, by design, the analysis compared cohort studies performed by different authors. Heterogeneity among these studies limits the interpretation of the results. Considerable changes have occurred between the publication of the first included trial (1993) and the last (2012) with respect to rehabilitation protocols, patient expectations, and reconstruction techniques. To mitigate this effect, only intra-articular reconstruction techniques were included, with the oldest study of reconstructive outcomes published in 2001. Second, as with any systematic review, the quality of the original data limits the quality of our findings. Third, the radiographic osteoarthritis grade was determined by a single unblinded orthopaedist or radiologist in most of the included studies, limiting the validity of this information. Fourth, limitation of the included data to published studies may have also introduced publication bias. Fifth, the majority of the studies included in this review were published outside the United States (reflecting a bias in the underlying literature), and thus our results may not be generalizable. However, as these five limitations affect both operative and nonoperative data equally, our results are unlikely to have been affected. Sixth, the extended length of time between injury and reconstruction (mean, twenty-one months) may have reduced the potential benefits of reconstruction. Finally, the majority of reconstructions in this review used patellar tendon autograft. Although several studies have noted equivalence between patellar tendon and hamstring tendon autografts, it remains unclear whether the results in the present review would also apply to hamstring grafts.

Although no baseline differences between the operative and nonoperative cohorts existed with respect to the percentage of female patients, preoperative Tegner score, or percentage of patients with concomitant meniscal injury, unmeasured residual bias likely exists between the treatment groups. Only randomization would be able to overcome this limitation; however, most physicians would agree that ACL reconstruction is the treatment of choice in a young, active patient who wishes to return to cutting and pivoting activities.

In conclusion, at a mean of 13.9 ± 3.1 years after injury, the patients who underwent ACL reconstruction had fewer subsequent meniscal injuries, less need for further surgery, and a significantly greater improvement in activity level as measured with the Tegner score. Clinical knee stability as measured with the pivot-shift test was also greater, but this difference did not reach significance. There were no significant differences in Lysholm and IKDC scores. Patients who had undergone operative ACL reconstruction did not have a lower rate of radiographically evident osteoarthritis compared with patients who had undergone nonoperative treatment.

Appendix

Tables summarizing the included studies and the number of studies with reporting of each outcome are available with the online version of this article as a data supplement at jbjs.org.

References


