Comparing Responsiveness Of Six Common Patient-Reported Outcomes To Changes Following Autologous Chondrocyte Implantation: A Systematic Review And Meta-Analysis Of Prospective Studies

By: Jennifer S. Howard, Christian Lattermann, Johanna M. Hoch, Carl G. Mattacola, and Jennifer M. Medina McKeon

Abstract

Objective: To compare the responsiveness of six common patient-reported outcomes (PROs) following autologous chondrocyte implantation (ACI). Design: A systematic search was conducted to identify reports of PROs following ACI. Study quality was evaluated using the modified Coleman Methodology Score (mCMS). For each outcome score, pre- to postoperative paired Hedge’s g effect sizes were calculated with 95% confidence intervals (CIs). Random effects meta-analyses were performed to provide a summary response for each PRO at time points (TP) I (1 year), II (1 year to 2 years), III (2 years to 4 years), IV (4 years), and overall. Results: The mean mCMS for the 42 articles included was 50.9 ± 9.2. For all evaluated instruments, none of the mean effect size CIs encompassed zero. The International Knee Documentation Committee Subjective Knee Form (IKDC) had increasing responsiveness over time with TP-IV, demonstrating greater mean effect size [confidence interval] (1.78 [1.33, 2.24]) than TP-I (0.88 [0.69, 1.07]). The Knee Injury and Osteoarthritis Outcome Score–Sports and recreation subscale (KOOS-Sports) was more responsive at TP-III (1.76 [0.87, 2.64]) and TP-IV (0.98 [0.81, 1.15]) than TP-I (0.61 [0.44, 0.78]). Overall, the Medical Outcomes Study 36-Item Short Form Health Survey Physical Component Scale (0.60 [0.46, 0.74]) was least responsive. Both the Lysholm Scale (1.42 [1.14, 1.72]) and the IKDC (1.37 [1.13, 1.62]) appear more responsive than the KOOS-Sports (0.90 [0.73, 1.07]). All other KOOS subscales had overall effect sizes ranging from 0.90 (0.74, 1.22) (Symptoms) to 1.15 (0.76, 1.54) (Quality of Life). Conclusions: All instruments were responsive to improvements in function following ACI. The Lysholm and IKDC were the most responsive instruments across time. IKDC and KOOS-Sports may be more responsive to long-term outcomes, especially among active individuals.

Comparing Responsiveness of Six Common Patient-Reported Outcomes to Changes Following Autologous Chondrocyte Implantation: A Systematic Review and Meta-Analysis of Prospective Studies

Jennifer S. Howard¹,², Christian Lattermann², Johanna M. Hoch³, Carl G. Mattacola¹, and Jennifer M. Medina McKeon¹

Abstract

Objective: To compare the responsiveness of six common patient-reported outcomes (PROs) following autologous chondrocyte implantation (ACI). Design: A systematic search was conducted to identify reports of PROs following ACI. Study quality was evaluated using the modified Coleman Methodology Score (mCMS). For each outcome score, pre- to postoperative paired Hedge’s g effect sizes were calculated with 95% confidence intervals (CIs). Random effects meta-analyses were performed to provide a summary response for each PRO at time points (TP) I (<1 year), II (1 year to <2 years), III (2 years to <4 years), IV (≥4 years), and overall. Results: The mean mCMS for the 42 articles included was 50.9 ± 9.2. For all evaluated instruments, none of the mean effect size CIs encompassed zero. The International Knee Documentation Committee Subjective Knee Form (IKDC) had increasing responsiveness over time with TP-IV, demonstrating greater mean effect size [confidence interval] (1.78 [1.33, 2.24]) than TP-I (0.88 [0.69, 1.07]). The Knee Injury and Osteoarthritis Outcome Score–Sports and recreation subscale (KOOS-Sports) was more responsive at TP-III (1.76 [0.87, 2.64]) and TP-IV (0.98 [0.81, 1.15]) than TP-I (0.61 [0.44, 0.78]). Overall, the Medical Outcomes Study 36-Item Short Form Health Survey Physical Component Scale (0.60 [0.46, 0.74]) was least responsive. Both the Lysholm Scale (1.42 [1.14, 1.72]) and the IKDC (1.37 [1.13, 1.62]) appear more responsive than the KOOS-Sports (0.90 [0.73, 1.07]). All other KOOS subscales had overall effect sizes ranging from 0.90 (0.74, 1.22) (Symptoms) to 1.15 (0.76, 1.54) (Quality of Life). Conclusions: All instruments were responsive to improvements in function following ACI. The Lysholm and IKDC were the most responsive instruments across time. IKDC and KOOS-Sports may be more responsive to long-term outcomes, especially among active individuals.

Keywords

articular cartilage, cartilage, knee, outcomes assessment, self-report

Introduction

The limited ability of articular cartilage to heal on its own has been a topic of discussion for more than 200 years.¹ The treatment and management of articular cartilage damage can be particularly challenging in the knee joint where such defects are frequently observed during arthroscopy.²⁻⁵ Restorative and reparative treatment of these defects is highly desirable to prevent the progression of osteoarthritis.⁵ Over the last three decades, approaches to treating chondral defects have shifted toward cell-based therapies. One of the most frequently used and well studied is the autologous chondrocyte implantation (ACI).⁷
**Treatment Evaluation**

As new methods for treating cartilage are developed, it is necessary to evaluate these treatments to determine their effectiveness. Although second look arthroscopies with cartilage biopsies may provide the most diagnostic method of evaluating cartilage repair, they are not always feasible or ethical to perform. Furthermore, biopsies allow for the assessment of the histological tissue repair but cannot be used to evaluate patient-oriented outcomes such as pain and function. To evaluate patient-oriented outcomes, investigators have relied on patient-reported outcome (PRO) instruments. Many PROs have been developed to address outcomes associated with a specific body part or region, a specific disease, or health-related quality of life as a whole. Numerous PROs have been used to document patient response to cartilage repair. Although the widespread use of PROs is beneficial for documenting treatment outcomes, the wide variety in PROs makes comparison across studies and instruments difficult. Ideally, a standard instrument or battery of instruments would be advantageous for reliable and valid assessment of patient response to treatment.

Some of the most commonly used PROs to evaluate articular cartilage repair outcomes include the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36), the International Knee Documentation Committee Subjective Knee Form (IKDC), the Lysholm Knee Scale (Lysholm), the Modified Cincinnati Knee Rating System (MCKRS), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), and the Knee Injury and Osteoarthritis Outcome Score (KOOS). Test–retest reliability and validity among cartilage patients has previously been established for a version of each of these instruments. Although all these instruments have been widely used to evaluate ACI treatment efficacy, there is no clear standard regarding which outcome instrument is ideal for evaluating treatment effects following ACI.

PRO responsiveness is the evaluation of change in the instrument score over time in response to treatment. The reported responsiveness in self-reported function following ACI has not been compared among instruments. Identification of the most responsive instrument for an ACI population will provide clinicians and researchers with a specific PRO instrument to compare treatment effects between therapies.

The purpose of this study was to systematically review and summarize the scientific literature evaluating changes in PRO scores after ACI. For analysis, we selected commonly used outcome instruments in cartilage repair studies, including the IKDC, Lysholm, MCKRS, KOOS, WOMAC, and SF-36. The outcome of interest was PRO responsiveness following ACI treatment. Meta-analyses of PRO score changes (“Hedge’s g effect sizes with 95% CIs) were plotted among instruments to visually reflect the responsiveness of each instrument at specified postoperative time points (TPs) forest plots of mean effect sizes for each instrument at each TP were used to provide a graphical representation of how responsive each instruments is to changes in self-reported knee function at varying postoperative TPs. A better understanding of the responsiveness of each instrument will allow for improved selection of outcome instruments in future cartilage research.

**Methods**

This nonregistered review was prepared in accordance with the “The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions.”

**Evidence Acquisition**

**Search strategy.** In February 2011, investigators conducted a systematic search of the literature using CINAHL (from 1981), Medline (from 1966), and SPORTDiscus (from 1980) to identify reports of PROs following autologous chondrocyte implantation/transplantation. Search terms used were autologous, chondrocyte, outcome, and knee. All abstracts were then reviewed for study inclusion/exclusion. In the event the abstract did not provide sufficient information to determine study eligibility, the full manuscript was reviewed. Additionally, the reference lists of all included studies were reviewed to identify other potentially eligible studies.

**Selection criteria.** All studies were required to meet the following inclusion criteria: (a) publication in English; (b) investigations with human participants; (c) prospective evaluation of patient outcomes following cell-based treatment of articular cartilage defects with some form of cultured autologous chondrocytes; (d) utilization of at least one of the following PRO instruments: IKDC, Lysholm, MCKRS as described by Browne et al., KOOS, WOMAC, or SF-36 Physical Component Scale (SF-36 PCS) preoperatively and at a minimum of one postoperative TP; and (e) reporting of statistics from which effect sizes and 95% confidence intervals (CIs) could be calculated.

**Assessment of Methodological Quality and Level of Evidence**

The quality of all studies included was assessed using the Coleman Methodology Score modified by Kon, Verdonk, and others. This assessment tool was specifically adapted to evaluate the quality of cartilage repair studies and includes 11 parameters on a 100-point scale (100 = highest quality): study sample size (10 points), average follow-up (10), number of concomitant surgical procedures (10), study design (15), description of the surgical procedure (5), description of postoperative rehabilitation (5), the
inclusion of MRI outcome (10), the inclusion of histological outcome (10), outcome criteria (5), procedure for assessing clinical outcomes (7), and description of subject selection process (8).24

Level of evidence was evaluated based on criteria from the Centre for Evidence Based Medicine (CEBM).25 Using this taxonomy, the quality of the evidence for the studies included was determined and a grade of recommendation was generated for the use of each PRO as a measure of ACI treatment effect.25 Methodological quality assessment and the rating of the level of evidence were assessed independently by two investigators. Discrepancies in scoring were discussed until a consensus score was agreed upon.

Data Extraction

The primary outcome variables of interest were scores on six specified PROs: the IKDC, Lysholm, MCKRS, KOOS, WOMAC, and SF-36 PCS. Because of the variation in Modified Cincinnati Knee Rating Systems reported in the literature, only the MCKRS presented by Browne et al.14 was reviewed. To avoid inappropriate comparisons of various versions of the MCKRS, the studies included had to have either published the scale in the article or provided a clear reference. From each study all data that could be used to calculate effect sizes for PROs were extracted.

For each outcome score, individual pre to postoperative standardized effect sizes were calculated using bias-corrected Hedge’s $g$ for paired samples26 with 95% confidence intervals (CIs) to examine the magnitude and precision of the difference between pre- and postoperative PRO scores. These effect sizes are unitless measures, corrected to represent a parametric distribution of the effects. For the purpose of this study, Hedge’s $g$ effect sizes were used as a measure of responsiveness with larger effect sizes representing increased responsiveness of an instrument—that is, greater change in the instrument score over time.18 Separate meta-analyses were performed to provide a summary response for each PRO at specified TPs. For the purposes of analysis, follow-up TPs were grouped into four categories, TP-I (<1 year), TP-II (1 year to <2 years), TP-III (2 years to <4 years), and TP-IV (4 years or more). Most studies made multiple comparisons across separate TPs, and each comparison was treated independently during statistical analyses. If a study reported multiple results within a given TP category (e.g., 3 months and 6 months are both within the <1 year TP category), only the latest data point (i.e., 6 months) was analyzed. Therefore, within each study, only a single result per instrument was included for a given TP category. For each PRO, an additional meta-analysis to determine the overall responsiveness across all TPs was conducted using the pooled standardized effects averaged across all available TPs. For each meta-analysis, a random effects model was employed.26

Effect sizes, 95% CIs, and $Z$-distribution $P$ values were calculated in Comprehensive Meta Analysis (Version 2.0, Biostat, Englewood, NJ). A positive effect size indicated improvement in postoperative PRO score compared with preoperative score. Effect sizes for which CIs did not overlap were considered to be substantially different. Effect size values were interpreted as small if they were between 0.20 and 0.49, moderate if between 0.50 and 0.79, and large if more than 0.80.

Assessment of Bias

Methodological bias was assessed using part B of the modified Coleman Methodology Score.24 To assess the likelihood of publication bias, a funnel plot of all measures was generated by plotting standard error against Hedge’s $g$ effect size for each included study. To assess the robustness of the observed overall effects of the variations in study design on PRO score, Orwin’s Fail-Safe $N$ test was employed.27 For this test, a Hedge’s $g$ effect size of 0.1 was assumed for all missing studies, or studies excluded due to publication bias, and the number of missing studies necessary to reduce the overall mean effect size for each instrument to a 0.4 was calculated.

Results

Study Selection

The initial literature search yielded 216 results. Application of inclusion and exclusion criteria resulted in the inclusion of 42 articles.14,28-68 Study selection and inclusion is depicted in fig 1. The included studies are summarized in Table 1. A total of 2,016 patients with a mean age of approximately 34.5 years are reported on across all studies. Overall, 16 studies reported outcomes using the IKDC, 11 used the KOOS (2 reporting only total KOOS scores), 18 used the Lysholm, 12 used the MCKRS, 9 studies used the SF-36 PCS, and 2 studies used the WOMAC. A single study reported on four instruments.58 All other studies reported on three or fewer instruments.

Methodology Scoring and Level of Evidence

The mean modified Coleman Methodology Score for all included articles was 50.9 ± 9.2 (range = 35-68). The mean modified Coleman Methodology Score for studies using each PRO instrument was as follows: IKDC 51.4 (standard deviation [SD] = 7.5), KOOS-Sports 51.9 (5.8), KOOS-all other subscales 53.7 (7.7), Lysholm 49.2 (8.8), MCKRS 48.2 (8.8), and SF-36 PCS 56.2 (7.8). The least reported parameters were inclusion of MRI and histological outcomes and description of the subject selection process. CEBM level of evidence was 2b for 38 articles.
and 1b for 4 articles included. Based on the consistent reporting of level 2 studies, a grade B recommendation was made for the use of the IKDC, KOOS, Lysholm, MCKRS, SF-36 PCS, and WOMAC as outcome measures following ACI. 

Assessment of Bias
The mean score for Part B of the modified Coleman Methodology Score assessing individual study bias was 13 ± 3 out of a possible 22 points. The assessment of publication bias revealed an asymmetrical distribution of studies with a disproportionate number of studies above the mean effect size at the bottom of the funnel (Fig. 2). This indicated a slight publication bias toward studies demonstrating large treatment effects, particularly for studies with smaller sample sizes. However, the results of Orwin’s Fail Safe N test (Table 2) demonstrate that an additional 14 (SF-36 PCS) to 196 (KOOS) studies with a trivial effect size of 0.10 are necessary to reduce the mean effect size for any of the PROs to a weak value of 0.40. Therefore, the observed overall effects are very robust and not likely to be artificially influenced by this potential publication bias.

Responsiveness of PROs
Mean effect sizes and 95% CIs for each instrument at each of the four TPs are reported in the forest plots in Fig 3. For an instrument to be included in the meta-analysis at a given TP, a minimum of four individual data points must have been reported. The WOMAC did not meet this requirement at any time point, and the SF-36 PCS only met this requirement at TP-III. The MCKRS could only be evaluated at TP-III and TP-IV and only the patient perception scale could be evaluated.

Responsiveness by Instruments across TPs
For all evaluated instruments, none of the mean effect sizes or CIs encompassed zero. This indicated that there was evidence of positive treatment effects following ACI regardless of the PRO used (Fig. 3). The IKDC increased responsiveness over time, with TP-IV demonstrating a greater mean effect size (mean effect size [95% CI]: 1.78 [1.33, 2.24]) than TP-I (0.88 [0.69, 1.07]). The responsiveness of the Lysholm varied little across TPs, with mean effect sizes only ranging from 1.29 to 1.69. There was also no difference in responsiveness for the MCKRS between TP-II and TP-III. Finally, the only KOOS subscale to show noticeable improvements in responsiveness over time was the KOOS-sports and recreation subscale (KOOS-Sports) for which TP-III (1.76 [0.87, 2.64]) and TP-IV (0.98 [0.81, 1.15]) demonstrated larger effect sizes than TP-I (0.61 [0.44, 0.78]). Effect sizes for the remaining KOOS subscales did not change over time and fell in the following ranges: KOOS-Activities of Daily Living 0.78 [0.27, 1.29] to 1.90 [1.02, 2.78], KOOS-Pain subscale 0.75 [0.40, 1.10] to 1.88 [1.12, 2.63], KOOS-Quality of Life 0.88 [0.32, 1.44] to 2.38 [1.20, 3.56], and KOOS-Symptoms 0.75 [0.50, 1.00] to 1.60 [0.79, 2.41].

Responsiveness by TP
At TP-I, the Lysholm (1.52 [0.92, 2.11]) appears more responsive than the KOOS-Sports subscale (0.61 [0.44, 0.78]). At TP-II, both the IKDC (1.37 [0.93, 1.80]) and the Lysholm (1.53 [0.96, 2.11]) were more responsive than the KOOS-Sports subscale (0.57 [0.23, 0.92]). There were no identifiable differences between any of the instruments at TP-III. Finally, at TP-IV the IKDC (1.78, [1.33, 2.24]) demonstrated a larger effect size than the KOOS-Sports subscale (0.98 [0.81, 1.15]).

Overall Responsiveness
The final comparison was of the overall responsiveness of each instrument with data from all available TPs averaged (Fig. 4). Both the Lysholm (1.43 [1.14, 1.72]) and the IKDC (1.37 [1.13, 1.62]) had overall mean effect sizes that
Table 1. Descriptive Variables for Autologous Chondrocyte Implantation Studies Included in Systematic Review and Meta-Analyses

<table>
<thead>
<tr>
<th>Study</th>
<th>Mean Age</th>
<th>Procedure Included</th>
<th>Follow-up Time (years)</th>
<th>Instrument Included in Review</th>
<th>Total N Analyzed</th>
<th>Lesion Locations</th>
<th>Average Lesion Size (cm²)</th>
<th>Level of Evidence</th>
<th>Mod. Coleman Methodology Score (Part B)</th>
<th>Potential Conflict of Interest Disclosed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basad et al. 2010</td>
<td>33</td>
<td>MACI</td>
<td>0.5, 1, 5, 2</td>
<td>L</td>
<td>39</td>
<td>FC, Troc, Pat</td>
<td>2.46</td>
<td>2b</td>
<td>45 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Behrens et al. 2006</td>
<td>35</td>
<td>MACI</td>
<td>2.87, 5</td>
<td>L</td>
<td>33</td>
<td>MFC, LFC, Pat</td>
<td>4.9</td>
<td>2b</td>
<td>52 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Bhosale et al. 2007</td>
<td>43</td>
<td>ACI-C w/Meniscus allograft transplant</td>
<td>1</td>
<td>L</td>
<td>8</td>
<td>MFC, LFC, Kissing</td>
<td>9.7 femoral, 3.7 tibial (median)</td>
<td>2b</td>
<td>47 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Briggs et al. 2003</td>
<td>30</td>
<td>ACI-C</td>
<td>2.825</td>
<td>L</td>
<td>14</td>
<td>MFC, LFC, Troc, Pat</td>
<td>6.27</td>
<td>2b</td>
<td>44 (14)</td>
<td>No</td>
</tr>
<tr>
<td>Browne et al. 2005</td>
<td>37</td>
<td>ACI-P</td>
<td>5</td>
<td>M, MP, MS</td>
<td>87</td>
<td>MFC, LFC, Troc</td>
<td>4.9</td>
<td>2b</td>
<td>61 (17)</td>
<td>Yes</td>
</tr>
<tr>
<td>de Windt et al. 2009</td>
<td>35</td>
<td>ACI—multiple versions</td>
<td>3</td>
<td>K—Total only</td>
<td>25</td>
<td>MFC, LFC</td>
<td>2.25</td>
<td>2b</td>
<td>49 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Della Villa et al. 2010</td>
<td>24.3</td>
<td>MACI in athletic compared to nonathletic males</td>
<td>1, 5</td>
<td>I</td>
<td>65</td>
<td>MFC, LFC, Troc</td>
<td>6.27</td>
<td>2b</td>
<td>36 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Ebert et al. 2008</td>
<td>36.9</td>
<td>MACI w/accelerated rehabilitation or traditional rehabilitation</td>
<td>0.25</td>
<td>K, S</td>
<td>62</td>
<td>MFC, LFC</td>
<td>3.3</td>
<td>2b</td>
<td>56 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Erggelet et al. 2000</td>
<td>33.7</td>
<td>ACI-P</td>
<td>0.5, 1</td>
<td>M</td>
<td>13</td>
<td>MFC, LFC, Troc, Pat</td>
<td>6.27</td>
<td>2b</td>
<td>36 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Farr et al. 2007</td>
<td>36.9</td>
<td>ACI-P c/Meniscus allograft transplant</td>
<td>4.5</td>
<td>L, M</td>
<td>29</td>
<td>MFC, LFC, Kissing</td>
<td>6.27</td>
<td>2b</td>
<td>46 (13)</td>
<td>Yes</td>
</tr>
<tr>
<td>Gobbi et al. 2006</td>
<td>30.5</td>
<td>MACI patellofemoral</td>
<td>2</td>
<td>I</td>
<td>32</td>
<td>Pat, Troc</td>
<td>4.7</td>
<td>2b</td>
<td>61 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Gobbi et al. 2009</td>
<td>31.2</td>
<td>MACI patellofemoral</td>
<td>2, 6.29</td>
<td>I</td>
<td>34</td>
<td>Pat, Troc</td>
<td>4.45</td>
<td>2b</td>
<td>61 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Henderson and Lavigne 2006</td>
<td>33.6</td>
<td>ACI-P patellofemoral with or without realignment</td>
<td>0.75, 1, 2</td>
<td>I, M, S</td>
<td>44</td>
<td>FC, Troc, Pat</td>
<td>3.07</td>
<td>2b</td>
<td>46 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Henderson et al. 2005</td>
<td>41</td>
<td>ACI-P</td>
<td>1.2</td>
<td>I</td>
<td>53</td>
<td>MFC, LFC, Troc, Pat</td>
<td>3.7</td>
<td>2b</td>
<td>58 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Henderson et al. 2006</td>
<td>38.8</td>
<td>ACI-P with or without reoperation</td>
<td>3.52</td>
<td>I, M, S</td>
<td>170</td>
<td>MFC, LFC, Troc, Pat</td>
<td>3.45</td>
<td>2b</td>
<td>50 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Horas et al. 2003</td>
<td>31.4</td>
<td>ACI-P</td>
<td>0.5, 1, 2</td>
<td>L</td>
<td>20</td>
<td>MFC, LFC, PFJ</td>
<td>3.86</td>
<td>2b</td>
<td>60 (11)</td>
<td>No</td>
</tr>
<tr>
<td>Knutsen et al. 2004</td>
<td>33.3</td>
<td>ACI-P</td>
<td>1.2</td>
<td>L, S</td>
<td>39</td>
<td>MFC, LFC</td>
<td>4.5</td>
<td>2b</td>
<td>68 (17)</td>
<td>No</td>
</tr>
<tr>
<td>Kon et al. 2009</td>
<td>29</td>
<td>MACI</td>
<td>5</td>
<td>I</td>
<td>40</td>
<td>MFC, LFC, Troc</td>
<td>8.5</td>
<td>2b</td>
<td>51 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Kreuz et al. 2009</td>
<td>35</td>
<td>MACI</td>
<td>0.5, 1, 4</td>
<td>I, K</td>
<td>19</td>
<td>MFC, LFC, Pat</td>
<td>4</td>
<td>2b</td>
<td>49 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Mandelbaum et al. 2007</td>
<td>37.1</td>
<td>ACI-P</td>
<td>4.91</td>
<td>M, MP, MS</td>
<td>40</td>
<td>Troc</td>
<td>4.5</td>
<td>2b</td>
<td>44 (11)</td>
<td>Yes</td>
</tr>
<tr>
<td>Marcacci et al. 2005</td>
<td>37.6</td>
<td>MACI</td>
<td>1.41, 3.17</td>
<td>I</td>
<td>141</td>
<td>FC, Troc, Pat, TP</td>
<td>3.5</td>
<td>2b</td>
<td>49 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>McNickle et al. 2009</td>
<td>30.3</td>
<td>ACI-P</td>
<td>4.3</td>
<td>I, K, L</td>
<td>122</td>
<td>MFC, LFC, Troc, Pat, Troc and Pat</td>
<td>4.21</td>
<td>2b</td>
<td>52 (13)</td>
<td>Yes</td>
</tr>
<tr>
<td>Micheli et al. 2006</td>
<td>15.5</td>
<td>ACI-P</td>
<td>4.3</td>
<td>M, MP, MS</td>
<td>32</td>
<td>MFC, LFC</td>
<td>5.4</td>
<td>2b</td>
<td>35 (7)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Study</th>
<th>Mean Age</th>
<th>Procedure Includeda</th>
<th>Follow-up Time (years)</th>
<th>Instrument Included in Reviewb</th>
<th>Total N Analyzed</th>
<th>Lesion Locationsc</th>
<th>Average Lesion Size (or Largest Lesion Size) (cm²)</th>
<th>Level of Evidenced</th>
<th>Mod. Coleman Methodology Score²/²⁴ (Part B)</th>
<th>Potential Conflict of Interest Disclosed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minas and Bryant 2005⁵⁰</td>
<td>36.9</td>
<td>ACI-P patellofemoral</td>
<td>3.95</td>
<td>M, S, W</td>
<td>45</td>
<td>Pat, Troc, Pat and Troc, FC and Pat, FC and Troc, FC and Pat and Troc</td>
<td>10.45</td>
<td>2b</td>
<td>56 (17)</td>
<td>Yes</td>
</tr>
<tr>
<td>Mithöfer et al. 2005⁵¹</td>
<td>15.9</td>
<td>ACI-P among adolescent athletes</td>
<td>3.91</td>
<td>L</td>
<td>20</td>
<td>MFC, LFC, Troc, Pat and Troc, FC and Pat, FC and Troc, FC and Pat and Troc</td>
<td>6.4</td>
<td>2b</td>
<td>43 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Moseley et al. 2010⁵²</td>
<td>37</td>
<td>ACI-P</td>
<td>9.2</td>
<td>M, MP, MS</td>
<td>72</td>
<td>MFC, LFC, Troc</td>
<td>5.2</td>
<td>2b</td>
<td>52 (13)</td>
<td>Yes</td>
</tr>
<tr>
<td>Nehrer et al. 2006⁶²</td>
<td>33</td>
<td>MACI</td>
<td>1.3</td>
<td>I, L, M</td>
<td>36</td>
<td>MFC, LFC, Pat, TP</td>
<td>1.5-8 (range)</td>
<td>2b</td>
<td>39 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Niemeyer et al. 2010⁵³</td>
<td>39.4</td>
<td>ACI-C to those 40 and older and younger than 40</td>
<td>0.5, 2</td>
<td>I, L</td>
<td>74</td>
<td>MFC, LFC, Troc, Pat</td>
<td>Not reported</td>
<td>2b</td>
<td>42 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Niemeyer et al. 2010⁵³</td>
<td>37.4</td>
<td>ACI-C</td>
<td>0.5, 1</td>
<td>I, L</td>
<td>66</td>
<td>MFC, LFC, Troc, Pat</td>
<td>4.3</td>
<td>2b</td>
<td>49 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Ochi et al. 2002⁵⁴</td>
<td>26.4</td>
<td>Atelocollagen-associated ACI with periosteum flap</td>
<td>2</td>
<td>L</td>
<td>28</td>
<td>MFC, LFC, Pat</td>
<td>2.93</td>
<td>2b</td>
<td>45 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Ossendorf et al. 2007⁵⁵</td>
<td>36</td>
<td>MACI</td>
<td>0.5, 1, 2</td>
<td>K</td>
<td>40</td>
<td>MFC, LFC, Pat, TP</td>
<td>4.6</td>
<td>2b</td>
<td>61 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Pascual-Garrido et al. 2009⁵⁶</td>
<td>31.8</td>
<td>ACI-P patellofemoral</td>
<td>4</td>
<td>I, K, L, S</td>
<td>52</td>
<td>PAT, Troc, bipolar, Troc and MFC</td>
<td>4.2</td>
<td>2b</td>
<td>46 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Peterson et al. 2010⁵⁷</td>
<td>33.3</td>
<td>ACI-P</td>
<td>12.8</td>
<td>L</td>
<td>58</td>
<td>FC, Pat</td>
<td>7</td>
<td>2b</td>
<td>39 (10)</td>
<td>No</td>
</tr>
<tr>
<td>Robertson et al. 2007⁵⁸</td>
<td>37.4</td>
<td>ACI-C</td>
<td>0.5, 1, 2</td>
<td>K</td>
<td>27</td>
<td>MFC, LFC, Pat</td>
<td>1-10 (range)</td>
<td>2b</td>
<td>56 (13)</td>
<td>No</td>
</tr>
<tr>
<td>Rosenberger et al. 2008⁵⁹</td>
<td>48.6</td>
<td>ACI-P over age 45</td>
<td>2.3</td>
<td>M, S, W</td>
<td>56</td>
<td>MFC, LFC, Troc, Pat</td>
<td>4.7</td>
<td>2b</td>
<td>61 (22)</td>
<td>Yes</td>
</tr>
<tr>
<td>Rue et al. 2008⁶⁰</td>
<td>23.4</td>
<td>ACI-P w/Meniscal allograft transplant</td>
<td>2</td>
<td>I, K, L</td>
<td>15</td>
<td>MFC, LFC, Pat, MTP LTP</td>
<td>3.93</td>
<td>2b</td>
<td>43 (10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Saris et al. 2008⁶¹</td>
<td>33.9</td>
<td>CCI</td>
<td>0.5, 1.5</td>
<td>K</td>
<td>51</td>
<td>FC</td>
<td>2.6</td>
<td>1b</td>
<td>68 (17)</td>
<td>Yes</td>
</tr>
<tr>
<td>Saris et al. 2009⁶²</td>
<td>33.9</td>
<td>CCI</td>
<td>3</td>
<td>K—Total only</td>
<td>41</td>
<td>FC</td>
<td>2.6</td>
<td>1b</td>
<td>62 (13)</td>
<td>Yes</td>
</tr>
<tr>
<td>Selmi et al. 2008⁶³</td>
<td>30</td>
<td>MACI</td>
<td>1.2</td>
<td>I</td>
<td>17</td>
<td>FC</td>
<td>3</td>
<td>2b</td>
<td>64 (17)</td>
<td>No</td>
</tr>
<tr>
<td>Tohyama et al. 2009⁶⁴</td>
<td>≥20</td>
<td>Atelocollagen-associated ACI with periosteum flap</td>
<td>0.5, 1, 2</td>
<td>I, L</td>
<td>27</td>
<td>MFC, LFC, Pat</td>
<td>3.2</td>
<td>2b</td>
<td>39 (13)</td>
<td>Yes</td>
</tr>
<tr>
<td>Zaslav et al. 2009⁶⁵</td>
<td>34.5</td>
<td>ACI-P following prior failed treatment w/n previous 3 years</td>
<td>0.5, 1, 3, 4</td>
<td>M, K, S</td>
<td>150</td>
<td>MFC, LFC, Troc</td>
<td>4.63</td>
<td>2b</td>
<td>52 (18)</td>
<td>Yes</td>
</tr>
<tr>
<td>Zeifang et al. 2010⁶⁶</td>
<td>29.1</td>
<td>ACI-P, MACI</td>
<td>1</td>
<td>I, L, S</td>
<td>21</td>
<td>MFC, LFC</td>
<td>4.20</td>
<td>2b</td>
<td>67 (13)</td>
<td>Yes</td>
</tr>
</tbody>
</table>


bI: International Knee Documentation Committee Subjective Knee Form (IKDC), L: Lysholm Knee Scale (Lysholm), K: Knee Injury and Osteoarthritis Outcome Score (KOOS), M: modified Cincinnati Knee Rating System (MCKRS) Patient Perspective: MP: MCKRS – Pain Scale; MS: MCKRS Swelling Scale, S: Medical Outcomes Study 36-Item Short Form Health Survey Physical Component Scale (SF-36 PCS), W: Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).


dScored 0 to 100 with 100 representing best methodology; Part B subscore is presented as an assessment of study bias scored 0 to 22 with 22 representing least bias.
Table 2. Orwin’s Fail Safe N Analysis to Evaluate Publication Bias

<table>
<thead>
<tr>
<th>Instrument</th>
<th>N(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKDC</td>
<td>95</td>
</tr>
<tr>
<td>Lysholm</td>
<td>83</td>
</tr>
<tr>
<td>KOOS</td>
<td>196</td>
</tr>
<tr>
<td>MCKRS</td>
<td>48</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>14</td>
</tr>
<tr>
<td>Overall across all instruments</td>
<td>399</td>
</tr>
</tbody>
</table>

IKDC = International Knee Documentation Committee Subjective Knee Form; Lysholm = Lysholm Knee Scale; KOOS = Knee Injury and Osteoarthritis Outcome Score; MCKRS = Modified Cincinnati Knee Rating System; SF-36 PCS = Medical Outcomes Study 36-Item Short Form Health Survey Physical Component Scale.

*Number of studies with an effect size of 0.1 needed to reduce the overall mean effect size to 0.4.

Discussion

Our purpose was to evaluate the responsiveness of common PROs to the treatment effects of ACI. An underlying assumption was that ACI would have a common effect across studies and varying ACI procedures. Evaluating ACI efficacy was not a purpose of this review, and the results are in agreement with previous reviews documenting ACI to be a viable procedure resulting in positive patient outcomes.\(^{24,69,70}\) The large mean effect sizes and narrow CIs observed in this review support the use of ACI for the generalized treatment of articular cartilage defects.

Responsiveness

The results of this review demonstrate that regardless of the duration of postoperative follow-up all instruments...
were responsive to patient improvement; however, the IKDC and Lysholm may be more responsive than the MCKRS, KOOS, or SF-36 PCS. There was insufficient data to adequately evaluate the WOMAC.

Responsiveness of PROs across TPs

The Lysholm demonstrated large mean effect sizes (1.30-1.70) with little variation across time (Fig. 2). The observed CIs for the Lysholm at all TPs overlap by more than 50%, suggesting little changes in responsiveness as time since ACI progresses. Common rehabilitation recommendations following ACI restrict return to sports participation for 12 to 18 months following surgery.76-79 This delayed return to physical activity may result in lower scores on instruments that emphasize higher sport demands. Because the Lysholm primarily assesses everyday activities (walking, squatting, stair-climbing) and does not address sports activity, delayed return to higher level physical activity has little influence on Lysholm score. Therefore, the Lysholm scale may be ideal for evaluating short-term outcomes or outcomes among patients not intending to return to sports, but less responsive to changes during long-term recovery as individuals return to higher demand activities. Additionally, it should be noted that some authors and investigations have called into question the current weighting system for Lysholm items.74,75

The IKDC also demonstrated large effect sizes. With a noticeable increase in mean effect size observed between TP-I and TP-IV, with mean effect size increasing from 0.88 to 1.78 with no overlap between CIs. This difference demonstrates increased treatment effects over time when evaluating outcomes with the IKDC. Greco et al.20 observed a similar trend with responsiveness of the IKDC increasing between 6 and 12 months in a cohort of surgical cartilage patients. It has previously been reported76-78 that functional and structural improvements following cartilage repair continue beyond 1 year postoperatively. The observed increases in mean effect size over time may represent the IKDC’s responsiveness to continual improvements in function that occur in the years following ACI. The responsiveness of the IKDC to continued improvements over time can be considered a strength of this instrument and may be due to its inclusion of sporting activities. The IKDC allows for continued improvement as individuals initiate return to strenuous activity and sports participation beyond 1 year postoperatively.

The KOOS-Sports subscale had the lowest mean effect at TP-I and TP-II, whereas the KOOS-Symptoms subscale had the lowest mean effect of all the KOOS scales at TP-III and TP-IV. Effect sizes for the KOOS-Sports subscale were lower at TP-I compared with TP-III and TP-IV. These results are similar to those observed with the IKDC, and this progressive increase in effect sizes over time may be related to the slow, progressive return to sports following ACI. For all other KOOS subscales no changes were seen for mean effect size between TPs. Overall, the KOOS was responsive to changes following ACI; however, the KOOS-Sports subscale was the only subscale to demonstrate increasing responsiveness over time, suggesting that it responded to increasing treatment effects as healing progressed and may be more sensitive to improvements in function among active individuals than other instruments or KOOS subscales.

There were only sufficient data to evaluate the MCKRS at TP-III and TP-IV, limiting any conclusions that can be drawn regarding the changes in its effect sizes over time. Our results suggest the MCKRS is responsive to changes in patient function following ACI; however, caution is urged regarding the use of this instrument. Many different versions of the MCKRS exist and many authors fail to reference the version of MCKRS they use. Several articles were excluded, at least in part, because the authors did not reference the version of the MCKRS used, or because a different version than the one presented by Browne et al.14 was used as an outcome measure.76,79-87 Because of ambiguity regarding the use of “modified” Cincinnati Knee Rating Systems, the developers of the original Cincinnati Knee Rating Scale discourage the use of any modified versions.88 However,
because of the frequency with which the Browne et al.\textsuperscript{14} version of the MCKRS has been clearly referenced in ACI outcomes studies, it was chosen for inclusion in this review.

Both the SF-36 PCS and the WOMAC had limited data available for analysis. For the SF-36 PCS, there was only sufficient data for analysis of responsiveness at TP-III. For this TP, the SF-36 PCS did demonstrate a positive mean effect 2 to 4 years following ACI with an effects size of 0.92 [0.55, 1.28]. There were insufficient data to include the WOMAC in any of the meta-analyses performed.\textsuperscript{50,61} Although additional studies have included the WOMAC as an outcome measure, the results were only reported using nonparametric statistics and/or without the reporting of means and standard deviations or other data necessary for calculating effect sizes.\textsuperscript{89-92} As a result no clear conclusions regarding the responsiveness of the WOMAC as an outcome instrument can be reached based on this review.

**Responsiveness between PROs**

The forest plots of PRO instruments for each TP can be seen in fig 3, whereas the overall mean effect sizes averaged across all TPs can be seen in fig 4. The IKDC and the KOOS-Sports were the only instruments to demonstrate changes in effect sizes over time. These changes may be related to activity restriction and gradual return to sports following ACI. The restrictions on sporting activity during the first year post-ACI may also explain the significant differences observed between the responsiveness of the KOOS-Sports and the Lysholm at TP-I and TP-II. At TP-II and TP-IV, the IKDC appears more responsive than the KOOS-Sports. These differences may be the result of the wider range of physical functioning addressed in the IKDC as compared with the KOOS-Sports. The responsiveness of the MCKRS was not different from any other instrument evaluated at both TP-III and TP-IV. The SF-36 PCS had the lowest responsiveness overall and at TP-III. This finding is not surprising as the SF-36 was the only included instrument not specifically designed for the knee.

The Lysholm and the IKDC demonstrated the largest overall effect sizes, regardless of TP. These had appreciably greater responsiveness than the KOOS-Sports subscale and the SF-36 PCS (Fig. 4). Should investigators or clinicians wish to explore patient outcomes for individual constructs (quality of life, activities of daily living, sports, etc.), the KOOS via its subscales is the only instrument that allows for this multifaceted investigation, and along with the IKDC has been recommended for use by the International Cartilage Repair Society.\textsuperscript{75} Although both the KOOS and IKDC include sports participation as components of evaluating knee function, the IKDC appears more responsive to overall changes in function following ACI (Fig. 4). This overall difference, combined with the observed differences in responsiveness between the IKDC and KOOS-Sports subscales at TP-II and TP-IV, leads us to propose that the IKDC may be the preferred outcome instrument for evaluating long-term outcomes following ACI, particularly among patients whose goals include return to sporting activity. Although all KOOS subscales are responsive to treatment effects following ACI, the IKDC and Lysholm are shorter instruments with single score outcomes and overall are more responsive to change than some subscales included in the KOOS. Based on these observations, the IKDC and the Lysholm may be preferable to the KOOS for documenting treatment effects following ACI.

**Study Quality**

The mean modified Coleman Methodology Score (50.9 ± 9.2) among studies was comparable to other recent reviews of ACI and other cartilage repair procedures.\textsuperscript{44,69,93} Although the modified Coleman Methodology Score provides a set of standardized criteria by which to evaluate cartilage research, it is not without limitations. The scale is heavily weighted toward diagnostic, clinician-oriented outcomes, with up to 25% of the score dependent on MRI and histological evaluation. The relationship between MRI and clinical outcomes is not definitive; some authors observed low to moderate correlations between MRIs and PROs,\textsuperscript{50,94,95} and others failed to observe such a relationship.\textsuperscript{96,97} Similarly, histological analysis can involve a wide variety of techniques and may not be ethical in cases where reoperation is not otherwise indicated. Of the 42 studies included, only a single study\textsuperscript{65} received full credit for both histological and MRI outcomes, suggesting that the requirement of these outcomes may not be applicable in a clinical research setting. Furthermore, only five studies scored a full 10 points for >90% of subjects undergoing one surgical procedure with <10% undergoing concomitant procedures.\textsuperscript{29,42,43,62,65} Notably, although concomitant procedures reduced the overall methodological score, studies that include concomitant procedures are more generalizable to real clinical practice than studies of single isolated defects.\textsuperscript{98}

In future research, more well-designed, well-documented, high-level clinical trials that use PROs with comprehensive data reporting are needed. Adopting uniform methodological reporting requirements for cartilage repair studies will improve the quality of the body of literature in this area. This review may provide a basis for this effort.

**Limitations**

The results of this review are limited by the quantity, quality, and strength of the studies and PROs selected for inclusion. Any recommendations made are based solely on the available evidence, and it should be noted that the IKDC and Lysholm were used in the literature more often than other instruments, strengthening the validity of
recommendations regarding these two instruments and limiting our ability to draw conclusions regarding other PROs. As evidenced by the low modified Coleman Methodology Score observed in this review and others, the quality of reporting in cartilage outcomes studies is variable and generally poor. A random effects analysis was used to account for the variability between studies allowing our results to be generalized to a broad clinical population.

A statistical limitation of our study is the use of multiple measures at multiple TPs from within the same study populations. We acknowledge that outcome scores obtained from within the same sample are likely correlated, but given that the correlation between outcome measures and TPs is rarely reported and no studies documented all instruments at all TPs, correction for this relationship was not feasible. Fortunately, the observed mean effect sizes are so large and the CIs so small for the included outcome instruments that we do not believe this assumption violation significantly influences the overall conclusions of this review.

Conclusions
Evidence for the use of ACI as a treatment for chondral defects consists primarily of level 2b observational cohort studies. The methodological quality of many of these studies is limited by the absence of diagnostic outcomes such as MRI and histological analyses, small sample size, short follow-up, and high frequency of concomitant procedures. In addition, documentation of recruitment rate and investigator independence was lacking from many studies. The IKDC, Lysholm, KOOS, MCKRS, and SF-36 PCS were all responsive to improvements in function following ACI. A positive treatment effect for ACI was observed using all instruments with follow-up ranging from <1 year to beyond 4 years. The Lysholm and the IKDC were the most responsive instruments across time. The Lysholm was highly responsive as early as <1 year following ACI and was consistently responsive throughout follow-up. However, this instrument may not be responsive to changes in function associated with the resumption of higher demand activities such as sports that occurs after 1 year. For the evaluation of long-term outcomes among patients with intent to return to physical activity, this review supports the use of the IKDC, which was able to detect increasing treatment effects over time. The use of the Lysholm and IKDC together represents a responsive combination of PRO instruments that are able to efficiently document both short-term and long-term treatment effects among patients of a variety of activity levels following ACI.

Acknowledgments and Funding
We acknowledge Heather Bush, PhD, Assistant Professor of Biostatistics, and the Applied Statistics Laboratory at the University of Kentucky for assistance with the preparation of this article. This research was completed at the University of Kentucky, Lexington, Kentucky, USA.

Declaration of Conflicting Interests
The author(s) declared the following potential conflicts of interest: JSH became a paid consultant for Zimmer Inc. and for Sanofi/Genzyme Corporation. After initial submission of this article, JSH became a paid consultant for Sanofi/Genzyme Corporation. However, the relationships with these corporations did not influence this research in any manner.

References


