A Comparison of Pain, Strength, Range of Motion, and Functional Outcomes After Hemiarthroplasty and Total Shoulder Arthroplasty in Patients with Osteoarthritis of the Shoulder

A Systematic Review and Meta-Analysis

BY DIANNE BRYANT, MSC, PHD, ROBERT LITCHFIELD, MD, FRCSC, MICHAEL SANDOW, BMBS, FRACS, GARY M. GARTSMAN, MD, GORDON GUYATT, MD, MSC, AND ALEXANDRA KIRKLEY, MD, MSC, FRCSC

Investigation performed at the Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

Background: A systematic review of the literature was performed to estimate the impact of hemiarthroplasty compared with total shoulder arthroplasty on function and range of motion in patients suffering from osteoarthritis of the shoulder.

Methods: We conducted an electronic search for relevant studies published in any language from 1966 to 2004, a manual search of the proceedings from five major orthopaedic meetings from 1995 to 2003, and a review of the reference lists from potentially relevant studies. Four randomized clinical trials, with similar eligibility criteria and surgical techniques, that compared hemiarthroplasty and total shoulder arthroplasty for the treatment of primary osteoarthritis of the shoulder were found to be eligible. Authors from three of the four studies provided original patient data. Analysis of covariance focused on the two-year outcome and included a comparison of the aggregate University of California at Los Angeles shoulder score, four University of California at Los Angeles domain scores, and range of motion.

Results: A total of 112 patients (fifty managed with hemiarthroplasty and sixty-two managed with total shoulder arthroplasty), who had a mean age of sixty-eight years, were included in this analysis. A significant moderate effect was detected in the function domain of the University of California at Los Angeles shoulder score (p < 0.001) in favor of total shoulder arthroplasty (mean [and standard deviation], 8.1 ± 0.3) compared with hemiarthroplasty (mean, 6.6 ± 0.3). A significant difference in the pain score was found in favor of the total shoulder arthroplasty group (p < 0.0001). However, the large degree of heterogeneity (p = 0.006, \( I^2 = 80.2\% \)) among the studies decreased our confidence that total shoulder arthroplasty provides a true, consistent benefit with regard to pain. There was a significant difference in the overall change in forward elevation of 13° (95% confidence interval, 0.5° to 26°) in favor of the total shoulder arthroplasty group (p = 0.008).

Conclusions: At a minimum of two years of follow-up, total shoulder arthroplasty provided better functional outcome than hemiarthroplasty for patients with osteoarthritis of the shoulder. Since continuous degeneration of the glenoid after hemiarthroplasty or glenoid loosening after total shoulder arthroplasty may affect the eventual outcome, longer-term (five to ten-year) results are necessary to determine whether these findings remain consistent over time.

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

N eer popularized shoulder hemiarthroplasty to treat humeral head fractures, osteonecrosis of the humeral head, and arthritis of the shoulder joint. Problems arose, however, if the surrounding musculotendinous rotator cuff was deficient or if the articular surface of the glenoid was abnormal. Additionally, patients who underwent hemiarthroplasty ran the risk of continued erosion of the glenoid surface, leading to increased pain and decreased function over time. Thus, hemiarthroplasty began to give way to total shoulder arthroplasty, which involves the replacement of both the humeral head and the glenoid. Total shoulder arthroplasty, however, is not without problems. These include the increased complexity of the procedure, increased blood loss, increased operating-room time, and increased cost. In addition, some patients who...
are followed for longer time-periods show signs of glenoid component loosening, which may lead to decreased function. Several cohort studies have compared the outcomes after hemiarthroplasty with those after total shoulder arthroplasty and had inconsistent conclusions as to which procedure is best. Since considerable controversy remains as to whether hemiarthroplasty or total shoulder arthroplasty is the better treatment option for patients with osteoarthritis, the purpose of this study was to conduct a systematic review and meta-analysis of randomized clinical trials that primarily compared the effectiveness of hemiarthroplasty and total shoulder arthroplasty at two years postoperatively in patients with primary osteoarthritis of the shoulder.

**Materials and Methods**

**Literature Search**

We undertook a computer-assisted search of the online bibliographic databases MEDLINE (1966 through week 2 of June 2004), CINAHL (1982 through week 2 of June 2004), and EMBASE (1980 through week 25 of 2004) and the Cochrane Central Register of Controlled Trials (second quarter of 2004) to identify relevant studies published in any language. We used the Cochrane strategy for randomized controlled trials and database-appropriate search terms, including arthroplasty and excluding articles on hip and knee arthroplasty, and we further limited the search by using the text word “shoulder.” In addition, we conducted a manual search of the proceedings from five major orthopaedic meetings from 1995 to December 2003. A review of the reference lists from potentially relevant studies complemented other searches. The specific search strategies are available from the authors.

**Selection of Studies**

Eligible studies included randomized trials that compared hemiarthroplasty and total shoulder arthroplasty in patients with a diagnosis of primary osteoarthritis of the shoulder who were followed for a minimum of two years and that used a measure of functional outcome and/or range of motion (forward elevation and/or external rotation). The electronic search produced 311 abstracts. Both reviewers identified only one eligible study and, on the basis of a review of the full text, this paper was included in the study. The agreement between reviewers was 100% for both the review of the abstract and the full text.

Both reviewers identified four abstracts from conference proceedings. One of these abstracts was from the published study that was included and three were from studies that had not yet been published. One of the abstracts was presented more than once at different conferences over the five years reviewed. Thus, only the abstract containing data for the two-year assessment was used.

**Validity Assessment**

The authors of all included studies were contacted before assessing the quality of each study since three of the four studies were present in abstract form only. This was done to request data and to obtain further details on the study design. Quality was assessed according to six criteria, including concealment of randomization, blinding of patients and data collectors, proportion of subjects randomized who dropped out, use of validated outcomes, and use of the intention-to-treat method of analysis. Quality assessment was not used to determine inclusion or exclusion of trials into this meta-analysis, but it served to generate hypotheses to explain potential heterogeneity between trials and helped to decide on the appropriate strength of inference from the data.

**Potential Heterogeneity**

Following quality assessment, we defined three a priori hypotheses to address potential differences in treatment effect between studies (variability in the magnitude of effect, also referred to as heterogeneity of results) that cannot be explained by chance alone. These hypotheses included a lack of concealment of randomization, whether data collectors were aware of patient group allocation (blinding), and use of uncemented or cemented fixation of the humeral prosthesis. These hypotheses were tested with use of sensitivity analyses.

**Data Abstraction**

Three of the four primary investigators of the studies that were included contributed their original data set to create a common data set for the meta-analysis. The University of California at Los Angeles (UCLA) functional shoulder assessment tool was common among all studies and was thus selected as the primary outcome for this meta-analysis. This instrument assigns a score to patients on the basis of five separate domains: pain, function, active forward flexion, strength of forward flexion, and overall satisfaction. There is one item for each of these areas. The weighting is such that pain accounts for 10 points; function, 10 points; forward flexion, 5 points; strength, 5 points; and overall satisfaction, 5 points, giving a total of 35 points (see Appendix). The overall satisfaction item was omitted since it can be collected only after an intervention and not before and after an intervention, as is ideal in most clinical trials. The fourth study did not collect UCLA data but did collect functional data with use of the scoring system of Constant and Murley. Other commonly used outcome measurements were range-of-motion measurements, including active forward elevation and external rotation.

**Statistical Analysis**

The intention-to-treat principle, whereby all patients are analyzed in the group to which they were randomized, guided all analyses. Use of the intention-to-treat principle has the potential to provide a conservative estimate of treatment effect. However, by conducting an efficacy analysis, which violates the intention-to-treat principle by analyzing patients according to the treatment they received (either by removing the data of patients who were randomized to hemiarthroplasty but crossed over to total shoulder arthroplasty, or by including these data in the total shoulder arthroplasty group), we almost certainly introduce a biased estimate of the treatment effect in favor of the hemiarthroplasty group. Specifically, since not only would the balance of the prognostic factors between the
groups that was established through randomization be compromised, but by removing the data from patients who had failure of the hemiarthroplasty and crossed over to total shoulder arthroplasty (especially if we include these data in the total shoulder arthroplasty group), we would bias the estimate of effect in favor of hemiarthroplasty.

The analysis was conducted with use of two statistical programs. The first program, Review Manager (version 4.1, 2003; The Cochrane Collaboration, Oxford, United Kingdom), uses the inverse of the variances for each study to weight its treatment effect in the pooled analysis, thereby assigning greater weight to larger studies with a more precise estimate of the treatment effect. Because it provides a more conservative estimate of the precision of treatment effect, we chose to conduct the pooled analyses with the random-effects model, which considers both the between-study and within-study variance in treatment effects, over the fixed effects model, which considers only the within-study variance. The random-effects model assumes that studies included in the meta-analysis are only a sample of all possible studies that could be conducted, while the fixed-effects model assumes that the studies that are included represent the entire population of studies that could be conducted on the topic.

When the estimates of effect across studies are pooled, the underlying assumption is that the effect is similar across the range of patients, interventions, and outcomes among the studies. A formal assessment of the variability of results across studies (heterogeneity) allows reviewers to test the validity of this assumption. For each pooled analysis, we included two statistical tests that provide insight into the degree of heterogeneity: the Cochran chi-square test\(^1\) and the \(I^2\) value\(^2\). The Cochran chi-square test provides an opportunity to reject the null hypothesis that the underlying effect is the same across studies but may be underpowered when only a small number of studies are available. The \(I^2\) value estimates the percentage of total variation in results across studies that is due to heterogeneity among studies rather than that due to chance. Low, moderate, and high levels of heterogeneity are roughly categorized by \(I^2\) values of 25%, 50%, and 75%, respectively.

The second analysis was conducted with use of SPSS software (release 11.0; SPSS, Chicago, Illinois) and the individual patient data from three of the four investigators. Analysis of covariance, which is analogous to a t test but allows for adjustment of the analysis for baseline scores, gives a more powerful estimate of the treatment effect than that provided by Review Manager.

Since we were uncertain of whether to include the data from the trial by Jonsson et al.\(^3\) (i.e., no concealment of randomization followed by unbalanced exclusion of randomized patients), we conducted a sensitivity analysis to determine the robustness of the results when data from that trial were added. In that study, five (9%) of the fifty-six patients randomized, or 17% of the patients managed with total shoulder arthroplasty, in whom the glenoid was found intraoperatively to be insufficient to support a glenoid prosthesis, were withdrawn from the trial. Similar patients were not withdrawn from the hemiarthroplasty group. This almost certainly introduces bias in favor of total shoulder arthroplasty and underscores the importance of including all patients who are randomized in the analysis and in their original groups. Specifically, the hemiarthroplasty group is now more likely to include patients with severe glenoid bone loss or defects, which will most likely bias this group toward poorer outcomes. To conduct the sensitivity analysis, we used the standardized mean difference of the change score since different functional instruments (the UCLA system\(^4\) and the Constant and Murley system\(^5\)) had been used. A similar sensitivity analysis was conducted with use of the weighted mean difference of the change score for the range-of-motion data since all measurements were conducted with use of a standard universal goniometer that measured in degrees.

### Results

#### Characteristics of the Studies That Were Included

Four trials were eligible for inclusion into the meta-analysis. The trial conducted by Gartsman et al.\(^6\) was the only one that had published results. Their study included fifty-one patients (fifty-five shoulders) who were randomly assigned to hemiarthroplasty or total shoulder arthroplasty. Four patients were lost to follow-up and were not counted in the analysis. Forty-seven patients (fifty-one shoulders) remained; twenty-four shoulders had a hemiarthroplasty, and twenty-seven shoulders had total shoulder arthroplasty. Hemiarthroplasty and total shoulder arthroplasty were performed with use of the Global Shoulder Arthroplasty System (DePuy Orthopaedics, Warsaw, Indiana). The criteria for inclusion into the trial included a diagnosis of osteoarthritis, an intact rotator cuff, and a concentric glenoid. Diagnoses other than osteoarthritis were criteria for exclusion. An unblinded data collector assessed shoulder function preoperatively and postoperatively at two weeks, six weeks, three months, six months, and one year after the operation and yearly thereafter. Shoulder function was assessed with use of the UCLA shoulder score\(^7\), the American Shoulder and Elbow Surgeons (ASES) standardized rating system\(^8\), and range-of-motion assessment. The investigators reported that the total shoulder arthroplasty group had significantly less pain \((p = 0.002)\) and significantly greater internal rotation \((p = 0.003)\) than the hemiarthroplasty group. Unsatisfactory results were reported by five patients who had a hemiarthroplasty and by one who had total shoulder arthroplasty. Three of the five shoulders with an unsatisfactory result after hemiarthroplasty were revised to total shoulder arthroplasty at nineteen, thirty-nine, and forty-eight months postoperatively.

Lo et al.\(^9\) conducted a trial (published in abstract form only) that included thirty-nine patients (forty-one shoulders) randomly assigned to hemiarthroplasty (twenty-one shoulders) or total shoulder arthroplasty (twenty shoulders) performed with use of the Neer-II prosthesis (35M, St. Paul, Minnesota). Patients were randomized intraoperatively if they had a diagnosis of osteoarthritis of the shoulder, had no previous shoulder surgery, and had no cuff tear arthropathy, no major trauma to the shoulder, and no major medical illness, infection, paralysis of the study limb, or other concurrent diag-
noses, such as osteonecrosis or rheumatoid arthritis. A blinded data collector assessed shoulder function preoperatively and at two weeks, six weeks, and at three, six, twelve, eighteen, and twenty-four months postoperatively and yearly thereafter using the Western Ontario Osteoarthritis of the Shoulder (WOOS) index, ASES rating system, Constant and Murley system, UCLA shoulder score, McGill pain score, Short Form-12 (SF-12), and range-of-motion assessment. Patients were also unaware of their group assignment. No significant differences were reported for any outcome measure at one or two years postoperatively. Complications included two intraoperative fractures (one humeral fracture in the hemiarthroplasty group and one glenoid fracture in the total shoulder arthroplasty group), two late rotator cuff tears (one in each group) requiring no further surgical intervention, and one deep infection in the total shoulder arthroplasty group that resolved following arthroscopic lavage and a course of antibiotics. One patient in the total shoulder group died from a thromboembolism two weeks following surgery. Three patients underwent revision to total shoulder arthroplasty at twelve, eighteen, and forty-eight months postoperatively. Randomization was stratified according to the type of arthritis. Five patients were excluded during surgery subsequent to randomization because the bone stock appeared insufficient for glenoid replacement. Two patients were excluded following surgery for complications, including one deep infection and one nerve injury. A significant improvement in the Constant score was reported for the total shoulder arthroplasty group compared with the hemiarthroplasty group (p < 0.03). Four patients in the hemiarthroplasty group underwent revision to a total shoulder arthroplasty at twenty-four months (two shoulders), thirty-six months (one shoulder), and forty-eight months (one shoulder). Two patients in the hemiarthroplasty group died at six and forty-eight months postoperatively, and one patient in the total shoulder arthroplasty group died at twelve months postoperatively. This study was terminated prematurely by the institutional review board for efficacy prior to reaching the estimated sample size.

Jonsson et al. conducted a trial (published in abstract form only) that included fifty-six patients with both rheumatoid arthritis and osteoarthritis randomly assigned to hemiarthroplasty (thirteen with rheumatoid arthritis and eleven with osteoarthritis) or total shoulder arthroplasty (twelve with rheumatoid arthritis and thirteen with osteoarthritis) before surgery. Randomization was stratified according to the type of arthritis. Five patients were excluded during surgery subsequent to randomization because the bone stock appeared insufficient for glenoid replacement. Two patients were excluded following surgery for complications, including one deep infection and one nerve injury. A significant improvement in the Constant score was reported for the total shoulder arthroplasty group compared with the hemiarthroplasty group (p = 0.01). No difference was detected between the groups with respect to active forward elevation (p = 0.065) and active external rotation (p > 0.05). Subgroup analysis revealed no significant differences in outcomes when patients with rheumatoid arthritis were compared with patients with osteoarthritis. Nonetheless, we included only data collected from patients with osteoarthritis.

Table I presents a summary of the results of the quality assessment.

### Table I Quality Assessment

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Method of concealment of randomization</td>
<td>Concealed table of random numbers</td>
<td>Opaque envelopes</td>
<td>Opaque envelopes</td>
<td>Not reported</td>
</tr>
<tr>
<td>Patient blinded</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data collector blinded</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Percentage of patients who dropped out or were lost to follow-up</td>
<td>8%</td>
<td>3%</td>
<td>7%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Function</td>
<td>29%</td>
<td>3%</td>
<td>3%</td>
<td>Not reported</td>
</tr>
<tr>
<td>Forward elevation</td>
<td>31%</td>
<td>Not measured</td>
<td>3%</td>
<td>Not reported</td>
</tr>
<tr>
<td>External rotation</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
<td>Yes*</td>
</tr>
<tr>
<td>Use of validated outcomes</td>
<td>Function</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Range of motion</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Intention-to-treat method of analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

*These studies each used several validated quality-of-life outcome tools. However, the UCLA shoulder score, which was the only instrument in common among the studies, is not a validated instrument.
A comparison of outcomes after hemiarthroplasty and TSA for osteoarthritis of the shoulder

Patients

In three of the four studies, if both shoulders underwent arthroplasty, both shoulders were randomized into the trial, which violates the assumptions of the statistical tests of independent observations. Thus, we excluded data collected from the second shoulder of patients who had been randomized twice, which resulted in the removal of data for six shoulder arthroplasties. One hundred and nineteen patients remained. Of those, four were patients who died in the early postoperative period (two in the hemiarthroplasty group and two in the total shoulder arthroplasty group) and three were patients with missing data, leaving 112 patients (fifty in the hemiarthroplasty group and sixty-two in the total shoulder arthroplasty group) with complete data for analysis of function with use of the UCLA shoulder score. No data on forward elevation and external rotation data were available for ninety-nine patients and fifty-nine patients, respectively. All patients with range-of-motion data contributed scores to the UCLA data. These values do not include patients from the trial conducted by Jonsson et al.15.

Of the 112 patients, sixty-eight were female (thirty had a hemiarthroplasty and thirty-eight had a total shoulder arthroplasty). The mean age was 67.7 years for both treatment groups. The groups were similar with respect to the mean baseline value for the UCLA total score (10.5 for the hemiarthroplasty group and 10.4 for the total shoulder arthroplasty group), forward elevation (87.9° and 82.2°, respectively), and external rotation (20.4° and 22.0°, respectively). Data on the patient characteristics for each study are presented in Table II.

The patients in the study by Gartsman et al.6 were, on the average, three years younger than those in the study by Sandow et al.21 and five years younger than those in the study by Lo et al.17 (p < 0.0001). Furthermore, the preoperative functional scores reported by Gartsman et al.6 were worse than those in the study by Lo et al.17 and Sandow et al.21, which is reflected by an average difference of approximately 4 points in the baseline UCLA score (p < 0.0001). In addition, although the patients in the trials were similar with respect to preoperative forward elevation, those in the trial described by Gartsman et al.6 had sig-

### Table II: Comparison of Patient Characteristics Before Intervention

<table>
<thead>
<tr>
<th>Study</th>
<th>Characteristic</th>
<th>Hemiarthroplasty</th>
<th>Total Shoulder Arthroplasty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gartsman et al.6</td>
<td>No. of randomized patients (no. of shoulders)</td>
<td>26 (28)</td>
<td>25 (27)</td>
</tr>
<tr>
<td></td>
<td>No. of patients included in analysis</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Male patients (%)</td>
<td>41</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Age* (yr)</td>
<td>65.0 ± 1.8</td>
<td>64.7 ± 1.7</td>
</tr>
<tr>
<td></td>
<td>UCLA preop. score*</td>
<td>8.2 ± 3.1</td>
<td>8.1 ± 2.5</td>
</tr>
<tr>
<td></td>
<td>Preop. forward elevation* (deg)</td>
<td>82.8 ± 37.5</td>
<td>84.4 ± 37.9</td>
</tr>
<tr>
<td></td>
<td>Preop. external rotation* (deg)</td>
<td>23.3 ± 26.1</td>
<td>25.4 ± 24.3</td>
</tr>
<tr>
<td>Lo et al.17</td>
<td>No. of randomized patients (no. of shoulders)</td>
<td>18 (21)</td>
<td>19 (20)</td>
</tr>
<tr>
<td></td>
<td>No. of patients included in analysis</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Male patients (%)</td>
<td>39</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Age* (yr)</td>
<td>71.3 ± 2.0</td>
<td>70.4 ± 2.0</td>
</tr>
<tr>
<td></td>
<td>UCLA preop. score*</td>
<td>12.8 ± 3.5</td>
<td>12.8 ± 3.8</td>
</tr>
<tr>
<td></td>
<td>Preop. forward elevation* (deg)</td>
<td>90.9 ± 24.4</td>
<td>85.3 ± 29.0</td>
</tr>
<tr>
<td></td>
<td>Preop. external rotation* (deg)</td>
<td>Not measured</td>
<td>Not measured</td>
</tr>
<tr>
<td>Sandow et al.17</td>
<td>No. of randomized patients (no. of shoulders)</td>
<td>11 (14)</td>
<td>18 (18)</td>
</tr>
<tr>
<td></td>
<td>No. of patients included in analysis</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Male patients (%)</td>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Age* (yr)</td>
<td>67.1 ± 2.7</td>
<td>69.0 ± 2.0</td>
</tr>
<tr>
<td></td>
<td>UCLA preop. score*</td>
<td>11.3 ± 2.3</td>
<td>11.1 ± 3.6</td>
</tr>
<tr>
<td></td>
<td>Preop. forward elevation* (deg)</td>
<td>91.5 ± 26.0</td>
<td>75.9 ± 34.1</td>
</tr>
<tr>
<td></td>
<td>Preop. external rotation* (deg)</td>
<td>15.0 ± 11.1</td>
<td>16.9 ± 12.4</td>
</tr>
<tr>
<td>Total</td>
<td>No. of randomized patients (no. of shoulders)</td>
<td>55 (63)</td>
<td>62 (65)</td>
</tr>
<tr>
<td></td>
<td>No. of patients included in analysis</td>
<td>50</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Male patients (%)</td>
<td>40</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Age* (yr)</td>
<td>67.7 ± 8.3</td>
<td>67.7 ± 9.2</td>
</tr>
<tr>
<td></td>
<td>UCLA preop. score*</td>
<td>10.5 ± 3.7</td>
<td>10.4 ± 3.8</td>
</tr>
<tr>
<td></td>
<td>Preop. forward elevation* (deg)</td>
<td>87.9 ± 30.2</td>
<td>82.2 ± 33.9</td>
</tr>
<tr>
<td></td>
<td>Preop. external rotation* (deg)</td>
<td>20.4 ± 22.1</td>
<td>22.0 ± 20.6</td>
</tr>
</tbody>
</table>

*The values are given as the mean and the standard deviation.
nificantly worse average external rotation at baseline than did the patients in the study by Sandow et al.²¹ (p = 0.023).

**Quantitative Data Synthesis**

Of the fifty patients randomized to hemiarthroplasty, ten patients (20%) crossed over to the total shoulder arthroplasty group because of excessive pain or stiffness; however, only three patients (6%) crossed over before the two-year assessment (at twelve, eighteen, and nineteen months). The remaining patients who had crossed over (as of the time of writing) did so at twenty-four months (two patients) and at thirty-six months (one patient), thirty-nine months (one patient), and forty-eight months (three patients).

**Functional Outcome**

When the mean change in the function domain of the UCLA scoring system was considered (Fig. 1), a significant moderate effect was detected in favor of total shoulder arthroplasty, with a difference of 1.4 points (95% confidence interval, 0.5 to 2.4) (p = 0.003). For this domain, the results were very homogeneous (p = 0.84, I² = 0%).

When the analysis was repeated with use of data on individual patients (analysis of covariance), the results were similar. A significant difference with respect to function at two years (p < 0.001) was detected in favor of the total shoulder arthroplasty group (8.1 ± 0.3; 95% confidence interval, 7.6 to 8.7) compared with the hemiarthroplasty group (6.6 ± 0.3; 95% confidence interval, 5.9 to 7.2). Thus, on the average, patients who underwent a total shoulder arthroplasty had only a slight restriction in activities and were able to do work above shoulder level at two years, whereas patients who underwent hemiarthroplasty were able to do most activities of daily living but were not able to do work above shoulder level. There was no evidence of heterogeneity between studies for this domain.

There was a significant difference (p = 0.04) in the mean change in the pain score of 2.0 points (95% confidence interval, 0.1 to 3.9) in favor of the total shoulder arthroplasty group, which represents a moderate effect. Both the Cochran chi-square test of heterogeneity (p = 0.006) and the I² value (80.2%) indicate a high degree of heterogeneity in the results among the studies (Fig. 1) and suggest that an interpretation of the results of this particular domain is difficult and could lead to spurious conclusions.

When this analysis was repeated with use of individual patient data, the results were similar. There was a significant difference (p < 0.0001) in the pain score in favor of the total shoulder arthroplasty group (8.6 ± 0.3; 95% confidence interval, 8.0 to 9.1) compared with the hemiarthroplasty group (6.5 ± 0.3; 95% confidence interval, 5.9 to 7.1), but there was also a significant interaction between which study the data came from and the mean change in the preoperative and postoperative scores (p = 0.004). A significant interaction of this sort reflects the heterogeneity among the study results and indicates that, for each study, the magnitude of the mean change between the preoperative and postoperative scores is inconsistent across studies, making interpretation of an overall pooled estimate for the ef-

---

**Fig. 1**

Comparison of the UCLA shoulder score at two years postoperatively. TSA = total shoulder arthroplasty, HA = hemiarthroplasty, SD = standard deviation, WMD = weighted mean difference, random = random effects model, 95% CI = 95% confidence interval, and df = degrees of freedom.
fect on pain difficult. Specifically, the mean changes in the pain score between the groups in the studies by Lo et al., (0.1; 95% confidence interval, –1.5 to 1.7), Gartsman et al. (2.1; 95% confidence interval, 0.9 to 3.4), and Sandow et al. (3.8; 95% confidence interval, 2.2 to 5.5) were quite different. Lo et al. found essentially no difference between the groups with respect to the mean change in the preoperative and postoperative pain scores, and the other two studies found a large or very large benefit in favor of total shoulder arthroplasty.

The other domains of mobility and strength did not demonstrate a significant difference between the groups with respect to the mean change in the preoperative and postoperative scores (Fig. 1) nor were significant differences detected when individual patient data were analyzed. Patients were able to gain, on the average, 30° in forward elevation and to progress from fair to good strength. There was no evidence of heterogeneity among the studies for these domains.

Although the overall aggregate of the domains is usually considered to represent an overall UCLA shoulder score, we believe that this is potentially misleading because of the heterogeneity in the pain domain among the studies. Nonetheless, we provided this analysis since overall scores are reported throughout the literature. Thus, there was a significant difference (p = 0.008) in the mean change in the overall UCLA shoulder score of 3.9 points (95% confidence interval, 1.0 to 6.8), which represents a moderate effect size, in favor of the total shoulder arthroplasty group. The test of heterogeneity was borderline in terms of significance (p = 0.12), and the I² value (52.9%) indicates moderate heterogeneity among studies. The heterogeneity in pain domains explains this moderate level of heterogeneity in the magnitude of effect.

When the analysis was repeated with use of individual patient data (analysis of covariance), the results were similar. A significant difference (4.2; 95% confidence interval, 2.4 to 5.9) was detected in the overall UCLA shoulder score at two years postoperatively in favor of the total shoulder arthroplasty group (p < 0.0001), and there was a significant effect with regard to which study the data came from (p = 0.034), which represented the heterogeneity among the studies.

The sensitivity analysis indicated that the addition of data on the mean change in the overall Constant score from the study by Jonsson et al.15 did not change the estimate of the treatment effect except to make it more precise (Fig. 2). The estimate of heterogeneity among the trials decreased (46% [p = 0.16] to 19% [p = 0.30]) when data from the study by Jonsson et al.15 were added.

Range of Motion

No significant difference between the groups at two years postoperatively was detected with respect to the change in forward elevation (p = 0.316) or the change in external rotation (p = 0.270) when individual patient data were used. On the average, forward elevation increased by 43° in the patients who underwent total shoulder arthroplasty compared with 31° in the patients who underwent hemiarthroplasty. The average increase in external rotation was 28° for the patients who had a total shoulder arthroplasty compared with 24° for those who had a hemiarthroplasty. Considerable variability in these measurements as well as a reduced number of patients for whom data were available contributed to the possibility of making a Type-II error (declaring no significant difference when one truly exists) and may explain why a significant difference (p = 0.008) in the change in forward elevation (13°; 95% confidence interval, 0.5° to 26°) is found (Fig. 3) when data from the study by Jonsson et al.15 are included. These data were included in the analysis since the estimated effect from this trial was not different from the other trials as evidenced by the unchanged I² value of 0% and the fact that the test for heterogeneity in both this analysis and the analysis for external rotation was not significant.

Discussion

We conducted a meta-analysis of four small randomized clinical trials to provide a more precise overall estimate of the effect of total shoulder arthroplasty compared with
hemiarthroplasty of the shoulder on pain and function at two years postoperatively. Our results demonstrate better function at two years in patients managed with total shoulder arthroplasty. Uncertainty remains with respect to the estimates of the relative effect of the two procedures, since no study included in this analysis is without methodological weaknesses.

Since we were uncertain of whether to include the data from the study by Jonsson et al. because of the high potential for biased results in favor of the total shoulder arthroplasty, we conducted a sensitivity analysis to determine the robustness of the results when data from this trial were added. At this early follow-up, we found no evidence of heterogeneous results among the studies and thus included this study in the pooled analysis. However, since glenoid wear has been shown to progress despite hemiarthroplasty, pooled analyses for long-term comparisons should again test for evidence of heterogeneity since the systematic effects of this efficacy analysis may become apparent over time.

An additional methodological limitation is the use of outcome measures that have not been sufficiently validated. Both the UCLA and Constant shoulder-rating tools were published before the advent of modern measurement methodology. Despite their popularity and widespread use, there is little published information documenting the reliability, validity, and responsiveness of either instrument. However, the results of these trials suggest that the instruments are able to demonstrate moderate differences in the change in function between two treatments with relatively small sample sizes. These results, therefore, provide compelling evidence of the responsiveness of the instruments. Longer-term follow-up studies of the current trial participants and future studies should use measures that have been shown to be both valid and reliable, such as the WOOS index, the ASES standardized tool, and the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire.

Our results reveal a large degree of heterogeneity among the studies with regard to the relative impact of the interventions on pain. The hemiarthroplasty group did far better with respect to pain in the trial described by Lo et al. than in the other studies, and this resulted in negligible differences between treatments with respect to pain. This degree of heterogeneity makes pooling to generate a single estimate of effect questionable. One possible reason for the difference is that patients participating in the trial conducted by Lo et al. had greater function before surgery than did the patients participating in the other two trials. When the average age and baseline functional scores in the study by Lo et al. are compared with the averages in the other two trials, there is no evidence to support this hypothesis.

Clinicians should view the trials conducted to date as preliminary observations. Wise decisions with regard to the choice for hemiarthroplasty or total shoulder arthroplasty will require trials with longer-term follow-up with documentation of outcomes that include glenoid erosion following hemiarthroplasty and its effect on important outcomes for patients. In the four randomized trials included in this meta-analysis, ten patients (20%) crossed over from hemiarthroplasty to total shoulder arthroplasty. The predictors of failure of hemiarthroplasty have not yet been adequately investigated. Longer-term review of the patients included in these trials is important to provide this potential prognostic data.

Conversely, in several studies on total shoulder arthroplasty, progressive radiolucent lines have been observed on plain radiographs, which is evidence of loosening of the glenoid component. At this time, the relationship between glenoid loosening and pain, function, range of motion, or quality of life has not been clearly established. If glenoid loosening occurs frequently following total shoulder arthroplasty and if the functional consequences of loosening are sufficiently serious, hemiarthroplasty may ultimately prove to be the superior procedure because it preserves bone stock, allowing conversion to a total shoulder arthroplasty in the case of failure. On the other hand, a failed total shoulder replacement...
is thought to have a much poorer prognosis as there are no good revision prostheses. In the study by Sandow et al., however, shoulders with a failed hemiarthroplasty were associated with glenoid erosion that compromised the ability to convert to total shoulder arthroplasty, and this finding may contradict the argument for predictable revision of a failed hemiarthroplasty to total shoulder arthroplasty. Similarly, Carroll et al., in a series of sixteen consecutive patients who underwent revision total shoulder arthroplasty for a failed hemiarthroplasty, reported that less than half (47%) of the patients had a satisfactory outcome at a mean of 5.5 years following revision. Furthermore, Antuna et al., in a series of forty-eight shoulders that required revision surgery following total shoulder arthroplasty, reported that 86% of the patients who underwent implantation of a new glenoid component had satisfactory pain relief at a mean of 4.9 years following revision.

It is important to assess the underlying costs associated with these complications in the long term. We know that total shoulder arthroplasty is more expensive at the time of implantation because the prosthesis is more expensive and a longer operating time is required. Total costs at five years, however, may favor total shoulder arthroplasty when the costs of revision surgery and rehabilitation are considered.

In conclusion, the results of this study indicate that, at a short-term follow-up of two years, total shoulder arthroplasty provides more consistent improvement in function than hemiarthroplasty for patients with primary osteoarthritis of the shoulder. Because of the potential for continued degeneration of the glenoid in patients managed with a hemiarthroplasty and for glenoid loosening in patients managed with total shoulder arthroplasty, longer-term (five to ten-year) results are necessary to determine whether these findings remain consistent over time.

Appendix

A table depicting the UCLA functional shoulder assessment tool is available with the electronic version 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).

NOTE: The authors thank Margaret McGee for her valuable review of the manuscript and Ulf Lillkrona for providing further details for one of the trials.

Dianne Bryant, MSc, PhD
Gordon Guyatt, MD, MSc
Department of Clinical Epidemiology and Biostatistics, Hamilton Health Sciences Center, McMaster University, 1200 Main Street West, Hamilton, ON L8N 3Z5, Canada. E-mail address for D. Bryant: jackowdm@mcmaster.ca. E-mail address for G. Guyatt: guyatt@mcmaster.ca

Robert Litchfield, MD, FRACS
Fowler Kennedy Sport Medicine Clinic, 3M Center, University of Western Ontario, London, ON N6E 3K7, Canada. E-mail address: rlitchfi@uwo.ca

Michael Sandow, BMBS, FRACS
Hand and Upper Limb Unit, Department of Orthopaedics and Trauma, Royal Adelaide Hospital and University of Adelaide, Adelaide SA 5000, Australia. E-mail address: michael.sandow@adelaide.edu.au

Gary M. Gartsman, MD
Fondren Orthopedic Group, Texas Orthopedic Hospital, 7401 South Main Street, Houston, TX 77030. E-mail address: gmgs@fondren.com

Alexandra Kirkley, MD, MSc, FRCS
Deceased

The authors did not receive grants or outside funding in support of their research or preparation of this manuscript. They did not receive payments or other benefits or a commitment or agreement to provide such benefits from a commercial entity. No commercial entity paid or directed, or agreed to pay or direct, any benefits to any research fund, foundation, educational institution, or other charitable or nonprofit organization with which the authors are affiliated or associated.

doi:10.2106/JBJS.D.02854

References


20. Ware JE, Kosinski M, Keller SD. SF-12: how to score the SF-12 Physical and Mental Health Summary Scales. 2nd ed. Boston: The Health Institute, New England Medical Center; 1995.


