ASSESSING THE RESULTS OF HIP REPLACEMENT

A COMPARISON OF FIVE DIFFERENT RATING SYSTEMS

JOHN J. CALLAGHAN, STANLEY H. DYSART, CARLTON F. SAVORY, WILLIAM J. HOPKINSON

From the Walter Reed Army Medical Center, Washington

One hundred hips in patients who had had primary uncemented replacements were followed up for one or two years, and assessed by five different methods. All produced different results. The Hospital for Special Surgery rating produced the most optimistic assessment and the Merle d'Aubigné rating the most pessimistic. The functional class of the patients, as defined by Charnley in 1979, significantly affected the ratings, and these should clearly be included in all rating systems. Moreover, if systems are to be compared, they should all use descriptive words, such as limp or pain, in precisely the same way.

A number of hip rating systems, designed to assess function after hip replacement, have been developed (Merle d'Aubigné and Postel 1954; Larson 1963; Harris 1969; Wilson et al 1972). A more recent method includes radiographic measurements (Kavanagh and Fitzgerald 1985). Unfortunately no uniform method has emerged, though this would clearly facilitate comparisons (Galante 1985; Chao et al 1988; Johnston 1988).

The purpose of the present study was to evaluate five frequently used rating systems to assess the one- and two-year results of patients who had undergone primary uncemented porous-coated total hip arthroplasty. Although Andersson (1972) compared different hip assessments after Austin Moore hip hemiarthroplasty, no recent comparisons after hip replacement have been made.

PATIENTS

The patient population consisted of 75 men and 25 women. The age range was from 22 to 81 years with a mean of 60. All patients had undergone a primary uncemented PCA total hip arthroplasty. The procedures were all performed in the same manner by one of two surgeons (or both), except that the first 60 procedures were performed through the direct lateral approach and the last 40 through the posterolateral approach.

RESULTS

The results of all hips for each rating as well as the patients' own assessments are illustrated in Figure 1. The Hospital for Special Surgery rating was the most optimistic and the Merle d'Aubigné rating the most pessimistic. Paired comparisons as analysed by Wilcoxon signed rank test is illustrated in Figure 2. Those with no statistically significant differences are underscored; all other assessments were significantly different with p < 0.05.

The patients' own assessments of the results were 84% excellent, 12% good, 4% fair, and 0% poor. Table I shows that the patient perception of the result did not necessarily agree with the hip score.
Any two means underscored by the same line are not statistically different at the p<0.05 level.

Table 1. Comparison of the patient’s own assessment with the ratings of the five systems (percentage). No patient considered the result to be poor.

<table>
<thead>
<tr>
<th>Patient's own assessments</th>
<th>Result by</th>
<th>Percentage in each category</th>
<th>Merle d'Aubigné</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>hip scoring</td>
<td>HSS</td>
<td>Mayo</td>
</tr>
<tr>
<td>Excellent (n=84)</td>
<td>Excellent</td>
<td>95.3</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>4.7</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>2.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Good (n=12)</td>
<td>Excellent</td>
<td>83.3</td>
<td>58.3</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>8.3</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>8.4</td>
<td>8.4</td>
</tr>
<tr>
<td>Fair (n=4)</td>
<td>Excellent</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

In comparing the results of hips performed through the lateral approach with those performed through the posterolateral approach, we found no statistically significant differences in any of the rating scales (p = 0.2 as measured by the Mann Whitney test between the two groups).

For all rating assessments, the proportion of hips judged to be excellent were significantly less (p = 0.02) for Charnley class C patients (Fig. 3). Of our 100 patients, 42 were Charnley class A, 41 class B, and 17 class C.

Thirty-five of the patients had no complaints at all, 29 complained of a limp, 18 of pain, 12 of limited function, and six of restricted movement. Of patients who were operated upon through the lateral approach 27% had no complaints, but 40% complained of limp; of those in whom a posterolateral approach was used, 48% had no complaints and only 13% complained of limp.

DISCUSSION

Although rating systems have been widely used in assessing the results of hip replacement, their usefulness has been questioned because of their lack of uniformity and objectivity. Although a functional analysis such as that reported by Chao et al (1988) may eliminate some bias, it is probably not practical in most clinical settings.

In our patients with uncemented porous-coated total hip arthroplasties we found no uniformity of results between ratings, nor any uniformity between the ratings and the patients’ impressions. When we compared lateral with posterolateral approaches, it became apparent that global hip scores tended to obscure important findings in individual factors such as limp. In addition, the patients’ Charnley functional class affected all rating scales. We recommend that, in reporting the results of uncemented hip procedures, emphasis should be placed on the functional classes and on important individual parameters, particularly pain and limp, rather than on global hip ratings.

No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

REFERENCES


