Does the new generation of high-flex knee prostheses improve the post-operative range of movement?

A META-ANALYSIS

A new generation of knee prostheses has been introduced with the intention of improving post-operative knee flexion. In order to evaluate whether this goal has been achieved we performed a systematic review and meta-analysis. Systematic literature searches were conducted on MEDLINE and EMBASE from their inception to December 2007, and proceedings of scientific meetings were also searched. Only randomised, clinical trials were included in the meta-analysis. The mean difference in the maximum post-operative flexion between the ‘high-flex’ and conventional types of prosthesis was defined as the primary outcome measure. A total of five relevant articles was identified.

Analysis of these trials suggested that no clinically relevant or statistically significant improvement was obtained in flexion with the ‘high-flex’ prostheses. The weighted mean difference was 2.1° (95% confidence interval -0.2 to +4.3; p = 0.07).

Pain is the most debilitating aspect of osteoarthritis of the knee. Limitation of movement is variable at presentation. Total knee replacement (TKR) is performed primarily to relieve pain, but the post-operative range of movement is an important aspect of outcome. Although many factors influence this, one of the most important predictors is the range present pre-operatively.1,2

Recently, alterations of prosthetic design have been proposed to improve the range of movement. The native knee combines rollback and translation to facilitate full flexion.3 Manufacturers have suggested that conventional designs have limited movement because of impingement of the articular surface of the tibia against the femoral condyle. Newer prostheses have been introduced with increased offset of the posterior femoral condyle intended to allow more flexion. Stretching of the extensor mechanism is another potential limiting factor impeding flexion. In order to address this some designs have incorporated a deep anterior patellar cut-out from the polyethylene insert in order to reduce tension on the extensor mechanism. Another modification has been to increase the posterior slope of the tibial polyethylene insert.

We undertook a meta-analysis to determine whether the new generation of ‘high-flex’ knee prostheses have provided increased movement.

Materials and Methods

Eligibility criteria. We identified randomised, clinical trials involving: a) patients undergoing primary TKR; b) the procedure was a ‘high-flex’ TKR; c) a comparison with conventional TKR was included, and d) the outcome was assessed on the basis of the post-operative range of movement or maximum flexion.

Data sources and search strategy. Initially, an English language literature search was carried out to ensure that no other systematic review of this topic had been conducted. We searched the Cochrane Database of Systematic Reviews,4 the Database of Abstracts of Reviews of Effectiveness5 and the Health Technology Assessment Database6 using the Medical subject heading (MeSH) heading ‘knee’. In addition two other health technology websites, the International Network of Agencies for Health Technology Assessment7 and the Canadian Agency for Drugs and Technologies in Health8 were searched.

Search protocol. The search used was OVID. Searches were carried out using a combination of a MeSH and free text terms to maximise the sensitivity of the search. Although ‘high-flex’ knee prostheses have been available for about five years, our search strategies were not limited by the year of publication. The MEDLINE and EMBASE databases were searched from their inception to December 2007, and the search strategy was conducted using the terms ‘knee prosthesis’, ‘range of movement’, ‘knee
arthroplasty’, ‘total knee replacement’ and high-flex. Additionally, the PubMed interface was used to search MEDLINE using the free text term high-flex. This process was combined with the ‘related articles’ query which had the highest yield of relevant articles compared with all other search strategies.

In order to increase our potential yield, the abstracts from the 2007 American Academy of Orthopaedic Surgeons’ Annual Meeting was hand-searched for relevant articles. In addition, staff at the orthopaedic manufacturer Zimmer (Warsaw, Indiana), were contacted to enquire whether they were aware of any studies evaluating their high-flex knee prostheses. The American and British volumes of the Journal of Bone and Joint Surgery, as well as the Journal of Arthroplasty, were hand-searched between January 2008 and February 2009 for relevant articles.

**Data selection.** All the articles were reviewed by one orthopaedic surgeon (RM) and initially evaluated based on their title and abstract. Irrelevant studies were eliminated. Duplicate articles were identified by reviewing the author, title and date of publication and subsequently eliminated. Potentially duplicate articles were scrutinised. After the preliminary evaluation, the full text of the chosen articles was reviewed and based on the outlined inclusion and exclusion criteria (Tables I and II), the final list of articles was determined.

**Quality assessment of the study.** The checklist of Downs and Black was adapted to assess the quality of the studies. This checklist consisted of 27 questions which evaluated the quality of reporting, the external validity, internal validity and power. Question 27 was replaced by the following: “If no significant difference is found between the two groups, did they perform a power analysis”? (1 - Yes, 0 - No, 0 - Not mentioned)”?

The highest possible score was 28.

In general, scores of assessment of quality may be used in several different ways. The score may be used to provide a cut-off value for exclusion of studies from the meta-analysis or it may be used for stratified analysis. It may also be used to calculate the weighted-study contribution to the overall effect. Component analysis entails evaluating the contribution of various aspects of the study to the overall effect size. For the purposes of our study, the study quality score was used only as an objective description of the studies which constituted the meta-analysis (Table III). Additionally,
evidence of industrial sponsorship of research was noted since this may have influenced the manner in which the outcome was reported.\textsuperscript{10,11} We also identified whether the measurement of the post-operative range of movement was conducted in a blinded fashion for the control and experimental groups.

**Data extraction.** The variables of interest were broadly categorised as the study design, surgical variables, patient factors and the results of the study. The information was extracted systematically and independently by one orthopaedic surgeon (RM).

**Synthesis of data**

**Quantitative data synthesis.** All the statistical evaluations were conducted using the metan module in STATA 10 (Stata Corp, College Station, Texas). For each study, the effect size or the primary outcome was calculated as the mean difference in maximum flexion between the high-flex and the conventional TKR groups. However, some of the studies reported only the range of movement and, in these cases, it was assumed that there was no significant post-operative flexion contracture and therefore the total arc of movement was interpreted as the maximum flexion. If the SD of the maximum flexion was not reported, it was indirectly calculated from the p-value using the test statistic, Z. The Z statistic is used to compare two means as follows:

\[
Z = \frac{X_1 - X_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

If a p-value is provided and assuming a two-sided test, the SD can be calculated as follows

\[
SD = \frac{X_1 - X_2}{Z \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}
\]

where \( Z = \theta^{-1} (P/2) \) and \( \theta^{-1} \) is the inverse normal distribution.

When the exact p value is provided, this is a very accurate estimate. However, when a range is provided (e.g. \( p < 0.01 \)), using \( Z = \theta^{-1} (P/2) \) (e.g. \( Z = 2.58 \)) will provide a conservative estimate of the variance.

Two of the studies did not provide the SD of the maximum flexion. The study of Weeden and Schmidt\textsuperscript{12} reported that the difference between the control and experimental group’s final range of movement was significant at \( p < 0.05 \). In order to calculate the SD for our study, a p-value of 0.05 was used to derive a z score of 1.96. Another study for which the SD had to be calculated was that of Kim, Sohn and Kim,\textsuperscript{13} in which an exact p-value of 0.41 was quoted from which a z score of 0.824 was derived. We acknowledge that using a range rather than an exact p-value overestimated the SD and yielded a wider confidence interval (CI).

The mean differences in maximum flexion between the conventional and high-flex groups (i.e. effect size) from each study were pooled and analysed in both a ‘fixed-effects’ model as well as a ‘random-effects’ model.\textsuperscript{14} The former is based on the assumption that the variation between trial results is due to differences in precision, i.e., if every study was infinitely large, it would yield an identical result. Because the range of movement is a continuous variable, the data were pooled using a weight which was merely the inverse of the pooled standard error of the mean (SEM).

The random-effects model makes the assumption that there are true differences between studies, i.e., that the individual studies are estimating different treatment effects. It further assumes that the distribution of these effects is normal.

**Descriptive data synthesis.** Tables IV and V outline pertinent information related to the design of the study as well as the relevant results.

**Evaluation of heterogeneity.** Statistical heterogeneity was assessed graphically using a forest plot and the I\textsuperscript{2} statistic.\textsuperscript{16} If the I\textsuperscript{2} statistic is greater than 50% this indicates a substantial level of heterogeneity. Statistical heterogeneity is different from study heterogeneity. The latter has been partially addressed using the standardised checklist for assessment of the quality of the study. Publication bias was evaluated using a funnel plot.\textsuperscript{17}

**Results**

**Study characteristics.** The combined yield of all the search strategies was 786 articles. After reviewing their title and
abstract, 749 were eliminated leaving 37 articles. Having removed 18 duplicated studies between the search strategies, 19 remaining articles were further evaluated. Of these, one was not used because patients who had more than 120° of flexion before surgery were included. It was felt that this significantly compromised the external validity of the study. Another study was removed because the outcome measure was the intra-operative range of movement rather than the post-operative range. One study compared very different prostheses (mobile-bearing cruciate-retaining vs posterior-stabilised high-flex) and was removed. Five other studies were removed since they did not compare the clinical outcome of a high-flex prosthesis with that of a comparable conventional model and six because they were not randomised controlled trials. This left five studies in the meta-analysis (Table III). All the trials except one were published in English. Of these five studies, only three involved measurement of the range of movement, in a blinded fashion.

Table IV. Qualitative description of the studies

<table>
<thead>
<tr>
<th>Authors</th>
<th>Country</th>
<th>Study design</th>
<th>Sample size (TKRs* in each group)</th>
<th>Mean follow-up (yrs)</th>
<th>Mean age (yrs)</th>
<th>Conventional High-flex</th>
<th>Prostheses†</th>
<th>Surgical approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al¹³</td>
<td>South Korea</td>
<td>RCT⁷</td>
<td>50 patients undergoing bilateral TKR</td>
<td>2.1</td>
<td>68 (53 to 81)</td>
<td>68 (53 to 81)</td>
<td>Zimmer LPS vs LPS Subvastus flex⁸</td>
<td></td>
</tr>
<tr>
<td>McCalden et al²⁹</td>
<td>Canada</td>
<td>RCT</td>
<td>50</td>
<td>2.7</td>
<td>72</td>
<td>70</td>
<td>Genesis II PS insert vs Genesis II HF flex⁸</td>
<td></td>
</tr>
<tr>
<td>Nutton et al²⁷</td>
<td>Scotland</td>
<td>RCT</td>
<td>28</td>
<td>1.0</td>
<td>68 (52 to 85)</td>
<td>71 (38 to 75)</td>
<td>Zimmer LPS vs LPS Medial parapatellar flex⁹</td>
<td></td>
</tr>
<tr>
<td>Weeden and Schmidt¹²</td>
<td>United States</td>
<td>RCT</td>
<td>25</td>
<td>1.0</td>
<td>62.3 (51 to 78)</td>
<td>62.8 (52 to 82)</td>
<td>Zimmer LPS vs LPS Subvastus, Medivastus flex⁹</td>
<td></td>
</tr>
<tr>
<td>Wohlrab et al²⁸</td>
<td>Germany</td>
<td>RCT</td>
<td>30</td>
<td>34.6</td>
<td>66 (SD 9)</td>
<td>67 (SD 5)</td>
<td>Zimmer LPS vs LPS Midvastus flex⁹</td>
<td></td>
</tr>
</tbody>
</table>

* TKRs, total knee replacements  † RCT, randomised, controlled trial  ‡ LPS, LPS flex, Zimmer, Warsaw, Indiana. Genesis II, Smith & Nephew, Memphis Tennessee

Table V. Quantitative summary of the studies. Range of movement (ROM, °) and maximum flexion (°) are reported as the mean (SD or calculated SD or range)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study design</th>
<th>Sample size (TKRs* in each group)</th>
<th>Mean pre-operative ROM</th>
<th>Mean post-operative ROM</th>
<th>Maximum pre-operative ROM</th>
<th>Maximum post-operative ROM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al¹³</td>
<td>RCT†</td>
<td>50 patients undergoing bilateral TKR</td>
<td>126 (55 to 150)</td>
<td>127 (55 to 150)</td>
<td>136 (calculated SD 18.1)</td>
<td>139 (calculated SD 18.1)</td>
</tr>
<tr>
<td>McCalden et al²⁹</td>
<td>RCT</td>
<td>50</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Nutton et al²⁷</td>
<td>RCT</td>
<td>28</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Weeden and Schmidt¹²</td>
<td>RCT</td>
<td>25</td>
<td>119</td>
<td>120.2</td>
<td>119.4</td>
<td>132.4</td>
</tr>
<tr>
<td>Wohlrab et al²⁸</td>
<td>RCT</td>
<td>30</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

* TKRs, total knee replacements  † RCT, randomised, controlled trial

Table VI. Effect of high-flex implants to improve range of movement. The outcome is the weighted mean difference (WMD) in maximum flexion, using a fixed-effects analysis model

<table>
<thead>
<tr>
<th>Study</th>
<th>WMD</th>
<th>95% CI</th>
<th>% weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim et al¹³</td>
<td>3.0</td>
<td>-4.1 to +10.1</td>
<td>9.56</td>
</tr>
<tr>
<td>McCalden et al²⁹</td>
<td>1.0</td>
<td>-1.8 to +3.8</td>
<td>63.28</td>
</tr>
<tr>
<td>Nutton et al²⁷</td>
<td>4.0</td>
<td>-4.9 to +12.9</td>
<td>6.13</td>
</tr>
<tr>
<td>Weeden and Schmidt¹²</td>
<td>13.0</td>
<td>-0.03 to +26.0</td>
<td>2.67</td>
</tr>
<tr>
<td>Wohlrab et al²⁸</td>
<td>3.0</td>
<td>-2.2 to +8.2</td>
<td>18.17</td>
</tr>
</tbody>
</table>

* CI, confidence interval

**Clinical diversity.** Although the concept of a high-flex prosthesis has common features, each company has their own distinct design. In four trials a Zimmer product was used, and a Smith and Nephew (Memphis, Tennessee) product (Table IV) in the remaining study. Two studies had industrial affiliation declared, two did not and one did not comment (Table III). Unfortunately, with so few studies, a meaningful sensitivity analysis to determine the importance of the industrial affiliation could not be performed. One of the studies was conducted in Canada, one in South Korea, one in Germany, one in Scotland and one in the United States.
The weighted mean difference between the two designs was calculated as only 2.1° (95% CI -0.2 to +4.3, Z test, p = 0.07). Not only is this not statistically significant, but more importantly it is not clinically significant. Additional flexion of 2° or even the upper limit of 4.3°, has no functional advantage to the patient.

The fixed-effect and random-effects models gave virtually identical results. Usually, the latter is more conservative than the fixed-effect model. The pooled models gave virtually identical results since there was no evidence of substantial statistical heterogeneity among the studies. Furthermore, the pooled-effect size calculated with a small number of studies in the meta-analysis.

Publication bias. The issue of publication bias was explored graphically using a funnel plot (not shown). Although there was no indication of publication bias, the results of the funnel plot must be interpreted with caution in the light of the small number of studies in the meta-analysis.

**Discussion**

Our meta-analysis suggested that the new generation of high-flex knee prostheses do not increase the post-operative maximum knee flexion compared with conventional implants. The main limitation of our study was the quality of the available studies. Only three of the five studies blinded those who measured the post-operative range of movement. There are not enough studies available to conduct relevant subgroup and sensitivity analysis of the clinical and methodological diversity of the studies.

The weighted mean difference between the two designs was calculated as only 2.1° (95% CI -0.2 to +4.3, p = 0.07). Not only is this not statistically significant, but more importantly it is not clinically significant. Additional flexion of 2° or even the upper limit of 4.3°, has no functional advantage to the patient.

**References**