Implant design influences patient outcome after total knee arthroplasty

A PROSPECTIVE DOUBLE-BLIND RANDOMISED CONTROLLED TRIAL


From University of Edinburgh, Edinburgh, United Kingdom

Total knee arthroplasty (TKA) is one of the most successful of all surgical procedures and is highly effective at reducing pain and enhancing function in those suffering from osteoarthritis. Each year, approximately 85,000 TKAs are carried out in the UK, while in the United States the number is thought to be in excess of 700,000. Although most patients do well, some describe a less successful outcome; moderate rates of dissatisfaction are consistently reported in around 20% of patients.

Over the last 30 years, there have been a considerable number of developments in the design of knee implants in an attempt to improve patient outcome. There is, however, little consensus whether any of these have had any effect. In 2012, Carr et al noted that the number of available implants had increased substantially but with little or no evidence of greater effectiveness. It is therefore essential to evaluate the comparative effectiveness of implants with prospective clinical trials.

Cite this article: Bone Joint J 2015;97-B:64–70.

Survivorship analysis is the standard method of evaluating the longevity of an implant but assesses failure rather than functional outcome. Revision surgery is rare within ten years of implantation. Consequently, trials which compare the functional outcome of modern designs with established implants in the early post-operative phase, as recommended by the British Orthopaedic Association, are needed to compliment survival studies.

The Triathlon TKA (Stryker, Mahwah, New Jersey) was introduced into the United Kingdom eight years ago to improve on the preceding devices offered by the manufacturer. It is now the third most commonly used knee implant reported in the National Joint Registry (NJR). However, no comparative trial exists to show the Triathlon TKA is better than the implant it superceded.

The aim of this study was to determine whether differences in outcome could be attributed to the implant patients received, and to test the null hypothesis that no difference in
outcome would be detectable between two different designs of prosthesis.

Patients and Methods
We carried out a prospective, double-blind randomised control trial to assess the influence of the design of a total knee implant on the functional outcome of a patient. Inclusion criteria were a diagnosis of osteoarthritis; a planned primary TKA with standard implants (i.e. without augments) and the capacity of the patient to give informed consent. Exclusion criteria were the presence of co-morbidities that would restrict post-operative recovery or subsequent physical performance (such as significant cardiovascular or neurological disease). Previous joint replacement or the radiographic presence of osteoarthritic changes in other major joints was accepted, provided that these were not reported as symptomatic by the patient and did not influence their physical function. This was assessed by review of the medical notes and discussion with the patient at time of consent and randomisation.

Ethical approval was granted by the Lothian Research Ethics Committee 03 (ref: 06/S1103/50) and the trial was logged with the International Standard Randomized Controlled Trial Number Register (ISRCTN85418379).

Recruitment took place at the Royal Infirmary of Edinburgh between February 2008 and August 2010. Patients were identified from the planned operating lists of six consultant surgeons with experience of both implants. Suitable patients were approached at the time of pre-operative assessment, and recruited with informed consent. Implant allocation was by a bespoke internet-based randomisation programme employing random block allocation, coordinated by the lead surgeon (PG). Both patient and researcher were blinded to implant allocation, and remained so throughout the trial period.

Implants and operative technique. The implants used were the Kinemax (Stryker) (group 1, old design) and Triathlon (group 2, new design) TKAs. The Kinemax was selected because it was the standard implant used in our unit at the onset of the trial, and the then recently-released Triathlon prosthesis because it represented a significant evolution in design. Particular changes in the design of the femoral component which were thought to be beneficial to patient outcome included a sided patellofemoral groove, flared posterior condyles and a single radius of curvature.

Each of the six trial surgeons used both implants. These were inserted using the same technique and standard instrumentation with intra-medullary referencing for the femur and extra-medullary alignment of the tibia. No patient recruited to this study underwent resurfacing of the patellofemoral joint. Cruciate retaining, fixed-bearing implants were used in each case and all components were cemented. The rotational alignment of the femoral implant was set with reference to the transepicondylar axis and Whiteside’s line. All other aspects of patient care were identical for each group.

Baseline descriptive data about the patients who were recruited are detailed in Table I. There were no significant differences in mean age or gender between the two groups. A total of 237 patients were assessed for eligibility: 212 consented to enrolment and were randomised pre-operatively (Fig. 1).

Outcome assessments. Patients were assessed pre-operatively, then at outpatient clinical review at six weeks, six months, one and three years’ post-operatively in a clinical testing facility at the Royal Infirmary of Edinburgh. A comprehensive protocol, consisting of patient-reported questionnaires and objective functional assessments, was used to evaluate patient outcome. All tests were carried out in the same way by a single assessor (DFH) who was blinded to the type of prosthesis implanted.

The primary outcome was the Oxford Knee Score (OKS), a validated and frequently used 12-item response questionnaire designed to assess the patient’s perceived pain and functional ability. Scores range from 0 (severe symptoms and dysfunction) to 48 (a well-functioning knee joint).

Global knee pain severity was assessed using an 11-point (0 to 10) numerical rating scale (NRS), where nought...
represents no pain and ten the worst possible pain. The excellent validity\textsuperscript{14} and sensitivity\textsuperscript{15} of this scale is well documented. Multiple measurements of pain status (as opposed to a single value of ‘current pain’) have been reported to provide a more realistic and meaningful measurement of intensity of pain\textsuperscript{16}. Separate assessments were made of ‘perceived worst pain’ and ‘perceived mean daily pain’ as has been recommended for clinical trials involving patients with osteoarthritis\textsuperscript{17}.

Active measurements of flexion and extension were made using universal goniometry, which has previously been shown to achieve a high level of clinical accuracy\textsuperscript{18}, specifically in patients after TKA\textsuperscript{19}.

The power of the patient’s lower limb was determined with a Leg Extensor Power Rig (LEP, Nottingham, United Kingdom), which has been validated for use with this population group\textsuperscript{20}. The LEP consists of a seat and footplate connected through a lever and chain to a flywheel. Application of force accelerates the flywheel from rest, and output is recorded as the maximal wattage (W) generated. Output is reported as maximal W generated in a single-leg extension, and is expressed as a proportion of the power output of the contralateral lower limb which acts as an internal control. Those who were unable to complete the test were assigned a score of zero as advocated by Lamb and Frost\textsuperscript{20}.

The ability to perform daily functional tasks was assessed using the aggregated locomotor function score (ALF)\textsuperscript{21}. This score is a composite timed measure of observed locomotor function which uses tests of walking, stair ascent/descent, and chair transfers: it has previously been shown to be valid, reliable and responsive. Specifically, patients were asked to walk over a flat eight-metre course, climb and then descend a platform consisting of seven fixed steps, and perform a chair transfer task. Time was recorded using a handheld stopwatch (Zeon, London, United Kingdom).

Patient satisfaction with outcome was assessed at a single point (year three assessment) on a four-point Likert scale: the response options were very satisfied; satisfied; unsure or dissatisfied.

**Statistical analysis.** The trial was powered to detect a difference between groups of three points in the change in OKS between the pre-operative and one-year post-operative assessment. Analysis of variance (ANOVA) with an alpha value of 0.05 and beta value of 0.8, estimated that 200 patients (in total, with a 50:50 split) would be needed, allowing for a drop-out of 10% at one year.

We performed a complete case analysis where data from participants which provided a primary outcome variable were included, using all available data to mitigate any effect of loss to follow-up\textsuperscript{13}. Data were manually checked for normality with histograms. Means with standard deviations (SD) or 95% confidence intervals (CI) as a measure of dispersion were used to describe the data. The primary
outcome of the trial (change in OKS between baseline and one year) was assessed by the repeated measures ANOVA. We then evaluated the patients three years post-operatively to reflect their clinical outcome more completely. Repeated measures ANOVA using generalised linear mixed models (suitable for assessing longitudinal data) were used for this post hoc analysis which incorporated data from all time points. Post-operative satisfaction response is assessed by Pearson’s Chi Square. Significance was accepted at \( p \leq 0.05 \). Data were collated and analysed using SPSS v.19 software (IBM, Armonk, New York).

**Results**

At one year, 183 patients (83 Kinemax TKA and 100 Triathlon TKA) were available for primary outcome analysis (Table I). Of the 29 patients lost to follow-up (drop-out rate 13%), eight did not undergo surgery. A total of six procedures were cancelled or delayed beyond the trial recruitment period and two were transferred to consultants who were not contributing patients to the trial as a part of a waiting list initiative. A further 20 patients were lost to follow-up by the time of the six-week post-operative assessment. One patient died of causes unrelated to the surgery. A total of three patients suffered an early infection requiring a revision, one patient suffered a post-operative flare-up of previously undiagnosed rheumatoid arthritis, one a progressive foot neuropathy and another a peroneal nerve complication, which prevented them undertaking the functional assessment. One patient revealed chronic pain pathology which had not been disclosed at recruitment and 12 refused to participate post-operatively, requesting routine follow-up by clinical teams at their local satellite outpatient’s facility as opposed to the central orthopaedic research facility. At 12 months, one patient could not be contacted and was lost to further follow-up.

At three years, 165 patients (75 Kinemax TKA and 90 Triathlon TKA) were available for post hoc analysis (consecutive dropout rate 22%): a further 18 patients had been lost to follow-up. Of these, 12 patients had died, four had moved out of the area and could not be contacted, one required revision of their implant for persistent pain and one developed co-morbidities which prohibited compliance with assessment. Loss to follow-up is summarised by implant allocation in Table II.

Those patients lost to follow-up displayed similar baseline characteristics to the wider trial cohort: 27 (58%) were female with a mean age of 70.0 (SD 13.2). There were no differences in patient characteristics between the two groups as a result of loss to follow-up. A full description of the trial population by implant allocation at the analysis timeframes is displayed in Table I.

Applying the primary outcome of the trial (the mean change in OKS between baseline and one year-follow-up), we identified the mean change in score was 17.1 points (-14 to 33) in group 1 and 20.1 points (-3 to 39) in group 2, (ANOVA, \( p = 0.05 \)) (Table III, Fig. 2a). Subsequent post hoc analysis of change in OKS over the three-year period (assessed using longitudinal mixed model ANOVA) did not achieve statistical significance (\( p = 0.09 \)), though was possibly underpowered due to the additional data time points.

### Table II. Summary of patients lost during the post-operative (post-op) period

<table>
<thead>
<tr>
<th>Time of drop out</th>
<th>Reason for drop out</th>
<th>Group 1 (n)</th>
<th>Group 2 (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery</td>
<td>Operation delayed/cancelled</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Non-trial surgeon reallocation</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Six weeks post-op</td>
<td>Death</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Early infection</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Patient withdrew (local follow-up)</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Comorbidity preventing assessment</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>One year post-op</td>
<td>Did not respond to correspondence</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>Three years post-op</td>
<td>Death</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Not contactable: moved out of area</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Revised</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Impaired health prohibiting assessment</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table III. Trial primary outcome: change in Oxford Knee Score between baseline and one year

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (n = 83)</th>
<th>Group 2 (n = 100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>20.6 (SD 7.73)</td>
<td>18.7 (SD 7.13)</td>
</tr>
<tr>
<td>One year</td>
<td>36.7 (SD 8.11)</td>
<td>38.8 (SD 7.62)</td>
</tr>
<tr>
<td>Change score</td>
<td>17.1 (SD 9.11)</td>
<td>20.1 (SD 8.84)</td>
</tr>
</tbody>
</table>
Significant between-group differences in range of movement (ANOVA, p = 0.009), lower limb power output (ANOVA, p = 0.026) and worst daily pain experienced (ANOVA, p = 0.003) were seen, with group 2 reporting better results in each case (Figs 2b, 2c and 2d, respectively). Mean scores for these three measures were similar at
baseline and diverged over the post-operative period, with the one and three-year results being notably different between groups. A mean $10^\circ$ (-11 to 23) difference in knee flexion was apparent at three years: the mean flexion was $99.8^\circ$ (SD 9.7) in group 1 and $108.5^\circ$ (SD 9.7) in group 2. The lower limb power output improved to 94% (SD 34) of the contralateral limb in group 1, and to 116% (SD 40) of the contralateral limb in group 2 at year three. It should be noted that this increase beyond the control limb reported in group 2 remains below values reported for healthy aged-matched individuals, reflecting the general reduction in function of patients who have undergone knee replacement.22,23

The mean ‘worst daily pain’ reported at three years was 2.5 points (SD 2.2) in group 1 and 1.2 points (SD 1.7) in group 2 (Fig. 2d) at three years.

The patients report of ‘average daily pain’ experienced had reduced substantially to 0.9 (SD 1.4) in group 1 and 0.5 (SD 1.2) in group 2 by year three (Fig. 2e). Differences across the trial assessment period were not significantly different (ANOVA, $p = 0.57$). The mean timed functional performance at three years was 26.4 seconds (SD 6.0) in group 1 and 25.6 seconds (SD 4.0) in group 2 (Fig. 2f).

Between-group differences across the trial period were not statistically significant at three years (ANOVA, $p = 0.23$).

Satisfaction with outcome at three years was significantly better in group 2, (Pearson’s chi-square, $p = 0.001$) (Table IV).

Table IV. Satisfaction with outcome by group at three years

<table>
<thead>
<tr>
<th>Satisfaction response</th>
<th>Group 1 (n)</th>
<th>Group 2 (n)</th>
<th>Total (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (very satisfied)</td>
<td>19</td>
<td>48</td>
<td>67</td>
</tr>
<tr>
<td>2 (satisfied)</td>
<td>51</td>
<td>41</td>
<td>92</td>
</tr>
<tr>
<td>3 (uncertain)</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>4 (unsatisfied)</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>75</strong></td>
<td><strong>90</strong></td>
<td><strong>165</strong></td>
</tr>
</tbody>
</table>

Crosstabulation table, Pearson’s chi-squared = 16.41, $p = 0.001$

Discussion
The main finding of this study was that, contrary to our null hypothesis, the design of a total knee replacement can influence functional outcome.

Our results show the substantial patient benefits of knee arthroplasty irrespective of implant, and chart the early changes in patients’ return to function in the months and years immediately after surgery. Patient satisfaction with outcome, report of worst daily pain experienced, range of movement and the ability to generate force with the lower limb musculature were all better in the group that received the Triathlon implant. Other generic outcome measures (timed functional tasks and report of average daily pain) were not significantly different between the groups.

As the failure rates of large-head metal-on-metal total hip replacements have shown, it cannot simply be assumed that a new design of implant gives a better outcome.24,26

Criticism has rightly been directed at implants which have been introduced without determining whether they are better than existing prostheses. This study was performed as part of a stepwise introduction of new technology at our unit as recommended by Malchau27 and supported by the BOA.7 The Kinemax implant used as the control in this study is a good example of a condylar implant that has been used successfully for many years:28-30 it was our unit’s standard prosthesis when this trial was developed (2007/8). Small single-centre cohort studies in the United States had suggested good early functional outcomes with the Triathlon implant in terms of range of movement and Knee Society Score.31,32 However, to the best of our knowledge, a prospective randomised controlled trial directly comparing the Triathlon implant to the Kinemax has not previously been reported.

Historically, implant survival is the preferred method of comparing implant designs. The NJR 10th annual report4 highlights the fact that short to medium-term survivorship is excellent after primary TKA regardless of fixation, constraint and type of bearing. The two implants investigated in this project show similar five-year survivorship figures in a tightly-controlled single-centre series,28,33 and in the current NJR report (2013) at nine years follow-up.2 However, detailed functional outcome trials, such as reported here, are needed to see if there are differences in functional outcome with different designs of prosthesis.

We found the change in OKS between the pre-operative and one-year scores to be significantly different between the groups with the difference reaching the accepted level of clinical significance of around three points.13 The additional analysis time points and further loss to follow-up of 18 patients between one and three year reviews is the likely to have substantially reduced the power of our study to detect a difference between groups. Despite the potential limitation of reduced power, we believe that the post hoc longitudinal analysis is the most robust and comprehensive means of assessing a dataset of this nature; investigating the ‘overall’ change in clinical outcome across the five assessment time points as opposed to artificially prising out individual time point comparisons. That pain, range-of-movement and power output variables remain significant despite this, is telling.

Loss to follow-up is a concern in all clinical trials as this has the potential to bias the outcome. Beyond reducing the power to detect differences, the most important consideration regarding participant attrition is whether it unbalances
the trial design and introduces bias by having different proportions of confounding variables in each group. To address this concern we have reported the patient characteristics (by implant group) at baseline and also at the one- and three-year analyses. Reassuringly, there appear to be no differences in the available known confounders that could influence the primary outcome, lending further credibility to the results. The reduction in the number of patients at three years is particularly interesting as, although implant survival is frequently reported, death rates amongst arthroplasty patients over time are not. The fact that 12 patients (5% of those recruited) died in the three years after surgery suggests that this additional loss to follow-up should be acknowledged when powering clinical studies of joint replacement beyond one year. The problems of survivorship analysis in the face of competing risks has been reported previously. The strengths of this study were the RCT design and that a single assessor performed all assessments, blinded to the underlying prognosis, and that other variables thought to impact outcome (such as bearing surface and fixation) were consistent between groups. Although tempting to attribute specific enhancements in outcome to specific design differences between the two implants, no direct causation can be implied by this study as the trial is not designed to evaluate these specific differences in isolation. Nevertheless, better patient function (lower limb power and knee flexion), pain levels and overall satisfaction with outcome were present in the patients treated with the newer Triathlon design.

References


