EMPIRICAL SUPPORT FOR RADIOGRAPHIC REVIEW: A FOLLOW UP STUDY OF TOTAL HIP ARTHROPLASTY

Lindsay K. Smith¹
Fiona Cramp²
Shea Palmer²
Nikki Coghill¹
Robert F. Spencer³

¹ Faculty of Clinical Sciences, University of Bristol, Bristol, UK
² Faculty of Life and Health Sciences, University of the West of England, Bristol, UK
³ Department of Trauma and Orthopaedics, Weston General Hospital, Weston-s-Mare, UK

Work attributed to: Faculty of Life and Health Sciences, University of the West of England, Bristol, UK

Presentation of work: British Hip Society, Torquay, England, 3 March 2011

Membership of EHS: Robert F. Spencer

Corresponding author: L K Smith, Department of Trauma and Orthopaedics, Weston General Hospital, Weston-s-Mare BS23 4TQ
Phone: 01934 647206
Fax: 01934 647219
Email: smith@unforgettable.com

Disclaimers: No sources of support were received in connection with this study. Conflict of interest: None of the authors have any proprietary interest. Abbreviations: Total hip arthroplasty (THA), Patient reported outcome measures (PROMs), Oxford Hip Score (OHS), EuroQol 5 dimension score (EQ-5D)

Short running head: Empirical support for radiographic review in hip arthroplasty surveillance
ABSTRACT

Routine post surgery surveillance of total hip arthroplasty (THA) is widely recommended to identify asymptomatic failure but inclusion of an x-ray adds to service costs. Evidence is needed to support orthopaedic opinion in order to identify what should be included in surveillance. An investigation was conducted to establish whether an x-ray is needed in addition to patient reported outcome measures. One hundred and fifty-four THA had been assessed at three years and were reviewed again at six to nine years (mid-term) when radiographic signs of deterioration commonly appear. Data were explored for associations between radiographic changes and changes in the participants’ Oxford Hip Score, age, EuroQol 5-D score or comorbidities. Hierarchical multiple regression analysis showed that the number of radiographic changes could not be predicted by any of the other variables. This supports the inclusion of an x-ray in THA surveillance and suggests that the state of the THA cannot be determined by the use of patient reported outcome measures alone. This has implications for future arthroplasty surveillance.

Keywords: Total hip arthroplasty, Surveillance, X-ray, Patient reported outcome measures.
**INTRODUCTION**

Total hip arthroplasty (THA) is a well established procedure and national joint registries show that over 90% of prostheses are still functioning successfully at ten years. Despite recent advances in materials and design, these same registries record that aseptic loosening and osteolysis still affect the long term survival of THA and are the major reason for revision surgery. The 'silent' nature of their development can be a threat to the fixation and stability of the prosthetic components (1-4).

Monitoring of all patients with a hip replacement is recommended in order to identify adverse changes and intervene appropriately to improve the chances of a good patient outcome at revision (1,2,5). However, the economics of healthcare delivery remain a significant challenge and increasing numbers of primary THA add to the burden (6-9). Many orthopaedic units now face pressure to discharge arthroplasty patients early in the postoperative period and to leave further care to the general practitioner. In this environment, a surveillance service should be both clinically effective and cost efficient, collecting only data of relevance at suitable review intervals.

The use of patient reported outcome measures (PROMs) to monitor the results of THA is now widely employed and tools such as the Oxford Hip Score (OHS) can provide useful information about large cohorts of patients over an extended period (10-12). The use of sequential scores to assess the state of the individual
THA before failure is less commonly reported but may provide an easily collected (and consequently attractive to service providers) indicator of the condition of the joint. The addition of radiographic review to any surveillance increases the cost and risk to patients and, although it is considered essential by the orthopaedic profession (5,13), service commissioners may require evidence for its inclusion.

This study was designed to answer the question: for patients with total hip arthroplasty undergoing a mid-term review, is an x-ray needed in addition to a specific hip outcome score? A prospective cohort study was designed for mid-term review (six to nine years) of two types of hip arthroplasty (cemented and hybrid). This period was chosen to represent the time at which signs of deterioration often appear (5,13,14). The choice of outcome tools reflects recommendations in the literature (12,15,16) and the information obtained is therefore potentially useful to orthopaedic professionals and service commissioners when planning THA surveillance.

**Patients and Methods**

The participants selected for this clinical study were a consecutive series that received a THA between 2000 and 2003 in a district general hospital and had previously been recruited to an observational study three years after the primary surgery. In common with other UK units at that time, an all cemented prosthesis was used for an older patient (cemented THA) and an uncemented metal cup with polyethylene liner and a cemented femoral stem (hybrid THA) was used in
younger patients. The cemented acetabular component was a Cenator cup (Corin Medical, Cirencester, UK), which is a high density polyethylene, flanged cemented device with option for an extended posterior wall. The uncemented cup was the EPF Plus (Plus Orthopedics AG, Rotkreuz, Switzerland) which is an equatorially expanded, spray coated pure titanium cup with screw options, a polished inside and a press-fit liner of ultra-high molecular weight polyethylene.

The original cohort attended three years after primary surgery for a check x-ray (plain film images) and completion of questionnaires including the OHS. All data were available for the current study and potential participants were identified from these records. Their mortality status was ascertained and those still alive were invited for mid-term review, six to nine years post surgery. Ethical approvals were obtained from the North Somerset and South Bristol NHS Research Ethics Committee and from the University of the West of England Ethics Sub-Committee, School of Health and Social Care.

Each participant completed outcome measures: the OHS, the EuroQol 5-dimension questionnaire (EQ-5D) and the Charlson comorbidity index (17,18). The OHS was selected because it is a patient reported outcome measure which is readily applied, is joint specific and has been shown to be sensitive to change (15). The inclusion of a generic health instrument and a comorbidity index is recommended to reduce the potential confounding effects of other health problems on a joint specific tool (15,16,19). The change in OHS from three years
to mid-term was computed for each participant as recommended for non-randomised studies (12). The five answers for the EQ-5D were each scored from one to three, assigned a value from a national value set and summed to produce a health index between 1.0 (full health) and -1.0 (a state worse than death) (20).

Two digitized x-rays were obtained for each THA, an antero-posterior view (full pelvis centred on the pubic symphysis) and an iliac oblique view (21) and all subsequent measurements were corrected for magnification using the known size of the femoral head. Femoral zones were numbered from one to fourteen (22,23) and acetabular zones from I to VI (24,25). The presence of radiolucencies, cortical hypertrophy and osteolysis were recorded by zone. Osteolysis was defined as a new or expanding radiolucent area in which no trabeculae were visible compared to adjacent bone (26,27). Radiolucencies greater than 2mm in width were noted as osteolysis (28) and the area of an osteolytic lesion was recorded by superimposing a simple morphometric grid designed for this purpose (29). Radiographic assessment of the femoral component included stem inclination, subsidence, calcar rounding and resorption (26,30,31). The inclination of the acetabular cup and any migration of the component were measured with reference to the inter-teardrop line (32,33). The presence of heterotopic ossification was recorded using the Brooker classification (34).
Measurement of linear wear of the acetabular cup or liner was made using the method of Dorr and Wan (35) using digital callipers (plain film images) or the integral software (digitised images). A steady state wear rate (mm per year) was calculated to allow for the initial ‘bedding in’ of components which occurs in the early post-operative stage, usually by two years (36,37). It was calculated as the difference between the linear wear at each radiographic review divided by the number of intervening years.

All data were collected by the primary author and reviewed at a later date by the senior orthopaedic author to reduce the bias introduced by a single set of observations. Interobserver reliability was assessed from a random sample of 20 results for acetabular inclination and wear rate measurements. A variable to represent radiographic changes was constructed from the difference between the number of changes observed at three years and the number observed at mid-term review. This included appearance or change in radiolucencies, osteolysis, cortical hypertrophy and component alignment. Steady state wear rate was treated as a separate variable.

**Statistical analysis.** Hierarchical multiple regression analysis was used to explore the ability of four independent measures to predict radiographic change after controlling for the type of THA. The four measures were the change in OHS score, age of the participant, EQ-5D score and Charlson comorbidity score. The sample size was calculated retrospectively using a recommended formula (38).
The number of participants available pragmatically determined the final study sample but it exceeded the minimum required for four variables (n = 82). Reliability of radiographic measurements between observers was assessed using an intraclass correlation coefficient (ICC) with a two-way mixed effects absolute model. The Kruskal-Wallis test was used to compare variables with a non-parametric distribution. The statistics were analysed using SPSS version 15.0 (SPSS Inc., Chicago, Illinois) and the level of significance was set at 0.05 for all tests.

**RESULTS**

*Participants*

The flow of participants to the final 147 is illustrated in Figure 1. Of those unable to attend due to age and/or comorbidities, none were experiencing problems with their THA or had required further surgery. Two participants were excluded from the final analysis, one due to the type of prosthesis (the only cementless THA in this cohort) and the other due to missing x-ray films from the three year review.
Fig. 1

Flow chart showing participants entered into study
The mean age of the 147 participants was 74.5 years (81.3 years in the cemented THA group and 69.4 years in the hybrid THA group) and 40% were males. The primary diagnosis was osteoarthritis in 92% of participants and the mean time from primary surgery was 7.5 years (range 6.2 to 9.1). There had been no revisions between the early follow up (mean 3.1 years post surgery) and mid-term. One case of deep sepsis was receiving conservative treatment with regular monitoring (included in study cohort).

All femoral components were cemented with metal heads: 83% were Exeter V.40 components with 28mm heads (Howmedica International, London, UK), 8.5% were CPS Plus stems with 28mm heads (Plus Orthopedics AG, Rotkreuz, Switzerland) and 8.5% were Charnley stems with 22.225mm heads (DePuy International, Leeds, UK).

Radiographic assessment
The interobserver reliability coefficients were high: ICC for acetabular inclination was ICC 0.97 (95% confidence interval: 0.93 to 0.99) and for wear was 0.87 (0.69 to 0.95). The mean acetabular inclination was 47 degrees (range 30 to 65) and the mean wear rate for cemented THA was 0.07mm/year compared with 0.12mm/year in the hybrid THA group. There was no acetabular component migration and of the 42 (65%) radiolucencies seen in zone I behind cemented acetabular cups at early follow up, only eight (12%) had progressed to zone II at
mid-term review. The proportion of acetabular components with radiographic changes was the same in both groups (62%).

Femoral stem alignment was neutral in 100 (65%) participants and less than four degrees valgus or varus in 50 (32%); three (2%) were in varus between four and five degrees, and one stem (1%) was in seven degrees valgus (sequelae of slipped upper femoral epiphysis). None of the positions had changed over time and no subsidence was greater than 3.0 mm. There were 94% of femoral stems with radiographic changes in the cemented THA group and 90% in the hybrid THA group.

There were 15 participants (3 cemented, 12 hybrid) with a total of 25 osteolytic lesions including one with known bony metastases in the pelvis (cemented) and the deep sepsis (hybrid). All had developed since the early review and the median size measured with the morphometric grid was seven points (range 2 to 34 points). The small amount of data collected prevented its inclusion as a variable in the regression analysis.

**Statistical analysis**

There were no instances of high strength correlation in preliminary testing of the test variables. Consequently, four independent variables were entered into the hierarchical regression analysis to explore the association with radiographic changes (dependent variable) and summary statistics are presented in Table I.
Hierarchical regression analysis was conducted in two steps with radiographic changes as the dependent variable (Table I). In Step 1, the type of THA was entered as a dichotomous variable and was not significantly related to the changes ($P = 0.06$).

In Step 2, the four variables of interest were entered after controlling for type of THA. Examination of the results showed that none of the variables contributed significantly to the model. The final model explained 5% of the variance in the radiographic changes.

Table I. Summary statistics for test variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>95% CI</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford Hip change score (from 3 to 7.5 years)</td>
<td>1.07 (6.77)</td>
<td>-0.01 to 2.15</td>
<td>0</td>
<td>-13 to 23</td>
</tr>
<tr>
<td>Age</td>
<td>74.5 (9.23)</td>
<td>73 to 75.9</td>
<td>75</td>
<td>47 to 94</td>
</tr>
<tr>
<td>EQ-5D score</td>
<td>0.76 (0.23)</td>
<td>0.79 to 0.87</td>
<td>0.74</td>
<td>-0.07 to 1.0</td>
</tr>
<tr>
<td>Charlson comorbidity</td>
<td>3.61 (1.47)</td>
<td>3.38 to 3.85</td>
<td>3</td>
<td>0 to 8</td>
</tr>
<tr>
<td>Number of radiographic changes</td>
<td>3.71 (2.29)</td>
<td>3.34 to 4.07</td>
<td>4</td>
<td>0 to 14</td>
</tr>
</tbody>
</table>

Key: SD = Standard deviation, CI = confidence interval, EQ-5D = EuroQol 5 dimension questionnaire
Table II. Results of regression analysis

<table>
<thead>
<tr>
<th>Step and variable</th>
<th>$R^2$</th>
<th>$R^2_{change}$</th>
<th>$F_{change}$</th>
<th>Standardised $\beta$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Type of THR</td>
<td>0.02</td>
<td>0.02</td>
<td>3.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Predictive variables</td>
<td>0.05</td>
<td>0.03</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OHS change score</td>
<td></td>
<td></td>
<td></td>
<td>-0.07</td>
<td>-0.84</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td>0.21</td>
<td>1.47</td>
</tr>
<tr>
<td>EQ-5D score</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.42</td>
</tr>
<tr>
<td>Charlson score</td>
<td></td>
<td></td>
<td></td>
<td>-0.18</td>
<td>-1.43</td>
</tr>
</tbody>
</table>

The Kruskal-Wallis test was used for the non-parametric steady state wear rate data. The participants were sub-divided by type of THA and presence or absence of osteolysis. There was a statistically significant difference in steady state wear rate between the four groups, $X^2 (3, n=154) = 15.72, p <0.001$ with higher rates in those with osteolysis (Table III).

Table III. Average wear rate for groups with and without osteolysis

<table>
<thead>
<tr>
<th>Group</th>
<th>Median wear rate in mm/yr (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cemented prosthesis and osteolysis (n = 3)</td>
<td>0.11 (0.1 to 0.26)</td>
</tr>
<tr>
<td>Cemented without osteolysis (n = 62)</td>
<td>0.05 (0.01 to 0.24)</td>
</tr>
<tr>
<td>Hybrid prosthesis and osteolysis (n=12)</td>
<td>0.16 (0.04 to 0.49)</td>
</tr>
<tr>
<td>Hybrid without osteolysis (n = 77)</td>
<td>0.08 (0.1 to 0.57)</td>
</tr>
</tbody>
</table>
The Kruskal-Wallis test was also used to explore the difference in wear rate between five age groups. There was no statistically significant difference between the groups, $X^2 (4, n=154) = 4.0$, $p = 0.41$ (Figure 2).

Key: Group 1= 45 to 54yrs (4 participants), 2= 55 to 64yrs (18), 3= 65 to 74yrs (50), 75 to 84yrs (63), 5= 85 to 99yrs (19)

**Fig. 2**

Box plot showing the difference in steady state wear rate between five age groups
**DISCUSSION**

The aim of this study was to determine if an x-ray was needed in addition to a hip specific patient reported outcome score at mid-term review. The results showed that, in a cohort of cemented (42%) and hybrid (58%) THA patients, the x-ray changes at mid-term could not be predicted by changes in the Oxford Hip Score, nor by a general health score or comorbidity score or the patient's age. Although the orthopaedic community repeatedly emphasise the importance of radiographic review, this study provides research evidence to support this claim and to refute suggestions that adequate surveillance can be achieved with the use of PROMs alone.

We acknowledge limitations in this study including potential bias introduced by the dual role of researcher and observer. However, interobserver reliability coefficients were high, indicating uniformity between observers (39). The variable for radiographic changes was simplistic but simple and innovative methods of combining information have been used elsewhere including the Swedish Registry (13,33). Such a method can be applied by any member of the specialist orthopaedic team, it allows for different types of degenerative change and it captures the progression that is essential for hip arthroplasty review (40). Our results for steady state wear rate were comparable with other studies with a similar association between higher wear rates and presence of osteolysis but no association between wear rate and age (1,32,33,41).
In the current economic environment, orthopaedic surgeons are facing significant cuts in the provision of services and arthroplasty follow up is one of the casualties. The suggestion that all follow up can be done in the community by general practitioners is perceived to be a cost saving exercise. However, the silent nature of aseptic arthroplasty failure and the need for a member of the specialist orthopaedic team to be involved in the surveillance of these patients has been documented in the orthopaedic literature (42). General practitioners are excellent at providing a front line service but a patient is unlikely to consult them for an asymptomatic condition. If the condition is ignored until symptomatic, the cost of revision is potentially considerable in financial and human terms (43)(44). Therefore, some form of stream-lined, long term surveillance should be maintained.

Within the orthopaedic community, benefits of long term surveillance are recognised but questions remain about when and where this should take place (45). Bolz et al (2010) (6) suggest that there might be no requirement for routine follow up in the first seven years post surgery. Others have suggested that the first signs of aseptic loosening appear in the mid-term (13,14,46) and that predictions about long term survival can be made from a review at this stage (47). The reduction of routine early follow up and replacement by mid-term review may provide sufficient screening for patients with prostheses that have an established track record. However, recent developments have highlighted the
need for more rigorous screening of components with newer and/or modified designs and materials (48). Further research is needed to define the criteria for surveillance that will reduce patient morbidity and the surgical cost associated with asymptomatic failure of THA.

The radiographic changes are an important indication of the state of a THA and this study has shown that the change in Oxford Hip Score cannot predict them even though it was constructed to be joint specific and sensitive to change. Similarly, there was no association between the number of radiographic changes and the age, general health or comorbid status of the participant, suggesting that none of these variables can be used to select THA patients for follow up. Long term surveillance of THA patients is needed to identify asymptotically failing joint replacements but the present economic climate requires any service to be cost-effective. This study provides strong evidence to support the inclusion of an x-ray in addition to patient reported outcome measures in such a service.
REFERENCES


MDA/2012/008 All metal-on-metal (MoM) hip replacements. http://www.mhra.gov.uk
(date last accessed 16 April 2012).
### Tables

**Table I. Summary statistics for test variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>95% CI</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxford Hip change score (from 3 to 7.5 years)</td>
<td>1.07 (6.77)</td>
<td>-0.01 to 2.15</td>
<td>0</td>
<td>-13 to 23</td>
</tr>
<tr>
<td>Age</td>
<td>74.5 (9.23)</td>
<td>73 to 75.9</td>
<td>75</td>
<td>47 to 94</td>
</tr>
<tr>
<td>EQ-5D score</td>
<td>0.76 (0.23)</td>
<td>0.79 to 0.87</td>
<td>0.74</td>
<td>-0.07 to 1.0</td>
</tr>
<tr>
<td>Charlson comorbidity</td>
<td>3.61 (1.47)</td>
<td>3.38 to 3.85</td>
<td>3</td>
<td>0 to 8</td>
</tr>
<tr>
<td>Number of radiographic changes</td>
<td>3.71 (2.29)</td>
<td>3.34 to 4.07</td>
<td>4</td>
<td>0 to 14</td>
</tr>
</tbody>
</table>

Key: SD = Standard deviation, CI = confidence interval, EQ-5D = EuroQol 5 dimension questionnaire
Tables…

Table II. Results of regression analysis

<table>
<thead>
<tr>
<th>Step and variable</th>
<th>$R^2$</th>
<th>$R^2_{\text{change}}$</th>
<th>$F_{\text{change}}$</th>
<th>Standardised $\beta$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Type of THR</td>
<td>0.02</td>
<td>0.02</td>
<td>3.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Predictive variables</td>
<td>0.05</td>
<td>0.03</td>
<td>1.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OHS change score</td>
<td></td>
<td></td>
<td></td>
<td>-0.07</td>
<td>-0.84</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td>0.21</td>
<td>1.47</td>
</tr>
<tr>
<td>EQ-5D score</td>
<td></td>
<td></td>
<td></td>
<td>0.04</td>
<td>0.42</td>
</tr>
<tr>
<td>Charlson score</td>
<td></td>
<td></td>
<td></td>
<td>-0.18</td>
<td>-1.43</td>
</tr>
</tbody>
</table>
### Table III. Average wear rate for groups with and without osteolysis

<table>
<thead>
<tr>
<th>Group</th>
<th>Median wear rate in mm/yr (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cemented prosthesis and osteolysis (n = 3)</td>
<td>0.11 (0.1 to 0.26)</td>
</tr>
<tr>
<td>Cemented without osteolysis (n = 62)</td>
<td>0.05 (0.01 to 0.24)</td>
</tr>
<tr>
<td>Hybrid prosthesis and osteolysis (n=12)</td>
<td>0.16 (0.04 to 0.49)</td>
</tr>
<tr>
<td>Hybrid without osteolysis (n = 77)</td>
<td>0.08 (0.1 to 0.57)</td>
</tr>
</tbody>
</table>