HIP

Retrieval analysis of 240 metal-on-metal hip components, comparing modular total hip replacement with hip resurfacing

This study compared component wear rates and pre-revision blood metal ions levels in two groups of failed metal-on-metal hip arthroplasties: hip resurfacing and modular total hip replacement (THR).

There was no significant difference in the median rate of linear wear between the groups for both acetabular (p = 0.4633) and femoral (p = 0.0872) components. There was also no significant difference in the median linear wear rates when failed hip resurfacing and modular THR hips of the same type (ASR and Birmingham hip resurfacing (BHR)) were compared.

Unlike other studies of well-functioning hips, there was no significant difference in pre-revision blood metal ion levels between hip resurfacing and modular THR.

Edge loading was common in both groups, but more common in the resurfacing group (67%) than in the modular group (57%). However, this was not significant (p = 0.3479). We attribute this difference to retention of the neck in resurfacing of the hip, leading to impingement-type edge loading. This was supported by visual evidence of impingement on the femur.

These findings show that failed metal-on-metal hip resurfacing and modular THRs have similar component wear rates and are both associated with raised pre-revision blood levels of metal ions.

Over the past 15 years, the benefits of hip resurfacing1 and large-diameter modular total hip replacement (THR)2 have been well-documented, leading to a growth in the use of metal-on-metal (MoM) components for hip arthroplasty. The annual report of the National Joint Registry for England and Wales for 2009 stated that 8% of all primary hip arthroplasties were resurfacings and 7% were large-diameter modular THR; the two procedures jointly representing 9708 patients annually.3

There have been many retrieval studies of MoM hips, but only four have investigated current-generation designs, and rates of wear have only been reported for 63 current-generation hip resurfacing couples (acetabular component and femoral head).7,11-13 The rate of wear of current-generation large-diameter modular THR has been reported for only one femoral component and no acetabular components.13

Blood metal ion levels have been shown to be significantly higher in patients with modular THR than in those undergoing resurfacing.14,15 This has been attributed to unspecified wear mechanisms at the modular junction. Fretting and crevice corrosion have been shown at the head-neck and stem-neck interfaces of modular femoral components,16 but it is not known to what extent this influences the accumulation of metal ions. No study has analysed rates of wear and blood metal ion levels of failed current-generation modular THR and compared these with resurfacings.


From Imperial College London, London, United Kingdom

©2011 British Editorial Society of Bone and Joint Surgery
do:10.1302/0301-620X.93B3. 25551 $2.00

Received 30 July 2010; Accepted after revision 17 November 2010

©2011 British Editorial Society of Bone and Joint Surgery
do:10.1302/0301-620X.93B3. 25551 $2.00

Received 30 July 2010; Accepted after revision 17 November 2010

©2011 British Editorial Society of Bone and Joint Surgery
do:10.1302/0301-620X.93B3. 25551 $2.00
The bearing surfaces of MoM hip resurfacing and modular THRs are cobalt-chromium-molybdenum alloy, and the design of the acetabular component is the same. It might therefore be expected that the wear of the bearing surface would be similar between the two types of joint replacement. In this study we aimed to quantify component wear and pre-revision blood metal ion levels in a large series of failed current-generation explanted modular MoM hips and compare these with a matched group of failed hip resurfacings. We also sought to investigate the influence of edge loading and component positioning upon the results.

**Materials and Methods**

In this retrospective study of 120 failed current-generation MoM hips (240 components), the rate of linear wear and blood metal ion levels were compared between explanted modular THRs and hip resurfacings. Ethical approval was obtained beforehand. We included a consecutive series of 60 failed large-diameter MoM THRs. These were matched for gender and head size to within 2 mm with 60 failed large-diameter hip resurfacings. Modular THRs that could not be matched in this manner were excluded. All hips in the study had been implanted for over 12 months and consisted of a large-diameter head (> 38 mm) and a resurfacing monoblock acetabular component. The inclination angle of the acetabular component was measured from the intersection of two lines on an anteroposterior (AP) plain radiograph, one across the face of the component and the other joining the two ischial tuberosities.\(^1\)

Table I summarises the important clinical data of the two groups. The cause of failure was listed according to the categories outlined in the 2008 report of the National Joint Registry for England and Wales.\(^1\)

Altman’s nomogram was used to perform a statistical power calculation in order to estimate the minimum number of components required in each group to show a difference in the rate of linear wear of 5 \(\mu\)m/year, representing a twofold increase over the average steady state wear rate of MoM hips.\(^1\) Linear wear was measured for an initial ten explants in order to estimate the expected SD, which in this pilot study was 7.7 \(\mu\)m/year, giving a standardised difference of 1.07. Statistical significance was set at 0.95, and in accordance with a standardised difference of 1.07, a minimum sample size of 46 (two groups of 23) provided a power of 95%.

All the components were visually inspected for evidence of \textit{in vivo} performance that might provide clues to the cause of failure.
The linear wear of each head and acetabular component was measured using a Taylor Hobson Talyrond 365 roundness machine (Taylor Hobson Ltd, Leicester, United Kingdom). This is a stylus instrument that measures the deviations from a perfectly round profile. The component is mounted and rotated on a spindle (accuracy ± 0.02 μm), while a stylus (ruby 2 mm diameter) in contact with the surface of the component measures deviations from a perfect circle (resolution of stylus gauge 10 nm). The measurements were analysed by the Taylor Hobson software package Ultra and using a specialist program written in Matlab (The Mathworks Inc., Natick, Massachusetts). Two measurements are required to characterise the wear of explanted acetabular components, namely the circumferential and polar measurements.

A total of 12 circumferential measurement profiles were taken along lines of latitude on the bearing surface, parallel to the rim of the acetabular component, at 1 mm increments to a depth of 12 mm below the rim. A maximum inscribed circle, representing the unworn shape of the implant, was fitted to each profile. The maximum distance between the measured profile and the maximum inscribed circle represents the maximum depth of linear wear.

A total of 12 polar measurement profiles were taken along the lines of longitude for each acetabular implant at 15° increments around the rim. The polar measurements allow wear and form to be distinguished from the circumferential measurements. Form errors are deviations from a perfect sphere that might result from the manufacturing...
process or from component deformation during implantation or explantation, whereas wear is material removed in vivo. In low-wearing components the wear and form can be of similar magnitude, and it is important to be able to separate wear and form in the analysis.

The heads were measured by taking 12 polar profiles, similar to those taken for the acetabular components. A minimum circumscribed arc was fitted through each profile, representing the ideal manufactured shape of the head. However, head components are subject to form error of up to 20 μm, making it impossible to identify wear scars from a single profile. In order to separate wear and form, all 12 profiles were superimposed using a program in Matlab. This allows the worn and unworn regions of each profile to be clearly identified. The maximum depth of linear wear can be calculated from the maximum deviation between worn and unworn profiles, thereby separating the form error from the wear calculation.

The measurements provide the maximum depth of linear wear (μm). The maximum rate of linear wear is calculated by dividing this by the time, in years, from implantation. This method allows the position, extent and depth of the wear patch to be measured, thereby identifying edge loaded acetabular components. Acetabular components were identified as edge loaded if the maximum linear depth of the wear scar occurred at its rim (Fig. 1).

**Statistical analysis.** As normality was not observed in the distribution of linear wear rate and blood metal ion levels, a non-parametric approach was used for statistical analysis. The Mann-Whitney U test compared median rates of linear wear and blood metal ion levels between the groups, Spearman’s rank test was used to assess correlations with the rate of linear wear, and Fisher’s exact test compared differences in the proportion of edge-loaded components in the two groups. A p-value < 0.05 was considered significant.

**Results**

A summary of the clinical parameters is shown in Table I. Imaging (CT or plain radiographs) was available for 103 of 120 patients and blood metal ion analysis was performed pre-operatively in 84 of 120.

There was no significant difference in median whole blood Co and Cr levels between the two groups of implants (Fig. 2).

Macroscopic inspection of the components commonly revealed fine scratching on the bearing surfaces in both groups, and in some cases the location of the wear patch could be seen. There was evidence of impingement on the femoral neck of three resurfacing heads (Fig. 3).

The linear wear data for the two groups are summarised in Table II. There was no significant difference in the median rate of linear wear between the modular THR and hip resurfacing groups for the acetabular and femoral components (Fig. 4). In both groups there was a strong positive
correlation between acetabular and femoral component wear, with Spearman’s rank correlation coefficient for modular THR and hip resurfacing 0.79 and 0.83, respectively (p < 0.001 in both groups).

Rates of linear wear were compared between the two most common types of modular and resurfacing hip arthroplasty (ASR (DePuy, Leeds, United Kingdom) and BHR (Smith & Nephew, Warwick, United Kingdom), Table III). There was no significant difference in the median rates of linear wear for both types.

Edge-loaded acetabular components had higher rates of wear. In the modular THR group the median rates of linear wear were 16.87 (0.82 to 119.15) and 0.00 (0.00 to 4.77) μm/year for edge-loaded and non-edge-loaded implants, respectively. In the hip resurfacing group the median rates of linear wear were 11.00 (0.77 to 173.81) and 0.00 (0.00 to 6.18) μm/year for edge-loaded and non-edge-loaded implants, respectively. The differences in median linear wear rates between edge-loaded and non-edge-loaded components in both groups were significant (Mann-Whitney U test, p < 0.001).

There was a significant positive correlation between the linear wear rate of the acetabular component and the angle of inclination of the component (Spearman’s rank correlation coefficient = 0.21, p = 0.043). However, in both groups, high-wearing edge-loaded acetabular components were observed across the entire range of inclination angles (Fig. 5).

**Discussion**

In our comparison of wear rates and blood metal ion levels between failed hip resurfacing and modular THR arthroplasties, the main findings were the lack of significant difference in the median rates of linear wear and pre-revision blood metal ion levels between the two groups. This has not been shown before and is clinically important, as it suggests that the problems relating to high wear in hip resurfacing are likely to be similar for all types of large-diameter MoM hip arthroplasty.

Several variables have been shown to influence either the failure or the rate of wear of MoM hips, namely gender, femoral head size, amount of time implanted, alignment of the acetabular component and component design. Along with the large sample size, evaluation of all these variables was a particular strength of our study.

Gender and head size have both been shown to affect the rate of failure in MoM hips. By matching the two groups for these two variables, their potential confounding effect was minimised. The wear rate of MoM hip bearings has been shown to be biphasic, consisting of a higher-wearing ‘bedding-in’ regime lasting approximately 12 months, followed by a lower-wearing ‘steady-state’ regime. In order to limit the effect of this variable, all included components were implanted for a minimum of 12 months. The effect of inclination of the acetabular component is well documented, but the effects of acetabular version are less well understood. There was a difference of 1° and 10° in the median inclination and version angles of the acetabular component, respectively, between the two groups of implants. It is likely that the effect of acetabular component position on our comparison was small, as the medians were similar and all within Lewinnek’s recommended range.

Component design has been suggested as an important variable influencing clinical performance, but owing to the nature of a retrieval study it is difficult to control. The numbers of each hip type retrieved are often too small to make a direct comparison between different brands.
In further analysis of the two most common hip types in our study, we showed that there was no significant difference in median linear wear rates between failed ASR resurfacing and modular ASR hips, and between BHR resurfacings and modular BHR hips. Because there were fewer components in these subgroups, the power was reduced from 95% for the main comparison between modular and resurfacing hips to 85% and 70% for the ASR and BHR comparisons, respectively. However, these comparisons have not previously been reported, with our report including more ASR and BHR hip couples than any other retrieval study of current-generation MoM hips.7,11-13

Langton et al13 published maximum out-of-roundness measurements for 20 explanted ASR components from 15 patients. It is difficult to draw direct comparisons with results for the ASR and ASR modular THR owing to differences between the measurement and analysis methods and differences in component selection. Their study included only five acetabular and one modular femoral component, and therefore results for acetabular and modular head wear are not readily comparable. Also, several components in their study had been implanted for less than a year, and were therefore likely to be in the initial bedding-in period of high wear.19

Our results support previous reports that edge-loaded MoM hips are higher wearing and that edge loading is positively correlated with the inclination angle of the acetabular component.11 Importantly, we have shown an overall parabolic relationship between wear rate and inclination angle (Fig. 5), with edge loading occurring above, below and within Lewinnek’s recommended range of acetabular component inclination. This does not support previous suggestions that implanting acetabular components at < 40° inclination could reduce wear in MoM hips.23

We also found that edge loading was common in all types of large-diameter MoM hips included in our sample. However, edge loading occurred 10% more often in the hip resurfacing group, and although this was not shown to be significant, it may become so if a larger sample were studied. We believe this is an important difference between hip resurfacing and modular THR.

Whereas hip resurfacing offers the advantage of bone preservation, retention of the femoral neck results in a reduced range of movement and an increased risk of impingement.24 Modular THRs have an increased head-neck ratio, thereby increasing the range of movement and reducing the risk of impingement.2

Our results show that six of nine hip resurfacings with an acetabular component inclination of < 40° were edge-loaded compared with three in ten of modular THRs. We suggest that low acetabular component inclination combined with neck retention increases the risk of anterior impingement, resulting in subluxation and impingement-type edge-loading. This is supported by the visual evidence

Table III. Linear wear data for the most common component designs in the study (ASR (DePuy) and the Birmingham hip (BHR; Smith & Nephew))

<table>
<thead>
<tr>
<th>Component design used</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASR (n = 15)</td>
<td>ASR THR* (n = 26)</td>
</tr>
<tr>
<td>Median acetabular component linear wear rate in μm/year (range)</td>
<td>10.60 (2.66 to 173.81)</td>
</tr>
<tr>
<td>Median head linear wear rate in μm/year (range)</td>
<td>5.62 (0.00 to 84.70)</td>
</tr>
<tr>
<td>BHR (n = 27)</td>
<td>BHR THR (n = 11)</td>
</tr>
<tr>
<td>Median acetabular component linear wear rate in μm/year (range)</td>
<td>2.73 (0.00 to 141.04)</td>
</tr>
<tr>
<td>Median head linear wear rate in μm/year (range)</td>
<td>2.80 (0.88 to 28.09)</td>
</tr>
</tbody>
</table>

* THR, total hip replacement

Scatter plot of linear acetabular wear rate versus component inclination angle. There was a weak but significant positive correlation between rate of wear and angle of inclination. The parabolic trend curve shows that both high and low inclinations are potentially disadvantageous for in vivo performance. Edge-loading was more common outside Lewinnek’s recommended range for inclination of the acetabular component (30° to 50°), but it also occurred within this range (THR, total hip replacement).
of impingement observed in hip resurfacings (Fig. 3). Although not associated with neck retention, impingement-type edge loading has been shown in explanted ceramic-on-ceramic hips.25 Our study also supports previous work on blood metal ions.26

As stated above, edge-loaded high-wearing acetabular components have been shown to occur across the entire range of acetabular inclination for modular THR and hip resurfacing. Adverse acetabular component version may be an important factor related to edge loading in these implants which had an acceptable inclination angle, and be responsible for elevated blood metal ions.26 We suspect the acetabular components of both hip resurfacing and modular THR are susceptible to edge loading with increased anteversion as a result of reduced cover. There are several examples of edge-loaded components with an acceptable inclination of 40° (± 10°) but excessive version (> 25°) in both groups of this study. In hip resurfacing, insufficient version (< 5°) may also be important, and again is related to the risk of anterior impingement due to neck retention. In this study there are also examples of edge-worn hip resurfacing acetabular components with acceptable inclination, but implanted with retroversion.

We suggest that acetabular component version is likely to have an important influence on the wear rate of MOM hips, although this is poorly understood. There are probably many other factors that adversely affect the wear rate which are not yet understood. These may include pelvic tilt, leg-length discrepancy and horizontal femoral offset. Further work is needed to evaluate the effect of all these variables on the rate of wear. Perhaps most importantly, there are several examples of high-wearing edge-loaded implants with acceptable inclination and version. Further work is needed to determine the cause of the variation in wear rate of a well-positioned hip.

Two recent studies demonstrated that blood metal ion levels are significantly higher in patients with a modular THR than with a resurfacing (Table IV).14,15 The authors attributed this to fretting and crevice corrosion at the modular junctions, as the bearing surfaces are the same. Our study has not shown this difference in blood metal ion levels between the two types of failed MoM hip. We attributed the different outcome to the difference in patient groups. The two previous studies were of patients with well-functioning hips, whereas ours was a retrieval study of failed hips.

The Co and Cr levels in our study of failed hips were approximately twice those reported in the two studies of well-functioning hips (Table IV) for which we would anticipate the wear of the bearing surface to be less than in failed explanted hips. This also explains why we did not observe higher blood metal ion levels in the modular THR group. In our study, high wear of the bearing surface is likely to be the dominant mechanism generating metal wear debris and masking the effects of fretting and crevice corrosion. However, in the studies of well-functioning hips,14,15 this might be more influential on the overall metal ion load, resulting in the observed difference between resurfacing and modular THR.

This is the first study of wear of the bearing surface in failed MoM modular hip replacements. The contribution of the elevated levels of metal ions attributable to unspecified wear mechanisms and corrosion at modular junctions is difficult to differentiate from bearing surface wear. Although there are differences in the modular head design between different hips in this study, all the femoral components in the modular hip group used a 12/14 taper design, although some designs also use a separate sleeve, thereby providing another site for wear and crevice corrosion.

The slightly higher median Co and Cr levels in the resurfacing group in our study may have been a result of ‘runaway wear’ in a minority of resurfacings with an inclination angle > 65°. Perhaps as a result of easier implantation, no modular THR acetabular components had an angle of inclination > 61°.

It is important to recognise that explant studies are not entirely representative of the population, as only the wear of failed implants is measured. Further limitations include the difficulty in quantifying visual damage and scratching on the bearing surfaces, and identifying evidence of impingement if insufficient femoral bone is removed with the component.

We evaluated the role of acetabular inclination on rates of wear and edge loading. Although we also included acetabular component version, this is more difficult, given the problems of assessing version on radiographs.27

In conclusion, failed modular THR and hip resurfacing components have similar rates of wear and blood metal ion levels are similar in both groups. There was a trend for edge-loading to occur more commonly following hip resurfacing which we attribute to impingement-type edge loading, which was most commonly observed in acetabular components implanted with low inclination and either excessive or insufficient version.

**Table IV. Studies that have compared whole blood metal ions between current-generation metal-on-metal hip resurfacing and modular total hip replacements (THRs)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Hip type</th>
<th>Patient type</th>
<th>Number</th>
<th>Modular THR</th>
<th>Hip resurfacing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venditti et al</td>
<td>Durom</td>
<td>Functioning well</td>
<td>82</td>
<td>Mean: 1.34</td>
<td>Mean: 1.61</td>
</tr>
<tr>
<td>Garbuz et al</td>
<td>Durom</td>
<td>Functioning well</td>
<td>104</td>
<td>Median: 2.14</td>
<td>Median: 5.09</td>
</tr>
<tr>
<td>This study</td>
<td>Multiple</td>
<td>Pre-revision</td>
<td>84</td>
<td>Mean: 9.05</td>
<td>Mean: 24.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Median: 4.22</td>
<td>Median: 8.64</td>
</tr>
</tbody>
</table>

VOL. 93-B, No. 3, MARCH 2011
The authors are extremely grateful for the help provided by G. Lloyd (Imperial College) and P. Coward (Royal National Orthopaedic Hospital).

This work was funded by the British Orthopaedic Association through an industry consortium of nine manufacturers: DePuy, Zimmer, Smith & Nephew, Biomet, JRI, Finsbury, Corin, Mathys and Stryker. The contract allows for freedom to publish all results.

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