Early clinical failure of the Birmingham metal-on-metal hip resurfacing is associated with metallosis and soft-tissue necrosis

B. Ollivere, C. Darrah, T. Barker, J. Nolan, M. J. Porteous

From West Suffolk Hospital, Bury St Edmunds and Norfolk & Norwich University Hospital, Norwich, England

The rate and mode of early failure in 463 Birmingham hip resurfacings in a two-centre, multisurgeon series were examined. Of the 463 patients two have died and three were lost to follow-up. The mean radiological and clinical follow-up was for 43 months (6 to 90).

We have revised 13 resurfacings (2.8%) including seven for pain, three for fracture, two for dislocation and another for sepsis. Of these, nine had macroscopic and histological evidence of metallosis. The survival at five years was 95.8% (95% confidence interval (CI) 94.1 to 96.8) for revision for all causes and 96.9% (95% CI 95.5 to 98.3) for metallosis.

The rate of metallosis related revision was 3.1% at five years. Risk factors for metallosis were female gender, a small femoral component, a high abduction angle and obesity. We do not advocate the use of the Birmingham Hip resurfacing procedure in patients with these risk factors.

As the patients requiring hip replacement become younger and life expectancy increases the required working life of a total hip replacement (THR) continue to rise. Although successful in the long term, traditional metal-on-polyethylene bearing surfaces suffer from wear and osteolysis. Attempts to reduce the effects of wear and to increase the working life of a THR have resulted in the use of resurfacing arthroplasty and hard-on-hard bearings.

Some of the earlier successful designs of THR involved metal-on-metal bearings but there were difficulties in manufacture and problems associated with the bearing surface. The Charnley low friction arthroplasty, using a metal-on-polyethylene bearing, had better results. However, the Charnley design of metal-on-polyethylene is associated with the production of wear debris and loosening. Recently, there has been renewed interest in hard-on-hard bearing surfaces in an attempt to reduce the effects of osteolysis which is the biggest cause of failure in joint replacement.

Hip resurfacing may have several advantages in young patients since minimal bone resection makes for easier revision and the design allows for an increased range of movement. The initial resurfacing designs were not successful and high rates of fracture, bone resorption and loosening were reported. Both resurfacing and the use of metal-on-metal bearings were abandoned in the 1980s, but the advent of second-generation resurfacings has renewed interest in these techniques.

The Birmingham hip resurfacing (BHR) (Smith & Nephew Orthopaedics, Warwick, United Kingdom) is the most widely used of the current generation of metal-on-metal hip resurfacings. It was developed as an evolution of the McMinn prosthesis (Corin Medical, Cirencester, United Kingdom) and is a cobalt-chrome polar metal-on-metal bearing with a stemmed femoral component and a hemispherical hydroxyapatite-coated acetabular implant. The design of the prosthesis with a large head and low surface roughness potentially allows for appropriate radial clearance and articulation through thick-film lubrication, dramatically reducing wear and the release of metal ions, thus possibly avoiding the metallosis and wear-related failure associated with previous metal-on-metal and resurfacing designs.

Recently, unexplained pain, early failure and the formation of pseudotumours have been recorded. Joint registry data have also shown an unexplained high early rate of failure for all designs of hip resurfacing. In addition to concerns about early loosening and soft-tissue reaction, there are conflicting reports on the amount of bone conservation, particularly on the acetabular side. We have therefore examined the rate and mode of early failure of the BHR in a two-centre series.

Patients and Methods

The records of 463 consecutive primary BHR metal-on-metal resurfacing procedures performed by five surgeons (AA, MJP, CM, JW and
JN) at two centres between 2001 and 2007 were reviewed. Two surgeons (MJP and AA) performed 341 of the arthroplasties and the other three (JN, CM and JN) the remaining 122. There were 307 men and 156 women with a mean age of 56 years (SD 7.17, 20 to 70). All the patients were recruited into the local arthroplasty follow-up programme at the time of the primary surgery. Clinical data and details of the components were collected at the time of surgery. The patients were reviewed clinically and radiologically at six weeks and at one, two and five years post-operatively. All the radiographs were taken using a standard protocol and included an anteroposterior and an oblique lateral view.

Patients with a failed BHR were identified from the databases at each centre. Case notes, the intra-operative record, radiographs, clinical findings and, when appropriate, histological samples taken at the time of revision were reviewed. The latter were reported initially by local histopathologists, but for the purposes of our study were reassessed by a histopathologist (TB) with experience in metal wear and debris.

The post-operative radiographs were reviewed by two independent observers (BO, CD) to determine the position of the components and to identify technical errors using the methods described by Amstutz et al.,17 De Smet, Pattyn and Verdonck,18 and Hing et al.19 The most recent radiographs were assessed in conjunction with pre-operative and immediate post-operative radiographs. The position of the component was measured using a digital PACS system by two observers (BO, CD) and the mean of both measurements was taken. The femoral and acetabular components were assessed for signs of loosening according to the criteria of Hodgkinson, Shelley and Wroblewski20 and Amstutz et al.17 Three patients underwent MRI before revision surgery with metal artifact reduction sequence (MARS) suppression according to the protocol of Toms et al.21

Data for the patients who had died or were lost to follow-up were included in the survival analysis. Those lost to follow-up were treated as failures for the purposes of survivorship analysis.

Statistical analysis. This was undertaken using Graphpad Prism software (GraphPad software, San Diego, California). Survivorship analysis was calculated using the Kaplan-Meier method with Wilcoxon-Gehan’s method for calculation of the confidence interval (CI). Survival curves were tested for differences using the Mantel-Cox (log-rank) test. Continuous data were analysed for differences using a two-tailed Student t-test for parametric data and the Mann-Whitney U-test for non-parametric data. A p-value < 0.05 was considered significant. Categorical data were analysed by Fisher’s exact test and, when appropriate, relative risks were calculated.

Results

Two patients died from unrelated causes and three were lost to follow-up. All of the remaining 458 patients completed a recent patient-reported outcome score (Harris hip22 or Oxford score23). However, 21 had not attended for their scheduled clinical review and radiography. Consequently, 437 were appropriately reviewed. The mean follow-up was for 43 months (6 to 90). The total observed component years were 1643. Intra-observer reliability for the radiological measurements had an excellent correlation (κ = 0.97).

We revised 13 BHRs (2.8%) in 12 patients (Table I). The case notes, radiographs and MR scans were available for all, but histological findings for only 12. This allowed us to identify the definitive cause of failure in these cases. Seven BHRs were revised for pain, three for fracture, two for dislocation and one for infection. All of the seven revised for pain and the two for fracture were found to have evidence of metallosis. In one case of fracture, the cause of failure could not be established since the histological

<table>
<thead>
<tr>
<th>Case</th>
<th>Age (yrs)</th>
<th>Gender</th>
<th>Presentation</th>
<th>Diagnosis</th>
<th>Head size (mm)</th>
<th>Component size (mm)</th>
<th>Interval to revision (mths)</th>
<th>Abduction angle (°)</th>
<th>Neck-shaft angle (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56</td>
<td>F</td>
<td>Dislocation</td>
<td>Dislocation</td>
<td>38</td>
<td>44</td>
<td>3</td>
<td>65.0</td>
<td>133.9</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>F</td>
<td>Dislocation</td>
<td>Dislocation</td>
<td>38</td>
<td>44</td>
<td>34</td>
<td>53.0</td>
<td>145.0</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>M</td>
<td>Fracture</td>
<td>AVN</td>
<td>50</td>
<td>56</td>
<td>3</td>
<td>48.4</td>
<td>127.0</td>
</tr>
<tr>
<td>4</td>
<td>47</td>
<td>M</td>
<td>Infection</td>
<td>Infected</td>
<td>50</td>
<td>56</td>
<td>32</td>
<td>59.0</td>
<td>135.0</td>
</tr>
<tr>
<td>5</td>
<td>58</td>
<td>F</td>
<td>Pain</td>
<td>ALVAL</td>
<td>42</td>
<td>50</td>
<td>34</td>
<td>62.0</td>
<td>126.0</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>F</td>
<td>Pain</td>
<td>ALVAL</td>
<td>42</td>
<td>48</td>
<td>24</td>
<td>72.0</td>
<td>122.0</td>
</tr>
<tr>
<td>7</td>
<td>59</td>
<td>F</td>
<td>Pain</td>
<td>ALVAL</td>
<td>46</td>
<td>54</td>
<td>12</td>
<td>59.7</td>
<td>137.7</td>
</tr>
<tr>
<td>8</td>
<td>56</td>
<td>F</td>
<td>Pain</td>
<td>ALVAL</td>
<td>42</td>
<td>48</td>
<td>12</td>
<td>58.2</td>
<td>137.6</td>
</tr>
<tr>
<td>9</td>
<td>55</td>
<td>F</td>
<td>Pain</td>
<td>ALVAL</td>
<td>42</td>
<td>48</td>
<td>13</td>
<td>55.6</td>
<td>140.0</td>
</tr>
<tr>
<td>10</td>
<td>56</td>
<td>M</td>
<td>Pain</td>
<td>ALVAL</td>
<td>50</td>
<td>56</td>
<td>7</td>
<td>49.2</td>
<td>131.0</td>
</tr>
<tr>
<td>11</td>
<td>64</td>
<td>F</td>
<td>Fracture</td>
<td>ALVAL</td>
<td>42</td>
<td>50</td>
<td>36</td>
<td>50.0</td>
<td>130.0</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>F</td>
<td>Fracture</td>
<td>ALVAL</td>
<td>42</td>
<td>48</td>
<td>53</td>
<td>49.1</td>
<td>128.7</td>
</tr>
</tbody>
</table>

* AVN, avascular necrosis; ALVAL, aseptic lymphocytic-vasculitis-associated lesion
samples were not available. The failure rate was 0.91 per 100 observed component years. All the patients were revised by the primary surgeon to a cemented taper slip femoral component (Exeter Universal Stem, Stryker, Newbury, United Kingdom) with a variety of bearing surfaces.

Survivorship. Survivorship analysis for all causes of revision (Fig. 1) gave a rate of survival of 95.8% (CI 94.1 to 96.8) at five years and 96.7% (CI 94.3 to 98.1) at a mean follow-up of 3.5 years. For metallosis requiring revision as an end-point the survival was 96.9% (CI 95.5 to 98.3) at five years making the likely incidence of metallosis-related failure 3.1% at five years. There was no statistically significant difference between the curves for all-cause and metallosis-related revision ($p = 0.43$).

Revision of pain. Seven of the 13 BHRs were revised for unexplained pain. In all cases the clinical review at six weeks and subsequent hip scores did not demonstrate a poorly performing or unduly painful replacement. The patients presented with thigh and groin pain at a mean of 15 months (6 to 24) after surgery. There was a similar history of a rapid onset of pain over a period of a few weeks with no history of trauma.

There was no evidence of radiological loosening in any patient before revision. Comparison with post-operative radiographs did not indicate migration or loosening of the component. All the patients had a normal level of CRP, ESR and white cell count indicating a low likelihood of infection.24

Operative findings in these cases typically included large quantities of pale-grey sterile fluid resembling thin purulent material, necrotic areas of bone (affecting a substantial portion of the roof of the acetabulum in one and areas of inflammatory tissue with black-staining soft tissue in five). In three cases well-circumscribed fluid-filled soft-tissue tumours were identified.

Other failures. Three patients were revised for fracture of the femoral neck at three, 36 and 53 months. None had any preceding pain and post-operative radiographs did not show any femoral notching. Pre-fracture radiographs showed collapse of the femoral neck in the patient with the early fracture. The radiographs of both patients presenting with late fractures showed no signs of loosening or collapse of the femoral neck. In both of these thinning of the femoral neck of approximately 20% was seen.19 No histopathological findings were available for the early presenting fracture, but an intra-operative diagnosis of avascular necrosis (AVN) was made based on the macroscopic appearance of large osteonecrotic lesions observed in the sectioned femoral neck. In both late-presenting fractures histopathological specimens were taken, which confirmed changes that were likely to be associated with the response to metal wear debris.

The two patients who were revised for dislocation had no evidence of metallosis or infection. The operative findings confirmed malposition of a component. One patient who was revised for deep acute Staphylococcus aureus infection presented nine months after surgery. Microbiological samples confirmed the infection but there was no histological evidence of metallosis or of a response to wear debris.

Histological examination. In all the patients except one (case 3) intra-operative histological and microbiological samples were taken. In nine the histological findings by both histopathologists were consistent with a response to...
wear debris. All seven BHRs revised for unexplained pain and both patients with a late fracture of the neck had features consistent with a metallosis reaction. In one patient with late fracture only bone samples were taken.

In all except one patient there were areas of necrosis (Fig. 3). The characteristic aseptic lymphocytic-vasculitis-associated lesion (ALVAL) was seen in seven patients, and necrosis in eight. Histological examination in all the remaining five revised patients showed no features of metal hypersensitivity or a response to wear debris.

**Risk factors.** The male-to-female ratio in the metallosis group was 1:3.5 as compared with 1:0.55 for the whole series (p = 0.018). The relative risk factor for the female gender was 4.94 (CI 1.33 to 18.31).

The mean size of the femoral component was 44 mm (median 42 mm) in the failure series and 48 mm (median 50 mm) in the surviving series (Mann-Whitney U test, p = 0.002). The mean abduction angle was 55.1° (38° to 72°) which was also significantly higher than that of the whole series (47.1°, p < 0.03). The mean body mass index (BMI) in the whole series was 27.7 (CI 27.5 to 27.9) compared with 30.4 (CI 29.3 to 31.6) in those which were revised. This difference was statistically significant (Student’s t-test, p = 0.034). There was no statistically significant difference in the mean age between the two groups.

**Discussion**
Modern metal-on-metal articulations have been associated with catastrophic failure,

but also good results in the short and intermediate terms.

Metallosis has not been widely reported with the BHR although concerns have been raised about the levels of metal ions and the possibility of toxicity in a wide range of metal-on-metal articulations.

This is the first report of a single-implant metal-on-metal resurfacing describing the causes and incidence of early failure and is also the largest independent series reported of a single-implant metal-on-metal resurfacing arthroplasty. Our series has shown a similar survivorship to those previously reported and a rate of failure of 0.91 per 100 observed component years is lower than that reported from the Australian Joint Registry. The rate of dislocation of 0.4% (n = 2/463) and infection of 0.2% (n = 1/463) are widely accepted complication rates and similar to those in other published series.

The rate of metallosis-related failure has been previously reported at approximately 1%. The results for our series suggest that the true rate may be higher, with an absolute rate of failure of 1.9% (n = 9/462) at a mean of 3.5 years, and survival analysis indicated a rate of failure of over 3% at five years. This difference may be explained by the mix of different resurfacing implants reported in the Oxford series. It is likely that the true rate of failure lies nearer to 3% since our study presented the findings of the use of a single implant in a two-centre, multisurgeon series.

Modern metal-on-metal articulations have been associated with raised serum and urine levels of metal ions. It is reasonable to suppose that an increase in the production of metal ions will increase the number and rate of any failures associated with their production. Our finding of an increased abduction angle in the failure group supports the suggestion that early failure is due to metal-ion production. However, the inclusion of two centres and five surgeons makes poor surgical technique unlikely, particularly since there were no statistically significant differences in the rates of failure between each surgeon and centre.

The position of the component is a surgeon-controlled factor, and while desirable, it is not always possible to...
achieve an optimal position. The abduction angle is only one factor in the optimal position. Version cannot be accurately measured on plain films, hence it was not possible to assess reliably the adequacy of version in our study. De Haan et al\textsuperscript{12} reported a series of 42 revised metal-on-metal resurfacing arthroplasties. They identified malposition of the acetabular component as the cause of failure in most (64\%) with abduction angles greater than 56\(^\circ\) being associated with early failure. Our findings support this. Six of the 13 patients who had revision had abduction angles of more than 56\(^\circ\). A relationship between the position of the acetabular component and serum levels of metal ions has also been reported.\textsuperscript{38,39}

Histopathological examination showed a range of changes, all of which were features that have been associated with the response to wear debris\textsuperscript{40} and metallosis.\textsuperscript{13} The precise nature of the metal-related wear-debris response and the relationship of ALVAL, metallosis and pseudotumours has been the subject of much debate.\textsuperscript{14} Tissue samples from our series showed a range of inflammatory changes including necrosis, inflammation, ALVAL and metal-containing macrophages. Not all of the features were seen in every patient and it is likely that they form part of the spectrum of the response to metal-wear debris. The existence of metallosis-associated necrosis and failure has been recognised since 1975,\textsuperscript{5,41} but the typical histological changes seen here of perivascular lymphocyte infiltration, inflammation, the accumulation of macrophages loaded with metal debris and surface ulceration have only been described recently.\textsuperscript{42}

While it is possible that ALVAL and metallosis may represent part of the normal spectrum of post-operative changes associated with a metal-on-metal articulation, the patients revised for other reasons had no evidence of any of these changes. This makes it unlikely that metal-debris disease is part of the normal post-operative spectrum associated with metal-on-metal articulations. De Haan et al\textsuperscript{12} identified changes associated with metallosis in 29\% in their series of revised metal-on-metal resurfacings. Our series would suggest that the rate of metallosis-associated failure is closer to 70\% (\(n = 9/13\)) of all failed BHRs. The high rate of metallosis observed in our series may be in part explained by a specific request to look for evidence of ALVAL or metallosis in the intra-operative samples or may have been due to our reporting on a single prosthesis, whereas other series had a variety of prostheses.

In our series, patients were more likely statistically to require revision surgery for pain if they had a high abduction angle, were female, had a small femoral component or were obese. All of these variables are associated with higher rates of wear.

Although in hard-on-soft bearings the rate of wear is inversely proportional to the size of the head\textsuperscript{43} this is not the case in hard-on-hard bearings such as the BHR. This is due to a thick-film lubrication,\textsuperscript{7} which results in a much reduced coefficient of friction and decreased wear debris. Thick-film lubrication is more likely to occur as the diameter of the head increases.\textsuperscript{10} An increased rate of wear would be expected in obese patients, those with small femoral heads\textsuperscript{44} and those with a high abduction angle. The association between female gender and a high rate of failure (relative risk 4.94) may be a confounding factor since a smaller femoral head is required in women. Similarly, it is possible that the demonstrated association between obesity and failure actually represents a confounding factor since the difficulty in obtaining adequate surgical access in the obese is likely to make adequate alignment of the component more challenging.

If metallosis is a response to metal-wear debris rather than a hypersensitivity reaction then, although we have not measured rates of wear, it is reasonable to suppose that earlier failure will be seen in patients with a higher rate of generation of wear debris. However, the correlation between the abduction angle and time to failure was poor (\(\kappa = 0.23\), Pearson's test). It seems likely that if failure is due to a dose-response effect to wear debris then the failure should increase dramatically with time.

We are concerned by the fact that two of the three fractures in our series occurred more than three years after surgery. Early rates of fracture of up to 2.4\% have been reported\textsuperscript{45} and concern has been expressed regarding thinning of the femoral neck,\textsuperscript{19} but we were not aware of reports until now that spontaneous late-presenting fracture of the femoral neck is a recognised complication of the BHR. In both patients with late fractures there was no radiograph available immediately before the fracture occurred because of the planned follow-up intervals. However, thinning of the femoral neck was visible on the most recent films. If thinning and fracture are related then the rate of fracture may be expected to increase with time. We have been unable to establish histologically the cause of failure in the patient with the early fracture of the neck, although it appeared macroscopically as avascular-necrosis-related collapse and it should be remembered that bone necrosis is a feature of metallosis.

The rate of early failure in our series was comparable with that of other published series in the literature.\textsuperscript{9,12} It gives us cause for concern, especially in the light of the young active patient group for which this prostheses is indicated. Other prostheses have a much higher observed rate of survival\textsuperscript{2,46-50} even in young patients.\textsuperscript{51,52}

The decision to revise a painful arthroplasty is always difficult and controversial. The lack of radiological changes seen in our patients despite widespread tissue necrosis and bone destruction found at revision is alarming. We would strongly advocate MRI for these patients and a low threshold for revision in painful metal-on-metal articulations to avoid the difficulties that tissue destruction can cause at the time of revision.

BHR is associated with a significant rate of early failure which may be as high as 3\% at five years. Histological changes associated with metallosis and a response to wear debris are seen in most cases of early failure. Risk factors for early failure include female gender, a small size of fem-
oral head, a greater abduction angle and a high BMI. We would advise caution in the use of this prosthesis in patients with these risk factors and urge a low threshold for revising patients presenting with a painful resurfacing.

The authors wish to thank A. August (West Suffolk Hospital), J. Wimhurst and C. Mann (Norfolk and Norwich Hospital) for contribution of patient data to this study.

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