We report the findings of an independent review of 230 consecutive Birmingham hip resurfacings (BHRs) in 213 patients (230 hips) at a mean follow-up of 10.4 years (9.6 to 11.7). A total of 11 hips underwent revision; six patients (six hips) died from unrelated causes; and 13 patients (16 hips) were lost to follow-up. The survival rate for the whole cohort was 94.5% (95% confidence interval (CI) 90.1 to 96.9). The survival rate in women was 89.1% (95% CI 79.2 to 94.4) and in men was 97.5% (95% CI 92.4 to 99.2). Women were 1.4 times more likely to suffer failure than men. For each millimetre increase in component size there was a 19% lower chance of a failure. The mean Oxford hip score was 45.0 (median 47.0, 28 to 48); mean University of California, Los Angeles activity score was 7.4 (median 8.0, 3 to 9); mean patient satisfaction score was 1.4 (median 1.0, 0 to 9). A total of eight hips had lysis in the femoral neck and two hips had acetabular lysis. One hip had progressive radiological changes around the peg of the femoral component. There was no evidence of progressive neck narrowing between five and ten years.

Our results confirm that BHR provides good functional outcome and durability for men, at a mean follow-up of ten years. We are now reluctant to undertake hip resurfacing in women with this implant.
Before beginning the review process the patients’ names, hospital record numbers and dates of surgery were provided to the Australian Joint Replacement Registry in order to confirm the number of BHRs that had been revised and the number of patients who had died. Registry privacy policy did not allow for individual identification of the patients who had undergone revision of their BHR.

All patients were initially contacted by mail and offered an appointment for a clinical and radiological review. Patients who failed to respond were contacted by telephone, and an appointment was made where possible. For patients who had moved to another city and were unable to be seen by the primary author (GC), arrangements were made for clinical and radiological assessment by an orthopaedic surgeon located in a city convenient to the patient. If the patient was either unwilling or unable to return for follow-up, we established whether their implant had been revised by telephone contact.

Operative technique. The three senior authors used a posterior approach, as previously described. The external diameter of the acetabular components ranged from 48 mm to 64 mm. The diameter of the femoral component was 6 mm or 8 mm smaller than the corresponding acetabular component. One dysplasia implant, which is an acetabular component permitting supplementary screw fixation, was used.

Clinical analysis. Post-operative assessment at a mean of ten years consisted of a patient satisfaction score (PSS), Oxford hip score (OHS), hip disability and osteoarthritis outcome score (HOOS), which includes elements drawn from the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Charnley grade, University of California, Los Angeles (UCLA) activity score, and assessment of range of movement.

Radiological assessment. Anteroposterior (AP) and lateral radiographs were taken at a standard magnification of 115%. Quantitative radiological assessment included measurement of femoral component stem–shaft angle, acetabular component abduction angle and femoral neck narrowing (Fig. 1). Qualitative evaluation of radiolucent lines (RLLs), sclerosis and osteolysis around the implant stem, scalloping of the femoral neck–implant junction, and osteolysis around the acetabular component was performed by a single observer (GC) using previously described methods. Owing to an inability to standardise patient rotation for the lateral radiograph, version of the components could not be accurately measured.

The change in femoral component stem–shaft angle between five years and ten years was measured to determine whether there had been any valgus or varus migration of the femoral component. The measurements of

Table I. Summary of the study cohort

<table>
<thead>
<tr>
<th>Gender (n, %)</th>
<th>Study cohort (n = 213 patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>140 (66)</td>
</tr>
<tr>
<td>Female</td>
<td>73 (34)</td>
</tr>
<tr>
<td>Mean age at operation (yrs) (range)</td>
<td>52.1 (18 to 82)</td>
</tr>
<tr>
<td>Mean height (cm) (range)</td>
<td>172.18 (150 to 195)</td>
</tr>
<tr>
<td>Mean weight (kg) (range)</td>
<td>80.62 (55 to 110)</td>
</tr>
<tr>
<td>Mean body mass index (kg/m²) (range)</td>
<td>27.02 (16.2 to 46.3)</td>
</tr>
<tr>
<td>Mean time to follow-up (yrs) (range)</td>
<td>10.4 (9.6 to 11.7)</td>
</tr>
</tbody>
</table>

Fig. 1a – anteroposterior radiograph showing the technique for measuring the diameter of the femoral neck. The diameter of the neck (N) at the junction with the femoral component was divided by the diameter of the femoral component (FC), producing a ratio (N/FC) to account for any changes in radiological magnification. Figure 1b – diagram showing the measurement of stem–shaft angle (S) and acetabular component abduction angle (C). (Reproduced from Hing CB, Back DL, Bailey M, et al. Narrowing of the neck in resurfacing arthroplasty of the hip: a radiological study. J Bone Joint Surg [Br] 2007;89-B:1019–24).
acetabular component abduction angle and stem–shaft angle were repeated by an independent reviewer to determine the inter-observer reliability of these measurements.

**Statistical analysis.** Data were collected using the SOCRATES database software v.3.5 (Ortholink Pty Ltd, Sydney, Australia). Statistical analysis was performed by an independent statistical consultant.

**Predictors of functional outcome.** Functional outcome variables were all highly skewed and therefore descriptive data, including median, minimum and maximum, were presented first. Potential predictors of functional outcome were assessed using exact Mann-Whitney U tests for dichotomous variables, exact Mann-Whitney tests for categorical variables, and Spearman’s rank correlation for continuous variables.

**Predictors of failure.** Exact chi-squared tests were used first to explore univariate predictors of failure. Generalised linear models (GLM) were used to analyse predictors of failure. Poisson regression with robust error terms was used to emulate a log binomial GLM, as the latter would not converge. In order to separate out the effects of confounding, a two-sided chi-squared test was performed. Allowance was made in the model for clustering by patient. The relationship between device size, gender and failure was further explored using a structural equation model with asymptotically distribution-free fitting. The model was repeated using a Bayesian Monte Carlo Markov chain (MCMC) procedure, treating failure as a dichotomous variable, and provided a similar pattern of results. Kaplan-Meier survival analysis was undertaken on the datasets, both overall and by gender, with 95% CIs. A log rank test was used to compare device survival by gender. Cox’s proportional hazards regression was undertaken to compare survival by gender after adjusting for clustering.

The impact of gender on the range of movement of the hip was assessed using independent samples *t*-tests.

Statistical analyses were undertaken using SPSS Statistics 19 (IBM, Armonk, New York), AMOS 19 (IBM) and MedCalc 11 (Medware Software bvba, Mariakerke, Belgium). A statistical significance level of 95% was chosen (p < 0.05).

**Results**

A total of six patients (six hips) died from causes unrelated to the BHR. All of these had well-functioning implants at the time of death. Our review and the registry search both identified 11 revisions; 13 patients (16 hips) were lost to follow-up. The mean time of follow-up for the remaining 184 patients (197 hips) was 10.4 years (9.6 to 11.7), which included 23 patients (25 hips) who owing to geographical constraints were unable to undergo a clinical assessment at ten years but confirm that they had not had a revision procedure. The survival data from these 25 hips are included in this report.

**Clinical analysis.** The early and mid-term clinical results have been published3,5; the results at ten years are presented in Table II. Patients had a high level of function at ten years; the mean OHS was 45.0 (median 47.0, 28 to 48), the mean HOOS WOMAC total score was 5.4 (median 1.0, 0 to 41.0), and the mean UCLA activity score was 7.4 (median 8.0, 3.0 to 9.0). The mean satisfaction score was 1.4 (median 1.0, 0 to 9.0) indicating a high level of satisfaction. The mean hip flexion was 106.7° (median 110°, 90° to 140°) as measured in 172 hips. From the 161 patients (172 hips) who underwent a clinical follow-up at ten years, 39 hips (23%) were classified as Charnley category A, 78 (45%) were Charnley B, 36 (21%) were Charnley C and 19 (11%) were unclassified.

| Table II. Clinical outcome scores at a mean of ten years |
|----------------|----------------|----------------|
| **Clinical outcome** | **Mean (range)** | **Median** | **SD** |
| Patient satisfaction normal scale | 1.4 (0.0 to 9.0) | 1.0 | 1.45 |
| Range of flexion (°) | 106.7 (90.0 to 140.0) | 110.0 | 9.84 |
| Oxford hip score | 45.0 (28 to 48) | 47.0 | 4.14 |
| **HOOS** | | | |
| Symptoms | 92.6 (60 to 100) | 95.0 | 10.34 |
| Pain | 95.6 (53 to 100) | 100.0 | 8.77 |
| Daily living | 94.1 (56 to 100) | 99.0 | 9.24 |
| Sports | 85.7 (13 to 100) | 94.0 | 18.96 |
| Quality | 88.7 (44 to 100) | 94.0 | 14.72 |
| **HOOS WOMAC** | | | |
| Pain | 0.8 (0 to 8) | 0.0 | 1.66 |
| Motion | 0.6 (0 to 4) | 0.0 | 1.00 |
| Function | 4.0 (0 to 30) | 1.0 | 6.26 |
| Total | 5.4 (0 to 41) | 1.0 | 8.40 |

* HOOS, hip disability and osteoarthritis outcome score
† HOOS WOMAC, elements of HOOS drawn from the Western Ontario and McMaster Universities Osteoarthritis Index
hip flexion was positively correlated with OHS \((r = 0.219; p = 0.066)\), HOOS daily living score \((r = 0.263; p = 0.001)\), HOOS sports score \((r = 0.224; p = 0.005)\), and negatively correlated with WOMAC function \((r = -0.263; p = 0.001)\) and WOMAC total \((r = -0.228; p = 0.004)\). There was no association between OHS and femoral or acetabular lysis \((p = 0.621, \text{exact Mann-Whitney test})\), acetabular component inclination angle \((r = -0.104; p = 0.194)\), range of internal and external rotation \((r = -0.022, p = 0.789)\); and \(r = 0.079, p = 0.334\), respectively), femoral component head size \((r = -0.006; p = 0.935)\), acetabular component size \((r = 0.019; p = 0.810)\) or gender \((p = 0.907, \text{exact Mann-Whitney test})\).

**Predictors of failure.** The Poisson regression found a significant effect of size on failure \((p = 0.005)\). Because of confounding between size and gender, the outcome of the Poisson regression depended on which variable (size or gender) was excluded from the model. In order to separate out the effects of confounding, a two-sided chi-squared test was performed using a size cut-off of 50 mm, which revealed an effect of size that was not significant \((p = 0.123)\), and a significant effect of gender whereby women were more likely to fail than men \((p = 0.019)\). An attempt was made to define the relationship between gender, size and failure more accurately using structural equation modelling, which identified that women were 1.4 times more likely to fail than men, and that for each millimetre increase in size of the femoral component there was a 19% lower chance of a failure.

**Other analyses.** Women had a significantly greater range of internal rotation than men \((p = 0.022)\), external rotation \((p < 0.001)\) and summation of rotations \((p < 0.001\); all independent samples \(t\)-test).

**Complications.** The immediate and five-year post-operative complications have been previously described.\(^5\) At ten years occasional painless clicking was audible in two patients \(\text{(two hips)}\), and a clicking sensation without audible noise was experienced by 13 patients \(\text{(13 hips)}\). None had ongoing squeaking.

**Failures.** A total of 11 hips in 11 patients \(\text{(eight women, three men)}\) underwent revision. In ten patients these were unilateral BHRs and one patient had a bilateral BHR in which only one component failed. Two revisions have been previously described.\(^3\) The mean age at the initial operation of patients who later required revision was 41.5 years \((19 \text{ to } 58)\). The femoral component head size was \(\leq 46 \text{ mm} in nine patients \(\text{(nine hips)}\) who required revision. Of the 11 patients \(\text{(11 hips)}\) whose BHR failed, two had avascular necrosis \(\text{(AVN)}\) of the femoral head; one had loosening of the femoral component; one had loosening of the acetabular component; one hip had acetabular osteolysis; two patients had unexplained pain; one patient had high blood cobalt/chromium levels; one patient had instability following spinal surgery to remove a tumour which has been previously reported\(^6\); and two hips demonstrated ARMD.

Of these 11 revision procedures, seven were revised at our institution and four at other hospitals. The seven revisions performed under our care were for a variety of reasons. All except one had both components revised to conventional total hip replacement \(\text{(THR)}\). All were uncomplicated procedures except one that involved acetabular bone grafting for osteolysis. All have had good outcomes to date, except for one patient who had a conventional stemmed femoral component and a modular metal head articulating with the original Birmingham acetabular component. This patient has continued to have pain but has not had further revision surgery. Of the four patients who were revised elsewhere, all were converted to conventional THRs. One of those patients developed recurrent instability and died of causes unrelated to the replacement two years later. One of the patients revised elsewhere has since been lost to follow-up.

**Radiological follow-up.** There were 155 patients \(\text{(155 hips)}\) \((73\%)\) with radiographs available for review. Inter-observer measurement reliability for the neck–shaft angle and acetabular inclination angle was determined using an intra-class correlation coefficient. The intra-class correlations for neck–shaft angle and acetabular inclination angle were 0.994 and 0.993, respectively, indicating that there was almost perfect agreement between the scorers. The mean femoral stem–shaft angle was 135.5° \((\text{median } 136.0°, 114° \text{ to } 156°)\). One patient \(\text{(one hip)}\) had late migration of the femoral component into \(> 5° \text{ of varus}\. The mean acetabular component inclination angle was 45.0° \((\text{median } 44°, 30° \text{ to } 66°)\).

A total of eight hips \(\text{(eight patients)}\) showed evidence of lysis in the femoral neck, predominantly in the cranial aspect of the head–neck junction \(\text{(Fig. 2)}\). These lytic lesions were not associated with scalloping or previous neck narrowing. There were two hips \(\text{(two patients)}\) with evidence of acetabular lysis. There was one hip with progressive radiological changes around the peg of the femoral component. There was no evidence of progressive neck narrowing between the five- and ten-year reviews. There was no association between osteolysis and acetabular component inclination angle \((p = 0.972, \text{Wilcoxon rank sum test})\). The presence of lysis had no effect on any of the measures of functional outcome, including the mean HOOS pain subcategory \((p = 0.829, \text{Mann-Whitney})\) and the mean OHS \((p = 0.621, \text{Mann-Whitney})\).

**Radiological risk of failure.** At the five-year follow-up a number of patients demonstrated radiological signs of unknown relevance. Two of the eight failures that have subsequently occurred had significant radiological change noted at our previous review. One had a progressive RLL around the acetabular component. This increased in severity until revision with extensive acetabular bone grafting was undertaken. Another with narrowing of the femoral neck and some lucency at the superior head–neck junction was also revised.

The current radiological review showed two components \(\text{(two patients)}\) thought to be at risk of failure. One femoral component demonstrated progressive varus angulation over a two-year period that was presumed to be secondary.
Survival rate. The cumulative survival rate at ten years was 94.5% (95% CI 90.1 to 96.9) with 11 hips revised (Fig. 3). The cumulative survival rate for women was 89.1% (95% CI 79.2 to 94.4) with eight hips revised (Fig. 4). The cumulative survival rate for men was 97.5% (95% CI 92.4 to 99.2) with three hips revised (Fig. 4). The difference in cumulative survival rate between men and women was statistically significant (log rank, p = 0.009).

Discussion
It has been reported that metal-on-metal hip resurfacing offers a number of advantages over primary THR, including more accurate recreation of the native biomechanics of the hip joint, high post-operative activity levels, a bearing couple that has high resistance to wear, a lower dislocation rate than THR due to the large articulation, and bone preservation on the femoral side so that it is technically easier to revise if necessary. These benefits suggest that third-generation hip resurfacings such as the BHR may be suitable for young, active patients. However, many of these benefits are open to debate and are often supported only by short- to mid-term data on survival and on functional outcome.

In 2007 we reported the five-year results of our first 230 primary BHRs, in which the results were favourable. Our results continue to be satisfactory at ten years. The functional outcome scores were excellent and have remained stable over time; the UCLA activity rating indicated that patients remain active and have a high level of function and their survival rates were high, similar to the findings of Treacy et al.

In our study the confounding variables of both gender and head size could be shown to affect survival. This was further investigated using structural equation models to separate the effects of each variable independently. Treacy et al also found these factors to be significantly associated with revision and had the same Kaplan-Meier survival rates for men (98%) as our study. These results are reinforced by reg-
Neck narrowing that was previously reported in both this series and others was not associated with failure of any of the BHRs. This is consistent with other reports in the literature and supports the theory that most neck narrowing occurs as a consequence of early remodelling, and is followed by stabilisation.

This study confirms that the BHR provides a good outcome in male patients. The results suggest that for men aged <65 years resurfacing is a suitable option offering a high level of function and survival, exceeding that which is expected for THR in younger men at ten years. There is little evidence to support the use of the BHR outside of this defined group. The incidence of femoral lysis, which has not been previously reported, is of some concern and will need continued monitoring.

The authors wish to thank Professor A. Esterman and Dr S. Pearce for statistical analysis of the study data, and A. Pearce for assistance with the preparation of the manuscript. 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