HIP

Do trainee surgeons have an adverse effect on the outcome after total hip arthroplasty?

A TEN-YEAR REVIEW

M. J. Reidy, A. Faulkner, B. Shitole, B. Clift

From NHS Tayside, Ninewells Hospital, Dundee, United Kingdom

Aims

The long-term functional outcome of total hip arthroplasty (THA) performed by trainees is not known. A multicentre retrospective study of 879 THAs was undertaken to investigate any differences in outcome between those performed by trainee surgeons and consultants.

Patients and Methods

A total of 879 patients with a mean age of 69.5 years (37 to 94) were included in the study; 584 THAs (66.4%) were undertaken by consultants, 138 (15.7%) by junior trainees and 148 (16.8%) by senior trainees. Patients were scored using the Harris Hip Score (HHS) pre-operatively and at one, three, five, seven and ten years post-operatively. Surgical outcome, complications and survival were compared between groups. The effect of supervision was determined by comparing supervised and unsupervised trainees. A primary univariate analysis was used to select variables for inclusion in multivariate analysis.

Results

There was no evidence that the grade of the surgeon had a significant effect on the survival of the patients or the rate of revision (p = 0.987 and 0.405, respectively) up to 12 years post-operatively. There was no significant difference in post-operative functional HHS or total HHS among consultants, junior and seniors up to ten years post-operatively (p = 0.401 and 0.331), respectively. There was no significant difference in hospital stay (p = 0.855) between different grades of surgeons. There was no evidence that the level of supervision had an effect on the survival of the patients or the rate of revision (p = 0.837 and 0.203, respectively) up to 12 years post-operatively. There was no significant difference between supervised and unsupervised trainee groups in post-operative functional HHS or total HHS up to ten years post-operatively (p = 0.213 and 0.322, respectively). There was no significant difference in the mean hospital stay between supervised and unsupervised trainees (p = 0.908).

Take home message: This study suggests that when trainees are appropriately supervised, they can obtain results comparable with those of their consultant colleagues when performing THA.

Cite this article: Bone Joint J 2016;98-B:301–6.

Total hip arthroplasty (THA) is one of the most commonly performed and successful orthopaedic procedures.1 There is a learning curve and a need for more operating time in those cases performed by trainees.2-4 The long-term functional outcome of THA performed by trainees compared with that of consultants is not known, but it has been suggested that junior surgeons have a higher rate of complications.5-8 In one study, the rate of dislocation of THAs by inexperienced surgeons was double that of those performed by their more experienced colleagues.6 In 2013, Schoenfeld et al11 found that in 43 343 arthroplasties of large joints, resident involvement conferred a mild to moderate risk of complications. There is, however, conflicting evidence in the literature with other recent studies finding no increased risk.2,3,9-14 This issue is particularly topical with the advent of individualised surgeon data available online to patients via the National Joint Registry,15 as well as the potential for surgeon-level data to be used as a means to assess surgical performance.16 While surgeons have always aimed for excellent care of their patients, the desire to avoid complications and perceived poor performance grows, potentially resulting in fewer opportunities for trainees to operate.

Palan et al10 reported no significant difference in Oxford Hip Scores (OHS) or rates of complications between trainees and consultant surgeons at five years after operation. There
remains, however, a need to examine the overall effect of the experience of the operating surgeon on long-term outcomes. We aimed to investigate whether there is any significant difference in complications, function and survivorship up to 12 years post-operatively between operations performed by surgeons in training and those performed by consultants.

**Patients and Methods**

The study includes all elective primary THAs performed in the three centres within our region in 2003 to 2004. Ethical approval to access the regional arthroplasty database was obtained.

All 879 patients undergoing elective primary THA in Ninewells Hospital, Perth Royal Infirmary and Stracathro Hospital during this time were included. Patients undergoing revision THA were excluded from the study.

Data recorded included age, gender, body mass index (BMI), pre- and post-operative Harris Hip Score (HHS), grade of trainee, supervision status, surgical approach, length of stay, survival of the implant and complications including dislocation and infection. The operating surgeon was graded as junior trainee (SpR years 1 to 3 or ST3 to 5), senior trainee (SpR 4 to 6 or ST6 to 8) or consultant.

**Statistical analysis.** This was performed using Statistical Package for Social Sciences (SPSS) version 20.0 manufactured by IBM corporation, Armonk, New York for Windows.

Univariate analysis was used to identify variables for multivariate analysis of variance (MANOVA) to compare length of stay between groups, with p < 0.30 being regarded as significant. Cox regression was used to compare both patient and implant survival. As with length of stay, univariate analyses were used to select variables for inclusion in the model.

The functional component and total HHS were investigated with generalised estimating equations (GEE) assuming a gamma log-linked distribution with first order autoregressive correlation between follow-up scores.

Binomial logistic regression was used to compare the rates of dislocation and infection between consultants and trainees.

Confidence intervals (CI) of 95% were used in all analyses and factorial models were fitted, with backwards regression where appropriate. The value for statistical significance was set at p < 0.05

### Results

Loss to follow-up, mortality and revision surgery are detailed at Table I. The mean age of the patients was 69.5 years (37 to 94) with 532 female (60.5%) and 347 male (39.5%) patients. The mean BMI was 24.5 kg/m² (13.7 to 44.0). The primary diagnoses, for which THA was being performed, were osteoarthritis (833, 94.8%), rheumatoid arthritis (26, 3.0%) and avascular necrosis (19, 2.2%). The relative paucity of HHS data from year three was due to one participating centre excluding the three-year review.

A total of 584 patients (66.4%) were operated on by consultants, 138 (15.7%) by junior trainees and 148 (16.8%) by senior trainees. The grade of surgeon was not available for nine patients (1.1%) and these were excluded from further analysis.

Trainees were supervised in 241 operations (85%), ranging from 129 (93.5%) performed by junior trainees to 112 (76.0%) for senior trainees.

The anterolateral approach was used in 510 THAs performed by consultants (87.3%), 114 of those performed by junior trainees (82%) and 146 by senior trainees (98.6%). The posterior approach was used in 71 (12.5%), 24 (17.4%) and two (1.35%) operations, respectively. One THA performed by a consultant was described as ‘using previous incision’.

Details of the types of implant used are given in Table II. All femoral heads were either 22 mm or 28 mm other than the single resurfacing procedure.

**Length of stay.** The mean length of stay for THAs undertaken by consultants was 8.15 days (2 to 48), 7.99 days (3 to 24) for junior trainees and 8.26 days (3 to 27) for senior trainees. In univariate analysis (Table III), age, gender, pre-operative HHS function and HHS total scores reached the threshold of significance for inclusion in the multivariate model. There was also a loose association between the choice of implant and length of stay, however, this was not significant at the 5% level (p = 0.062). In multivariate analysis – only age was significant (p < 0.001).

Neither surgeon grade nor supervision had a significant effect on length of stay (p = 0.855 and 0.908, respectively) and there were no interactions with other dependent variables.
Patient survival. In univariate analysis, age; BMI; pre-operative HHS function and HHS scores; diagnosis; incidence of complications and length of stay reached significance for inclusion in the multivariate model. In multivariate analysis, age; gender; diagnosis; incidence of complications and length of stay were significant for the survival of the patients (Table IV). There was some evidence that the surgical approach was associated but this finding was not significant (p = 0.065). Neither the grade of surgeon nor supervision had a statistical effect on survival (p = 0.987 and 0.837, respectively) and there were no interactions with other dependent variables.

Implant survival. A total of 28 patients required revision surgery within the study period; 21 THAs had been performed by consultants, four by junior trainees and three by senior trainees (Fig. 1). The primary surgery in the seven patients operated on by trainees had been supervised while none of the 44 THAs performed by unsupervised trainees required revision.

In univariate analysis, gender, surgical approach, incidence of complications and primary pathology reached significance for inclusion in the multivariate model. In multivariate analysis, gender and the incidence of complications were significant in survival to revision (Table V). Neither the grade of surgeon nor supervision had an effect on survival to revision (p = 0.405 and 0.203, respectively) and there were no interactions with other dependent variables.

GEE analysis of post-op function and score. The HHSs are shown in (Table VI). In univariate analysis, age, BMI, gender, the grade of surgeon, implant, surgical approach, pre-operative scores and incidence of complications reached the significance for inclusion in the multivariate model (Table VII). In multivariate analysis, these factors remained significant with the exception of the grade of surgeon. Supervision was not significant in either outcome. There were no interactions between the grade of the surgeon or supervision and other dependent variables.

Complications and dislocations. A total of 13 patients (1.5%) developed a deep infection: eight had been operated on by consultants, three by junior trainees and two by senior trainees. The surgery in the five patients in the trainee group was supervised.

A total of 22 patients (3.0%) had a dislocation, 17 of whom had been operated on by consultants, two by junior trainees and seven by senior trainees. Of the nine dislocations that occurred in all trainee-led surgery, eight (89.9%) were in the supervised group and one (10.1%) in the unsupervised group.

Age and length of stay reached significance for inclusion in the multivariate model (Table VIII). In multivariate analysis, only length of stay was significant. Neither the
grade of the surgeon, nor supervision, was associated with the likelihood of developing complications, either as main effects or interactions with other dependent variables.

In univariate and multivariate analyses, there were no significant factors in the development of dislocation.

In our study, there was no statistical association between the grade of surgeon and mortality (p = 0.987) or the survival of the implant up to 12 years post-operatively (p = 0.405). Our findings are in keeping with the shorter term findings of Palan et al.10 who found no significant differences in OHS or the rate of infection or complications up to five years post-operatively.

No significant differences were observed in the rate of dislocation between the consultants and trainees in this study. Moreover, we found no significant difference between junior and senior trainees. This is not in keeping with the findings of Hedlundh et al,6 which suggested inexperienced surgeons had twice the number of dislocations as more experienced surgeons. Importantly, the level of supervision of less experienced surgeons in this study was not described. In a more contemporaneous study that may better represent current training behaviours, Moran et al9 found no difference in the rates of dislocation between trainees and consultants in short-term follow up.

The way that training has evolved over the past number of decades as led to a move away from early independent operating by trainees to a more structured and supervised environment.18 This may, in part, account for the differences seen between older research and more contemporary studies.

We found no significant differences between the groups of surgeon for the HHSs pre-operatively and up to ten years post-operatively. Moran et al9 found the same up to six and 18 months post-operatively (p = 0.990 and p = 0.670, respectively). We would suggest that whilst a difference in outcome is not demonstrated between the supervised and unsupervised trainees, caution should be exercised in drawing firm conclusions from this section of the analysis due to the low number of patients in the unsupervised group.

Among the 138 THAs undertaken by junior trainees, 129 (94%) were supervised and nine were unsupervised (6%), a level which may be expected in this group. Among the 148 undertaken by senior trainees, 113 were supervised...
and 35 were unsupervised (24%). This high level of supervision is of interest and may reflect the local need for a consultant to assist due to a lack of other available assistants. Differentiating this scenario from those trainees who were judged to require supervision, however, is not possible.

One of the strengths of this study is the subdivision of trainee-led THAs, something that has not been investigated in most previous studies. This study has limitations. Patients were not randomised to different groups and consultants may have chosen to operate on more difficult cases and those patients with a higher American Society of Anesthesiologists grade, although these data are unfortunately not available. Radiological follow-up would also have been advantageous but, due to the number of different combinations of implants used, fair comparison would be challenging and open to interpretation, and would feature small numbers in some groups. Since the period of the study, however, our trust has rationalised the number of implants used for THA, and therefore, a radiological assessment would be more achievable in the future. The addition of such a parameter would make it possible to assess the quality of cementation and the alignment of the components, which would permit a more comprehensive assessment of surgical performance.

It is also worth noting that this study began before the introduction of The European Working Time Directive (EWTD). It has been suggested that this has had a particularly deleterious effect on surgical training and it would...

| Table V. Significance of independent variables in implant survival, with hazard ratio (HR) and upper and lower 95% confidence intervals |
|-----------------------------|------------------|------------------|------------------|------------------|------------------|
| Independent variable | Univariate analysis | Multivariate analysis |
| | HR | Lower | Upper | p-value | HR | Lower | Upper | p-value |
| Age | 0.97 | 0.94 | 1.01 | 0.76 | 0.97 | 0.94 | 1.03 | 0.175 |
| Body mass index | 0.94 | 0.98 | 1.11 | 0.241 | 1.04 | 0.95 | 1.13 | 0.452 |
| Gender (F:M) | 0.36 | 0.16 | 0.77 | 0.009 | 0.44 | 0.19 | 0.99 | 0.047 |
| Surgeon grade | 0.381 | 0.405 |
| Supervision (N:Y) | 0.96 | 0.42 | 2.19 | 0.928 | 0.26 | 0.03 | 2.08 | 0.203 |
| Approach | 0.448 | 0.988 |
| Harris Hip Score (HHS) pain pre-op | 0.041 | 0.109 |
| HHS function pre-op | 0.734 | 0.610 |
| HHS pre-op | 1.00 | 0.98 | 1.20 | 0.761 | 0.99 | 0.97 | 1.02 | 0.629 |
| Complications (Y:N) | 6.51 | 2.94 | 14.4 | <0.001 | 6.68 | 2.87 | 16.1 | <0.001 |
| Diagnosis | 0.15 | 0.185 |
| Length of stay | 1.02 | 0.98 | 1.06 | 0.327 | 1.02 | 0.96 | 1.07 | 0.548 |

Table VI. Harris Hip Score (HHS) by surgeon grade throughout the study period.

<table>
<thead>
<tr>
<th>Consultants</th>
<th>Junior trainees</th>
<th>Senior trainees</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>HHS pre-op</td>
<td>519</td>
<td>42.55 (11.32)</td>
</tr>
<tr>
<td>HHS yr 1</td>
<td>505</td>
<td>87.39 (11.45)</td>
</tr>
<tr>
<td>HHS yr 3</td>
<td>267</td>
<td>86.49 (12.22)</td>
</tr>
<tr>
<td>HHS yr 5</td>
<td>392</td>
<td>87.06 (12.67)</td>
</tr>
<tr>
<td>HHS yr 7</td>
<td>80</td>
<td>85.02 (12.58)</td>
</tr>
<tr>
<td>HHS yr 10</td>
<td>181</td>
<td>85.29 (12.90)</td>
</tr>
</tbody>
</table>

Table VII. Factors in post-operative Harris Hip Scores (HHS) over ten years.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Function</td>
<td>Score</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.013</td>
<td>0.032</td>
</tr>
<tr>
<td>Gender (F:M)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Surgeon grade</td>
<td>0.030</td>
<td>0.036</td>
</tr>
<tr>
<td>Supervision (N:Y)</td>
<td>0.140</td>
<td>0.174</td>
</tr>
<tr>
<td>Implant</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Approach</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HHS function (pre-op)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HHS pre-op</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Complications (Y:N)</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>0.192</td>
<td>0.101</td>
</tr>
</tbody>
</table>
Table VIII. Factors in the development of complications, with odds ratio (OR) and upper and lower 95% confidence intervals

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Univariate analysis</th>
<th>Multivariate analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR</td>
<td>Upper</td>
</tr>
<tr>
<td>Age</td>
<td>1.03</td>
<td>1.01</td>
</tr>
<tr>
<td>Body mass index</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>Gender (F:M)</td>
<td>0.88</td>
<td>0.85</td>
</tr>
<tr>
<td>Surgeon grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supervision (N’Y)</td>
<td>1.21</td>
<td>0.87</td>
</tr>
<tr>
<td>Implant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris Hip Score (HHS) pain (pre-op)</td>
<td>1.00</td>
<td>0.98</td>
</tr>
<tr>
<td>HHS function (pre-op)</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>Diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of stay</td>
<td>1.14</td>
<td>1.09</td>
</tr>
</tbody>
</table>

be of interest to carry out a comparative analysis with a post-EWTD cohort in order to evaluate any effect of reduced training time on patient outcomes.

While the results suggest that trainers are capable of judging when a trainee can operate competently without supervision, a firm conclusion cannot be drawn from the relatively small numbers in this subgroup. The study does show, however, that trainees can perform THA safely and effectively, achieving comparable long-term clinical results with their consultants whilst under appropriate supervision.

Author contributions:
M. J. Reidy: Design, data collection, results analysis, write-up.
A. Faulkner: Design, data collection, results analysis, write-up.
B. Shitole: Design, data collection, results analysis, write-up.
B. Clift: Design, data analysis, write-up.

The authors would like to thank the TAAG staff and L. Cochrane for her help and advice with the statistical analysis.

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.

This article was primary edited by P. Page and first proof edited by J. Scott.

References