A systematic review and meta-analysis of trainee- versus consultant surgeon-performed elective total hip arthroplasty

Prashant Singh¹
Suroosh Madanipour²
Andreas Fontalis³
Jagmeet Singh Bhamra⁴
Hani B. Abdul-Jabar¹

Introduction

Total hip arthroplasty (THA) is one of the most commonly performed orthopaedic procedures worldwide. Demand for THA is set to increase over the next decade with a 174% increase in the USA estimated by 2030,¹,² largely due to longer life expectancy, an ageing population and higher functional demand in the developed world.³ Training future orthopaedic surgeons is clearly crucial if supply is to meet demand. However, such training raises several conflicting issues. Although trainees must have adequate operative experience to be deemed competent in a procedure, this must be balanced with increasing scrutiny of outcomes, the widespread use of joint registries and pressure on surgical teams to maximize efficiency. Concerns have been raised that trainee-performed operations may be associated with poorer outcomes for patients, a reduction in efficiency and a consequent rise in healthcare delivery costs.⁴⁻⁸ Schoenfeld et al conducted a retrospective review of outcomes using registry data for 12 commonly performed orthopaedic operations, noting a mild to moderate risk of complications for operations involving a resident.⁵ Similarly, Marston et al compared outcomes following trainee- and consultant-performed THA, noting a higher revision rate among trainee-performed procedures.⁹ The growing use of validated simulation packages for training has helped to familiarize trainees with orthopaedic procedures and offset the learning curve for THA.¹⁰ However, there is no substitute for gaining competence by repetitively undertaking a procedure.¹¹

Keywords: total hip arthroplasty; surgical training; supervision
The aim of the present study is to compare outcomes following trainee- and consultant-performed THA by analysing the existing evidence relating to this important question.

Methods

Study selection

This study was registered with the PROSPERO international database of systematic reviews (CRD42018086012) and followed the Preferred Reporting Items of Systematic Reviews and Meta-analyses (PRISMA) guidelines. A systematic search of all published literature was performed using The Medical Literature Analysis and Retrieval System Online (MEDLINE via PubMed), Excerpta Medica (EMBASE), the Ovid database, Google and Google Scholar, and the Cochrane Controlled Trials Register. The following items were used for the search, both alone and in various combinations: ‘total hip arthroplasty’; ‘total hip replacement’; ‘resident training’; ‘resident performed’; ‘trainee’; ‘trainee performed’; ‘outcomes’; ‘trainee lead’; and ‘functional outcome’. The ‘related articles’ function in PubMed was used to widen the search. The titles, abstracts and citations resulting from each search were systematically scanned by the authors and assessed for inclusion. A minimum of two authors conducted a manual search of all references. Reviewers independently assessed full texts to determine whether the study met our inclusion criteria. Date limits were between January 2000 and October 2017. No language restrictions were imposed.

Inclusion and exclusion criteria

The following inclusion criteria were applied:

1) compared outcomes following trainee-performed and consultant-performed THA in an elective setting;
2) trainees must have performed part of or the whole procedure. Studies which reported outcomes for resident involvement in the procedure only (i.e. participated as an assistant but did not perform part of or the whole procedure) were excluded. A consultant was defined as a board-certified, independently operating surgeon;
3) randomized controlled trial, prospective observational or retrospective study;
4) reports data on at least one of the primary or secondary outcomes;
5) a minimum sample size of ten patients in each group.

Studies were excluded from the analysis if:

1) extractable data relating to any of the outcomes were not available;
2) the study reported outcomes for operations involving trainees as assistants;
3) the study did not directly compare outcomes for consultant- and trainee-performed THA;
4) the study was a review article, correspondence or conference abstract.

Data extraction and outcomes

Two authors independently identified studies for inclusion and extracted data for the outcomes. Discrepancies in extraction of data were resolved by re-examination of the literature until consensus was achieved. A standardized data extraction spreadsheet was used by each of the reviewers to ensure consistency in the method of data extraction. The primary outcome for the meta-analysis was the rate of revision surgery following THA. Secondary outcomes included the rate of dislocation, rate of deep tissue infection, mean operation time, length of hospital stay and Harris Hip Score (HSS) up to one year. Where meta-analysis was not possible due to insufficient data or unacceptable clinical heterogeneity, a qualitative analysis and review of the available data were performed.

Statistical analysis

Dichotomous variables were compared using odds ratios (ORs) with 95% confidence intervals (CI). The OR was defined as the probability of an event occurring in the trainee group compared with the consultant group. Continuous variables were compared using weighted mean differences (WMD) with 95% CIs. The Mantel–Haenszel method was used for the meta-analysis. Heterogeneity was assessed using the chi-squared (X²) test, with p < 0.050 being regarded as significant. The I² statistic was also used, with < 50% being regarded as a low degree of heterogeneity. In such cases, a fixed effects (FE) model of meta-analysis was used. Studies with I² > 50% were considered to be associated with a high degree of heterogeneity and a random effects model was applied. Random effects models assume that variation in effect size between studies exists, and account for differences in study population, co-morbidities and surgical protocol which would otherwise lead to a significant risk of bias.

Statistical analysis was performed using Review Manager 5.3 (The Cochrane Collaboration, Copenhagen, Denmark).

Quality assessment and sensitivity analysis

A quality assessment of all studies was performed using the Newcastle-Ottawa Scale (NOS) for non-randomized studies. High-quality studies were defined as scoring ≥ 7 on the NOS. A planned sensitivity analysis was performed for high-quality studies and a separate subgroup analysis
was performed comparing outcomes for supervised trainees versus consultants.

**Results**

**Eligible studies**

After exclusions, seven studies were included in the final analysis (Fig. 1), involving 40,810 THAs, of which 6,393 (15.7%) were performed by trainee surgeons and 34,417 (84.1%) were performed by consultant surgeons (Table 1). Two of the included studies were prospective, non-randomized studies and five studies were retrospective. A total of 5,651 (88.4%) THAs within the trainee group were performed under supervision of consultant surgeons. One study reported outcomes following minimally invasive THA.15 Dates of publication for included studies ranged from 2004 to 2017. All studies were performed in developed-world settings, with four British, one German, one North American and two Australasian studies. The mean age of patients was in the range of 65.9 to 70.0 years for the trainee group compared with 63.8 to 70.0 years for the consultant group (Table 2). In one study, trainees operated on significantly older patients when compared with consultants (70 years versus 66 years, respectively; p < 0.01).16 In the remaining studies, there was no significant difference in baseline age or ASA grade between the two groups. The mean follow-up interval for the primary outcome was 42 months (Table 1).

Meta-analysis was conducted for the following outcomes: rate of revision surgery; rate of dislocation; rate of deep infection; operation time; length of hospital stay; and HHS at six months. Two studies defined supervision as having the consultant scrubbed with the trainee and acting as first assistant.16,17 The remaining studies did not give a clear definition of supervision (Table 3). Although the definition of a trainee varied between studies, all were on a recognized training programme for orthopaedic surgery. Two studies separated trainees into junior and senior groups with 40.7% (215/528)18 and 48.3% (138/286)19 being assigned to the junior trainee groups. There was wide variation in implant choice and surgical technique between studies, described in Table 4. Woolson et al compared operations done in the private (by consultants) and public (by trainees) sectors. While the components used were the same, there was a difference in the utilization of screw fixation for the acetabular component.
Regarding approaches, one study reported that both consultants and trainees used the anterolateral approach for all cases. One study reported on the anterolateral minimally invasive approach, again for both consultant and trainee groups in all cases. In the studies where differing approaches were used, the anterolateral approach predominated (Table 4). Of the data available in the three studies where differing approaches were employed, consultants used the anterolateral approach in 57% (1601/2797) of cases and the posterior approach in 31% (871/2791) of cases. Trainees used the anterolateral approach in 72.6% (1215/1697) of cases and the posterior approach in 28% (476/1697) of cases. One of these studies also reported that 4% (52/1240) of consultants used the anterior approach versus 0.7% (6/1032) of trainees.

**Risk of bias assessment**
Each study was assessed for quality using the NOS tool for non-randomized studies. Studies were deemed to be of high quality if they scored ≥ 7 out of 9 points on the NOS. Results of the quality assessment are displayed in Table 5. Five out of seven studies were deemed to be of high quality.

**Comparison of outcomes**
Comparison of rate of revision surgery
Data from five studies were included in the meta-analysis for the rate of revision surgery. There was no difference in

---

**Table 1. Characteristics of included studies**

| References | Year | Journal | Location | Study type | Comparison | Trainee (n) | Consultant (n) | FU/months | Outcome measures |
|------------|------|---------|----------|------------|------------|------------|---------------|-----------|------------------|
| Inglis 2012 | 2012 | Journal of Surgical Education | New Zealand | Retrospective, registry data | Consultant vs supervised or unsupervised trainee vs consultant | 4049 | 30344 | 72 | Revision rates; OHS |
| Moran 2004 | 2004 | Journal of Arthroplasty | UK | Prospective, non-randomized | Supervised trainee vs consultant | 139 | 397 | 18 | HHS at 6 and 18 months; blood loss; transfusion requirements; revision; dislocation; death; acetabular and femoral component alignment; cementation |
| Palan 2009 | 2009 | BJJ | UK | Prospective, non-randomized | Trainer vs trainee | 528 | 973 | 60 | Revision rate; change in OHSS; dislocation; length of stay; operating time |
| Reidy 2016 | 2016 | BJJ | UK | Retrospective | Consultant vs senior or junior trainee | 286 | 584 | 144 | HHS; dislocation; length of stay; deep infection |
| Weber 2017 | 2017 | Nature Scientific Reports | Germany | Retrospective, registry data | Senior surgeon vs supervised trainee | 240 | 768 | 12 | EQ-SD Score; WOMAC score; dislocation; operating time; deep infection; intraoperative fracture |
| Wilson 2016 | 2016 | BJJS | Australia | Retrospective | Consultant vs trainee, junior vs senior trainee | 1032 | 1240 | 12 | Surgical complication; medical complication; wound complication; transfusion; readmission |
| Woolson 2007 | 2007 | JBJS | USA | Retrospective | Supervised trainee vs consultant | 119 | 111 | 49 | HHS; length of stay; operating time; units of transfusion required; estimated blood loss; femoral component alignment: varus/neural/varus; femoral component fit: good/fair/poor |

---

**Table 2. Study demographics**

| References | Age | Gender Male/Total | ASA | Notes |
|------------|-----|-------------------|-----|-------|
| Inglis 2012 | N/A | N/A | 50/139 | Trainees operated on significantly older patients (p < 0.001) |
| Moran 2004 | Trainee | 70 | 9 | 50/139 |
| Consultant | 66 | 11 | 155/397 |
| Palan 2009 | Trainee | 68.8 | 17 | N/A |
| Consultant | 68 | 18 | N/A |
| Reidy 2016 | N/A | N/A | 125 | No significant difference in age, gender, ASA on multivariate analysis |
| Weber 2017 | Trainee | 65.9 | 10.1 | 121/240 |
| Consultant | 63.8 | 10.8 | 365/768 |
| Wilson 2016 | Trainee | 69 | 3.25 | 406/1032 |
| Consultant | 70 | 3.5 | 517/1240 |
| Woolson 2007 | N/A | N/A | 698 | No significant difference in age, gender, ASA on multivariate analysis |
Comparison of rate of deep infection
The rate of deep infection was reported by five studies. There was no difference in the infection rate between consultants and trainees (OR 1.49; 95% CI 0.93 to 2.41; \( p = 0.10 \), Table 6, Fig. 2a). These results were associated with low heterogeneity (I\(^2\) = 0%).

Comparison of rate of dislocation
Six studies compared rates of dislocation following THA between the two groups. There was no overall difference in dislocation rate (OR 0.96; 95% CI 1.76 to 1.67; \( p = 0.10 \), Table 3. Description of supervision characteristics for included studies

| Study      | Trainee group | Definition of trainee | Level of involvement | Definition of supervision | Definition of senior trainee | Total | Supervised | Unsupervised | Junior | Senior |
|------------|---------------|-----------------------|----------------------|---------------------------|----------------------------|-------|-------------|--------------|-------|--------|
| Inglis     | Unclear       | Performed / primary operator | Unknown             | N/A                       | 4049                       | 2982  | 1067        |              |       |        |
| Moran      | Year 1 to 4 registrar (UK) | Performed / primary operator | Trainer scrubbed and acting as first assistant. Intervenes if trainee about to make a critical mistake that could jeopardize the final outcome | N/A                       | 139                        | 139   | N/A         |              |       |        |
| Palan      | All grades registrar (UK) | Performed / primary operator | Unknown             | Post FRCS Exam Year 4 to 6 registrar | 528                        | 528   | N/A         |              | 215   | 148    |
| Reidy      | Year 1 to 6 registrar (UK) | Performed whole procedure / primary operator | Unclear             | N/A                       | 286                        | 241   | 44          |              | 138   | 148    |
| Weber      | Year 3 to 5 of surgical training (Germany) | Performed / primary operator | Unclear             | N/A                       | 240                        | 240   | N/A         |              |       |        |
| Wilson     | Unclear       | Performed / primary operator | Unclear             | N/A                       | 1032                       | Unclear | Unclear  |              |       |        |
| Woolson    | Resident or joint replacement fellow (USA) | Performed either femoral or acetabular component, other performed by attending | Attending present for entire procedure on trainee side of table | N/A                       | 109                        | 109   | N/A         |              |       |        |

Table 4. Surgical techniques

| References | Implant | Femoral component | Acetabular component | Head | Acetabular screw fixation | Approach | Consultants | Trainees |
|------------|---------|-------------------|----------------------|------|---------------------------|----------|-------------|----------|
| Inglis     | N/A     | N/A               | N/A                  | N/A  | N/A                       | Anterolateral | Anterolateral |         |
| Moran      | Cemented Charnley (De Puy) | Cemented Charnley (De Puy) | Cementless and cemented Exeter and Charnley | Anterolateral 57% (402) | Anterolateral 77% (291) |
| Palan      | Cemented Exeter (Stryker) | Exeter 285, Charnley 209 CPT 190 Aesculap 103 Lubinus 50 | Exeter polyethylene cup 160 Trilogy 149 Aesculap 102 | Not recorded 316 Exeter polyethylene cup 160 Trilogy 149 Aesculap 102 | Anterolateral 43% (301) | Anterolateral 23% (88) |
| Reidy      | Exeter 285, Charnley 209 CPT 190 Aesculap 103 Lubinus 50 | Mayo 21 ABG I 17 Biomet 2 ABG II 1 Birmingham Resurfacing 1 | ZCA 78 Ogee 38 ABG II 15 TOP 13 Charnley Elite plus Ogee 5 Birmingham 1 ABG 1 Charnley LPW 1 | Not recorded 316 Stainless steel 335 Cobalt chrome 167 Ceramic 21 | Anterolateral 88% (510) | Anterolateral 91% (260) |
| Weber      | Uncemented Corail Uncemented Trilock (Depuy) | Pinnacle Cup (Depuy) | Pinnacle Cup (Depuy) | Minimally invasive anterolateral | Minimally invasive anterolateral |
| Wilson     | Uncemented | Uncemented | Uncemented | Consultant group 3/111 | Trainee group 98/119 |              |              |         |
| Woolson    | Uncemented AML and Replica | Uncemented Duraloc (Depuy) | Uncemented Duraloc (Depuy) | Anterolateral 56% (689) | Anterolateral 64.3% (664) |
|            | Cemented Endurance (Depuy) | | | Anterior 4% (52) | Anterior 35% (362) |
|            |         | | | Anterior 0.7% (6) | Anterior 0.7% (6) |
Fig. 4). These results were associated with low heterogeneity ($I^2 = 0\%$).

Comparison of operation time
Three studies reported the mean operation time taken to complete THA for trainees versus consultants. Using a random effects model of meta-analysis, the consultant group was associated with lower mean operation times (WMD 12.9 minutes; 95% CI 6.63 to 19.17; $p < 0.01$, Fig. 5). There was a high degree of heterogeneity associated with this result ($I^2 = 78\%$).

Comparison of length of stay
Three studies compared the results for length of stay between the two groups. There was no difference in length of stay for THA patients when trainees or consultants...
performed their operations (WMD -0.03; 95% CI -0.45 to 0.48; p = 0.92, Fig. 6). There was low heterogeneity associated with this result (I² = 0%).

Comparison of Harris Hip Score at six months
The HHS at six months following THA was reported by three authors. No significant difference was observed between the two groups (WMD -0.29; 95% CI -2.53 to 1.95; p < 0.80, Fig. 7). There was high heterogeneity associated with this result (I² = 71%).

Sensitivity analysis

Supervised trainees versus consultants
In the subgroup analysis for supervised trainees versus consultants, the trainee group was associated with a longer operation time (two studies, WMD 9.48; 95% CI
trainee- versus consultant surgeon-performed elective THA

6.33 to 12.62; p < 0.01; I² = 0, Table 6, Fig. 2b). There was no significant difference between the two groups for rate of revision, rate of dislocation or rate of deep infection.

Analysis of high-quality studies

In the sensitivity analysis of high-quality studies, there was no significant difference in rate of revision, rate of dislocation, rate of infection and length of stay between the two groups. Once again, trainees were associated with a longer operation time which was more pronounced than in the overall analysis (two studies; WMD 13.68; 95% CI 3.69 to 23.66; p < 0.01) (Table 7). Trainees were also associated with a less favourable HHS at six months (two studies; WMD -1.61; 95% CI -2.49 to -0.72; p < 0.01). In the high-quality analysis of supervised trainees versus consultants, there was no difference in rate of revision, rate of dislocation or rate of infection.

Qualitative analysis of functional outcome

Oxford Hip Score

Two studies assessed functional outcome using the Oxford Hip Score (OHS). Inglis et al noted a significantly superior OHS for consultants compared with supervised trainees (40.7 vs 38.95; p < 0.001) (Table 8) and unsupervised trainees (38.27; p < 0.001). There was no significant difference in OHS between supervised and unsupervised groups. Palan et al measured mean change in OHS pre- and post-operatively for trainee and

| Study or Subgroup | Trainee Mean | SD | Total | Expert Mean | SD | Total | Mean Difference IV, Random, 95% CI | Mean Difference IV, Fixed, 95% CI |
|-------------------|--------------|----|-------|-------------|----|-------|-----------------------------------|----------------------------------|
| Palan 2009        | 104          | 50 | 528   | 85          | 56.5| 973   | 19.0 [13.45, 24.55]               |                                  |
| Weber 2017        | 78.1         | 25.4| 240   | 69.3        | 23.8| 768   | 8.80 [5.17, 12.43]                |                                  |
| Woolson 2007      | 72.5         | 22  | 119   | 61          | 26.25| 111   | 11.50 [5.22, 17.78]              |                                  |
| Total (95% CI)    | 387          | 887| 1852  | 100.0%      |     |       | 12.90 [6.63, 19.17]               |                                  |
| Heterogeneity:    | TAU² =23.72; | Chi² = 9.10, df = 2 (P = 0.01); I² = 78% Test for overall effect: Z = 4.03 (P = 0.0001) |                                  |                                  |

Fig. 4 Forest plot of meta-analysis of rate of dislocation for trainee- versus consultant-performed THA.

| Study or Subgroup | Trainee Mean | SD | Total | Expert Mean | SD | Total | Mean Difference IV, Fixed, 95% CI | Mean Difference IV, Fixed, 95% CI |
|-------------------|--------------|----|-------|-------------|----|-------|-----------------------------------|----------------------------------|
| Palan 2009        | 11           | 27.7| 528   | 9.8         | 10.2| 973   | 1.20 [–1.25, 3.65]                |                                  |
| Weber 2017        | 8.13         | 5.65| 286   | 8.15        | 11.5| 584   | –0.02 [–1.16, 1.12]               |                                  |
| Woolson 2007      | 5.1          | 1   | 119   | 61          | 3   | 111   | –0.10 [–0.69, 0.49]               |                                  |
| Total (95% CI)    | 933          | 1668| 1668  | 100.0%      |     |       | –0.03 [–0.54, 0.48]               |                                  |
| Heterogeneity:    | Chi² = 1.02, df = 2 (P = 0.60); I² = 0% Test for overall effect: Z = 0.11 (P = 0.92) |                                  |                                  |

Fig. 5 Forest plot for meta-analysis of operation time for trainee- versus consultant-performed THA.

| Study or Subgroup | Trainee Mean | SD | Total | Expert Mean | SD | Total | Mean Difference IV, Random, 95% CI | Mean Difference IV, Random, 95% CI |
|-------------------|--------------|----|-------|-------------|----|-------|-----------------------------------|----------------------------------|
| Palan 2012        | 30           | 4049| 147   | 30344       | 50.2% | 1.53 [1.03, 2.27] |                                  |                                  |
| Moran 2004        | 2            | 139 | 6     | 397         | 2.5% | 0.81 [0.17, 3.96] |                                  |                                  |
| Palan 2009        | 7            | 528 | 13    | 973         | 7.5% | 1.11 [0.48, 2.55] |                                  |                                  |
| Reidy 2016        | 7            | 286 | 21    | 584         | 11.2%| 1.08 [0.48, 2.46] |                                  |                                  |
| Weber 2017        | 2            | 119 | 3     | 111         | 2.5% | 0.29 [0.02, 5.24] |                                  |                                  |
| Woolson 2007      | 2            | 119 | 3     | 111         | 2.5% | 1.36 [0.50, 3.71] |                                  |                                  |
| Total (95% CI)    | 5361         | 33177| 100.0%| 1.30 [0.96, 1.76] |                                  |                                  |
| Total events      | 60           | 198 |       |                                  |                                  |                                  |
| Heterogeneity:    | Chi² = 2.39, df = 5 (P = 0.79); I² = 0% Test for overall effect: Z = 1.67 (P = 0.10) |                                  |                                  |

Fig. 6 Forest plot for meta-analysis of length of stay for trainee- versus consultant-performed THA.
consultant groups (Table 8). No significant difference in change of OHS was noted between trainees and consultants at three-month follow-up as well as at the one-, two-, three-, four- and five-year follow-ups. The consultant group was noted to have a higher post-operative OHS at five years compared with trainees (40.5 vs 39.2; \( p = 0.02 \)), but the consultant group had a higher pre-operative OHS.

When comparing senior and junior trainee groups, Palan reported no difference in change in OHS between the two groups at three months, two years and three years, but a significantly better change in OHS for senior trainees at one, four and five years. Meta-analysis of long-term functional outcome was not performed due to variation in type of data reported (i.e. change in OHS versus OHS) and variation of follow-up intervals between studies.

**Harris Hip Score**

Three studies compared HHS between consultants and trainees at various follow-up intervals.\(^ \text{16,17,19} \) Two of these studies were included in the meta-analysis of HHS scores at six months.\(^ \text{16,17} \) Reidy et al found no significant difference between trainees and consultants both pre-operatively and at one-, three-, five-, seven- and ten-year follow-up. There was a significant loss to follow-up in the analysis at seven and ten years. Moran et al reported no significant differences in HHS at 18 months between the two groups (84.6 vs 86.4).

**Other outcomes**

Weber et al reported no difference in WOMAC score and EQ-50 scores between trainee and consultants at one-year follow-up.\(^ \text{15} \) Moran et al noted a significant difference in cup anteversion between trainees and consultants.\(^ \text{16} \) Trainee-sited cups were in 6.1° anteversion compared with 11.4° for consultants, when measured using Pradhan’s technique (\( p < 0.001 \)). Other outcomes were no different between the two groups.

---

### Table 7. Results of meta-analysis of outcomes for trainee- versus consultant-performed THA – sensitivity analysis for high-quality studies

| Outcome (dichotomous) | 95% CI | Heterogeneity |
|------------------------|--------|---------------|
| Revision rate          | OR     | Lower Upper   | \( \chi^2 \) | \( p \) | \( I^2 \) | FE/RE |
| 3                      | 0.82   | 0.46 1.47     | 0.5       | 0.82   | 0   | FE   |
| 3                      | 1.24   | 0.55 2.79     | 0.6       | 0.91   | 0   | FE   |
| 4                      | 0.98   | 0.58 1.67     | 0.94      | 0.83   | 0   | FE   |

| Outcome (continuous)   | No studies | WMD | Lower Upper | \( p \) | \( \chi^2 \) | \( p \) | \( I^2 \) | FE/RE |
|------------------------|------------|-----|-------------|-------|-------------|-------|-------|-------|
| Operation time         | 2          | 13.68 | 3.69 23.66  | 0.007 | 9.09        | 0.003 | 89    | RE    |
| Length of stay         | 2          | 0.2  | -0.384 1.23 | 0.71  | 0.78        | 0.38  | 0     | FE    |
| HHS                    | 2          | -1.61 | -2.49 -0.72 | \( <0.01 \) | 1.37        | 0.24  | 27    | FE    |

**Supervised trainees vs consultants**

| Outcome (dichotomous) | 95% CI | Heterogeneity |
|------------------------|--------|---------------|
| Revision rate          | OR     | Lower Upper   | \( \chi^2 \) | \( p \) | \( I^2 \) | FE/RE |
| 2                      | 0.83   | 0.39 1.79     | 0.64      | 0.03   | 0   | FE   |
| Dislocation rate       | 3      | 0.93 | 0.47 1.9     | 0.84    | 0.88  | 0.64 | 0   | FE    |
| Deep infection rate    | 2      | 1.09 | 0.34 3.43    | 0.89    | 0.33  | 0.57 | 0   | FE    |

| Outcome (continuous)   | No studies | WMD | Lower Upper | \( p \) | \( \chi^2 \) | \( p \) | \( I^2 \) | FE/RE |
|------------------------|------------|-----|-------------|-------|-------------|-------|-------|-------|
| Revision rate          | 2          | 0.83 | 0.39 1.79   | 0.64  | 0.03        | 0.85  | 0     | FE    |
| Dislocation rate       | 3          | 0.93 | 0.47 1.9    | 0.84  | 0.88        | 0.64  | 0     | FE    |
| Deep infection rate    | 2          | 1.09 | 0.34 3.43   | 0.89  | 0.33        | 0.57  | 0     | FE    |

HHS = Harris Hip Score; OR = odds ratio; FE = fixed effects model; RE = random effects model; WMD = weighted mean difference.
Discussion

The present study has demonstrated no difference in the rate of revision surgery, the rate of deep infection and the rate of dislocation when trainees perform THA compared with consultant surgeons. We can infer from this that trainees are safe to operate in selected cases under supervision from consultants. These results are reassuring for both patients and trainers. Demand for THA is forecast to rise significantly over the next decade.1 Orthopaedic trainees must be sufficiently competent to operate independently as consultants to provide a safe service for patients and must become competent in a variety of technical skills and procedures. Changes to training over the last decade include the increasing use of simulation, and in some settings, a move towards a competency-based rather than time-based training model.22-24 There is no real substitute for the opportunity to perform operations in a real-life setting. The findings of this paper demonstrate that the ‘apprenticeship model’ of surgical training is safe and should be maintained.

This study has shown that trainees take significantly longer to perform THA compared with consultants. While some service providers may have reservations about efficiency, the data show that training need not hamper efficiency to a large extent. The difference of 13 minutes represents the learning curve of the trainee, and while not insignificant, is a relatively short time and an acceptable ‘cost’ of training. There are estimates in the literature regarding the added cost associated with trainee-performed surgery.25 Weber et al estimated an additional $33,000 for 230 THAs performed by trainees, which equates to $140 per case.15 Clearly, this is a crude estimate and is likely to vary significantly; however, it does provide a rough indication of the low cost. A recent BMJ paper, evaluating the cost of operating theatre time per hour, estimated this as approximately £1200 per hour, equating to £240 for 13 minutes.26

Interestingly, our subgroup analysis for supervised trainees versus consultants showed no significant difference between the two groups, compared with the overall analysis. This is likely attributable to a faster intra-operative decision-making process with a consultant present. The unsupervised trainee may be slower to deal with unexpected steps whereas consultant presence, even when a trainee is performing the procedure, keeps the operation moving.

With regards to our secondary outcomes, we found no difference in rates of deep infection. In fact, overall rates of infection were low in both groups, in the range of 0.9% to 1.7% for the trainee group and 0.2% to 1.7% for the consultant group; infection rates across all studies were within limits accepted in the literature.27 It is possible that rate is underreported due to the retrospective nature of the data; however, the key finding of no difference between trainees and consultants is reassuring.

In terms of functional outcome, there was no difference in HHS between consultants and trainees at

### Table 8. Functional outcomes

| References | Oxford Hip Score | Harris Hip Score | Other scores |
|------------|------------------|-----------------|-------------|
| Inglis     | Significantly superior OHS at 6-month follow-up for consultants (40.7) vs supervised trainees (38.95; p < 0.001) and unsupervised trainees (38.27; p = 0.001). No significant difference in OHS between supervised and unsupervised groups. | N/A | N/A |
| Moran      | N/A | No significant difference in HHS at 6 months between consultant (80.2) vs trainees (80.2), no significant difference in HHS at 18 months between consultant (84.6) and trainees (85.4). | N/A | N/A |
| Palan      | No significant difference in change of OHS between two groups at 3 months, 1, 2, 3, 4 and 5 years. Superior post-operative OHS scores at 5 years in consultant group compared with trainee group (40.5 vs 39.2; p = 0.02); however, consultants had higher OHS pre-operatively. Mean change in OHS at 5-year follow-up for senior registrars was 25.2 compared with 21.8; p = 0.001. | No significant difference in HHS pre-operatively and up to 10 years post-operatively between consultants, junior and senior trainees. Significant decrease in number of patients at 7- and 10-year follow-up (n = 136 and 277, respectively) | N/A |
| Reidy      | N/A | No significant difference in HHS at 6 months between consultants (94.3) and trainees (96). | N/A |
| Weber      | N/A | No difference in WOMAC and EQ-5D scores between trainee or senior surgeon operators at 1 year | N/A |
| Wilson     | N/A | | N/A |
| Woolson    | N/A | | N/A |
six-month follow-up in the overall analysis; however, in the analysis of high-quality studies, consultants were associated with small but significantly improved HHS compared with trainees. While it is statistically significant, it is unlikely to be clinically significant, especially given the inherent case selection bias associated with predominantly retrospective data, e.g. Moran et al demonstrated that trainees were operating on significantly older patients. A recent retrospective analysis of 8158 THAs by Jolback et al found no association between surgeons’ experience and EQ-5D (Euroqol group) index, EQ-VAS (Euroqol group visual analogue scale) and pain VAS (visual analogue scale) one year after surgery. The authors did, however, find lower VAS scores one year after THA for trainees when compared with surgeons who had > 15 years’ experience.

Several studies within the orthopaedic literature, which were not suitable for inclusion in our analysis, have examined the impact of trainee involvement in arthroplasty and scoliosis surgery. These studies have also found no increase in adverse event rates associated with operations involving or being performed by trainees. Schoenfield et al reviewed data relating to total joint arthroplasties (hip and knee) with or being performed by trainees. The authors found a significantly higher rate of one or more complications and major systemic complications for operations involving a trainee compared with ones with no involvement. However, the cohorts in each group were subject to unknown case-mix and unknown levels of resident involvement. Given these added variables, which were not adjusted for, it is unreasonable to attribute this difference in outcome purely to trainees. To mitigate this added source of bias, our analysis only included studies where trainees were performing the THA (i.e. as primary operator), as opposed to merely being involved in the procedure. This makes our findings more specific to trainee outcomes.

This study was subject to a number of limitations, the most important being selection bias due to the use of non-randomized data. Consultants would therefore be more likely than trainees to operate on more challenging cases, thus subjecting both groups to a different case-mix. Adjusting for variations in case-mix between the groups was not possible due to lack of reported data and a low number of studies precluding a meta-regression model of analysis.

An additional limitation is the uneven size of the consultant and trainee groups in the overall analysis. However, when considering the high-quality sensitivity analysis, the groups were more balanced with 36.0% (2225/6187) in the trainee group and 64.0% (3962/6187) in the consultant group. The fact that few differences were observed in the results of the sensitivity analysis compared with the overall analysis suggests that this unevenness had a limited effect.

The definition of supervision varies widely across all studies and in many it is not explicitly described. This reflects the nature of real-life training where supervision can take many forms, depending on the experience and seniority of the trainee as well as the relationship between trainer and trainee. Supervision is a spectrum rather than a binary value. The arbitrary division of trainees into junior and senior in some studies was based on heterogeneous definitions. Some studies based this on year of training/residency, which is a reasonable method. This may not always reflect ability and levels of confidence, e.g. a year 1 resident would both be classified as junior residents but there is likely to be a difference in ability.

The incidence of complications such as revision and infection rate are subject to the length of follow-up. The studies included in our analysis had a mean follow-up of 42 months for the primary outcome. Clearly, longer follow-up intervals may yield a higher number of revision surgeries.

It is important to note that studies such as the ones included will always be limited in terms of quality given the hypothesis being tested. Randomized data are unlikely to become available for this type of comparison and this analysis summarizes the best available evidence. There is, however, a lack of data relating to long-term outcome for trainee-performed THA and these data would certainly be a valuable addition to the literature.

**Conclusion**

The present meta-analysis has shown that, in selected cases, trainees are safe to perform THA under supervision, with no adverse impact on patient outcomes or short-term functional outcome. This provides reassurance to trainees and their trainers as well as patients. In addition, the lack of difference in length of stay and acceptable increase in operation time should be viewed favourably by managers, service providers, trainers and, most importantly, patients.

**AUTHOR INFORMATION**

1 London North West University Healthcare NHS Trust, UK.
2 Barts Health NHS Trust, UK.
3 St. George’s Hospital NHS Trust, UK.
4 Guy’s and St. Thomas’s NHS Trust, UK.

Correspondence should be sent to: P. Singh, Department of Trauma and Orthopaedic Surgery, London North West University Healthcare NHS Trust, Watford Road, Harrow HA1 3UJ, UK.

**ICMJE CONFLICT OF INTEREST STATEMENT**

None declared.
FUNDING STATEMENT
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.

LICENCE
© 2019 The author(s)
This article is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 International (CC BY-NC 4.0) licence (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed.

REFERENCES
1. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg [Am] 2007;89-A:780–5.
2. Fehring TK, Odum SM, Troyer JL, et al. Joint replacement access in 2016: a supply side crisis. J Arthroplasty 2010;25:175–81.
3. MaraditKremers H, Larson DR, Crowson CS, et al. Prevalence of total hip and knee replacement in the United States. J Bone Joint Surg [Am] 2015;97-A:1386–97.
4. Hammond JW, Queale WS, Kim TK, McFarland EG. Surgeon experience and clinical and economic outcomes for shoulder arthroplasty. J Bone Joint Surg [Am] 2003;85-A:2318–24.
5. Schoenfeld AJ, Serrano JA, Waterman BR, Bader JO, Belmont PJ Jr. The impact of resident involvement on post-operative morbidity and mortality following orthopaedic procedures: a study of 43,343 cases. Arch Orthop Trauma Surg 2013;133:1483–91.
6. Lavernia CJ, Sierra RJ, Hernandez RA. The cost of teaching total knee arthroplasty surgery to orthopaedic surgery residents. Clin Orthop Relat Res 2000;(380):99–107.
7. van der Leeuw RM, Lombarts KM, Arah OA, Heineman MJ. A systematic review of the effects of residency training on patient outcomes. BMC Med 2012;10:65.
8. Kelly M, Bhanu A, Singh P, Fitzgerald JE, Tekkis PP. Systematic review and meta-analysis of trainee- versus expert surgeon-performed colorectal resection. Br J Surg 2014;101:750–9.
9. Marston RA, Cobb AG, Bentley G. Stannmore compared with Charnley total hip replacement. A prospective study of 413 arthroplasties. J Bone Joint Surg [Br] 1996;78-B:878–84.
10. Kalun P, Wagner N, Yan J, Nousiainen MT, Sonnadara RR. Surgical simulation training in orthopedics: current insights. Adv Med Educ Pract 2018;9:125–31.
11. Cameron JL. William Stewart Halsted. Our surgical heritage. Ann Surg 1997;225:445–58.
12. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.
13. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22:719–48.
14. Schmidt FL, Oh IS, Hayes TL. Fixed- versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. Br J Math Stat Psychol 2019;72:97–128.
15. Weber M, Benditz A, Woerner M, et al. Trainee surgeons affect operative time but not outcome in minimally invasive total hip arthroplasty. Sci Rep 2017;7:6152.
16. Moran M, Yap SL, Walmsley P, Brenkel IJ. Clinical and radiologic outcome of total hip arthroplasty performed by trainee compared with consultant orthopedic surgeons. J Arthroplasty 2004;19:853–7.
17. Woolson ST, Kang MN. A comparison of the results of total hip and knee arthroplasty performed on a teaching service or a private practice service. J Bone Joint Surg [Am] 2007;89:601–7.
18. Palan J, Guliati A, Andrew JG, et al. The trainer, the trainee and the surgeons’ assistant: clinical outcomes following total hip replacement. J Bone Joint Surg [Br] 2009;91:928–34.
19. Reidy MJ, Faulkner A, Shirrole B, Clift B. Do trainee surgeons have an adverse effect on the outcome after total hip arthroplasty? A ten-year review. Bone Joint J 2016;98-B:301–6.
20. Wilson MD, Dowsey MM, Spelman T, Choong PF. Impact of surgical experience on outcomes in total joint arthroplasties. ANZ J Surg 2016;86:967–72.
21. Inglis T, Dalzell K, Hooper G, Rothwell A, Frampton C. Does orthopedic training compromise the outcome in total hip arthroplasty? J Surg Educ 2013;70:76–80.
22. Hoang NS, Lau JN. A call for mixed methods in competency-based medical education: how we can prevent the overfitting of curriculum and assessment. Acad Med 2018;93:996–1001.
23. Knox AD, Gilardino MS, Kasten SJ, Warren RJ, Anastakis DJ. Competency-based medical education for plastic surgery: where do we begin? Plast Reconstr Surg 2014;133:702e–10e.
24. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med 2008;355:2664–9.
25. Farnworth LR, Lemay DE, Wooldridge T, et al. A comparison of operative times in arthroscopic ACL reconstruction between orthopaedic faculty and residents: the financial impact of orthopaedic surgical training in the operating room. Iowa Orthop J 2001;21:31–5.
26. Fletcher D, Edwards D, Tolichard S, Baker R, Berstock J. Improving theatre turnaround time. BMJ Qual Improv Rep 2017;6:u219831.
27. Lenguerrand E, Whitehouse MR, Beswick AD, et al. Revision for prosthetic joint infection following hip arthroplasty: Evidence from the National Joint Registry. Bone Joint Res 2017;6:391–8.