



## Hip resurfacing arthroplasty : A new method to assess and quantify learning phase

Tajeshwar S. AULAKH, Narlaka JAYASEKERA, Rohit SINGH, Amit PATEL, Jan H. KUIPER, James B. RICHARDSON

*RJAH Orthopaedic Hospital, Oswestry, UK*

Hip resurfacing had initially gained acceptance and popularity as it helps preserve femoral bone stock. In this study we tried to answer the following questions ; 1. Whether there is a learning curve for hip resurfacing ? 2. Is it present in surgeons from non-developer centres ? 3. Is it present in surgeons from developer centres as well ?

The Oswestry outcome centre was setup to serve an independent international registry for collecting, analysing and reporting outcomes following hip resurfacing. Over a 10 year period, 4535 patients (5000 hips) were recruited from different countries and within the UK from different centres in this study by 139 surgeons from 37 different countries.

Our study has shown that function can be used to assess the level of surgical competence. The results from this multilevel analysis have helped to answer the questions posed in the introduction.

Hip resurfacing is a surgical procedure with a learning phase and this learning effect is more pronounced in non-developer surgeons as compared to developer surgeons. Hip scores can be used to assess proficiency and competence of surgeons undertaking hip resurfacing arthroplasty.

**Keywords :** Hip ; resurfacing ; learning ; quantify.

### INTRODUCTION

Hip resurfacing had initially gained acceptance and popularity as it helps preserve femoral bone stock. It also provides better range of motion which

translates into active sports participation. So far good results have been reported from developer surgeons (1,18) but similar results have not been achieved by the non-developer surgeons (2,7,8). One of the reasons cited for this difference is that surgeons from non-developer centres are still in a learning phase (12,15) and thus have more complications as compared to surgeons from developer centres. Operative technique can have a major influence on post-operative hip function and post-operative pain score can be used to assess the improve-

- 
- Tajeshwar S. Aulakh, MBBS.  
*PhD Student, Keele University, Keele, UK.*
  - Narlaka Jayasekera, FRCS.
  - Rohit Singh, BMEDSCI, MBCHB, MRCS.
  - Amit Patel, MBCHB, FRCS.  
*Specialist Registrar, RJAH Orthopaedic Hospital, Oswestry, UK.*
  - Jan H. Kuiper, PhD.  
*Lecturer in Biomechanics, Keele University, Keele, UK.  
Supervisor, Joint Reconstruction Unit, RJAH Orthopaedic Hospital, Oswestry, UK.*
  - James B. Richardson, MD FRCS.  
*Professor of Orthopaedic Surgery, Keele University, Keele, UK.  
Consultant Orthopaedic Surgeon, RJAH Orthopaedic Hospital, Oswestry, UK.*
- Correspondence : Rohit Singh, Specialist Registrar, RJAH Orthopaedic Hospital, Oswestry, UK.  
E-mail : rohitamolsingh@hotmail.com  
© 2014, Acta Orthopædica Belgica.
-

ment in surgical technique (11). We used serial post-operative hip scores after each procedure as an indicator of surgical technique.

In this study we tried to answer the following questions.

1. Whether there is a learning curve for hip resurfacing ?
2. Is it present in surgeons from non-developer centres ?
3. Is it present in surgeons from developer centres as well ?

### PATIENTS AND METHODS

The Oswestry outcome centre was setup to serve an independent international registry for collecting, analysing and reporting outcomes following hip resurfacing. Over a 10 year period, 4535 patients (5000 hips) were recruited from different countries and within the UK from different centres in this study by 139 surgeons from 37 different countries. All patients were entered in this study after written informed consent and demographic and clinical details of each were provided by the respective operating surgeon. These details were stored on a secure electronic database in the outcome centre and patients were followed annually in context of function, complications, revision and death. This was done via a standard questionnaire which included a Harris hip score (6) modified for patient self-assessment. This questionnaire was divided into separate domains of pain, mobility, range of motion and satisfaction. The patient details were always kept up to date by postal, telephonic and electronic channels. In case the patients did not post a reply, they were either contacted via email or telephone and any change in address was updated. In case this failed the respective surgeons were contacted and they helped to restore communication. This helped to make sure that minimum patients were lost to follow-up. All patients were operated using the Birmingham Hip Resurfacing implant (Smith & Nephew, Warwick, UK). This study was approved by the institutional review board.

The data was checked for errors and normality. Out of 139 surgeons in this cohort, two surgeons were involved in design, development and improving surgical technique of the implant. On this basis the cohort was divided into two groups with one group containing developer surgeons and other group comprised of non-developer surgeons. The non-developer cohort included only those surgeons who had done at least 40 procedures, thus the number of

surgeons in this group was 25. The data was analysed comparatively in terms of function, complications and implant survival. We used Kaplan Meier method (9) to analyze implant survival and log rank test to compare survival between the two groups. In this study each patient reported a functional score annually after their operation up to the latest follow-up. This data is called a nested repeated measure where the scores are nested within each patient. The patients in turn may be nested within each surgeon who operated upon them. Simple regression techniques are not capable to analyse data to see the effect of different levels. Therefore, we used a new method which encompasses this technique and is called hierarchical regression or multilevel modelling.

Multilevel data has been shown to be clustered and correlated within each level e.g. the scores of patients following hip arthroplasty will tend to correlate with each other for every single patient. Similarly, if the patients have been operated by different surgeons then patient means will tend to correlate with each other for each individual surgeon. Using normal regression techniques for such data is possible by modelling a separate term for each surgeon. This can be considered inefficient as it involves estimation of many more coefficients than a multilevel analysis and does not provide any information about variation among surgeons. The simple regression techniques are unable to recognize the multilevel nature of this data and thus unable to quantify the variance due each level. This might lead to underestimation of standard errors which might result in an effect when there is no real effect. Another advantage of multilevel modelling over repeated measures ANOVA is that the latter is unable to handle missing values in data and thus tends to delete the whole observation (4,5,10,16). Multilevel modelling is capable of handling this and this helps to use all the available data for analysis.

In this analysis we used first year scores following hip resurfacing as an indicator of surgical technique. The data was sorted for each surgeon in ascending fashion based on the procedure date. This gave us serial scores for each surgeon. These scores were considered as level one and the surgeon was considered as level two. The number of procedures done by developer surgeons was greater as compared to the non-developer surgeons. This could have resulted in some shift towards the developer surgeons, therefore we chose only first 500 procedures from each of the developer surgeon. In the first model we looked at the effect of each individual procedure on first year scores. This model can be considered as level one model. In the next step we added second level with surgeons but the score for each surgeon was allowed to vary

randomly between the surgeons. This is also called as the random intercept model where scores for each surgeon have a different intercept but similar slope. In the next model we allowed the scores to vary within each surgeon and thus added a random slope for each surgeon. This model is also called random coefficients model.

All statistical analysis was done using Stata SE 9.0 (Stata Corp, Texas, US) and MLWIN 2.02 software package (Centre for Multilevel Modelling, Bristol, UK). A p value of less than equal to 0.05 was considered significant in all analysis.

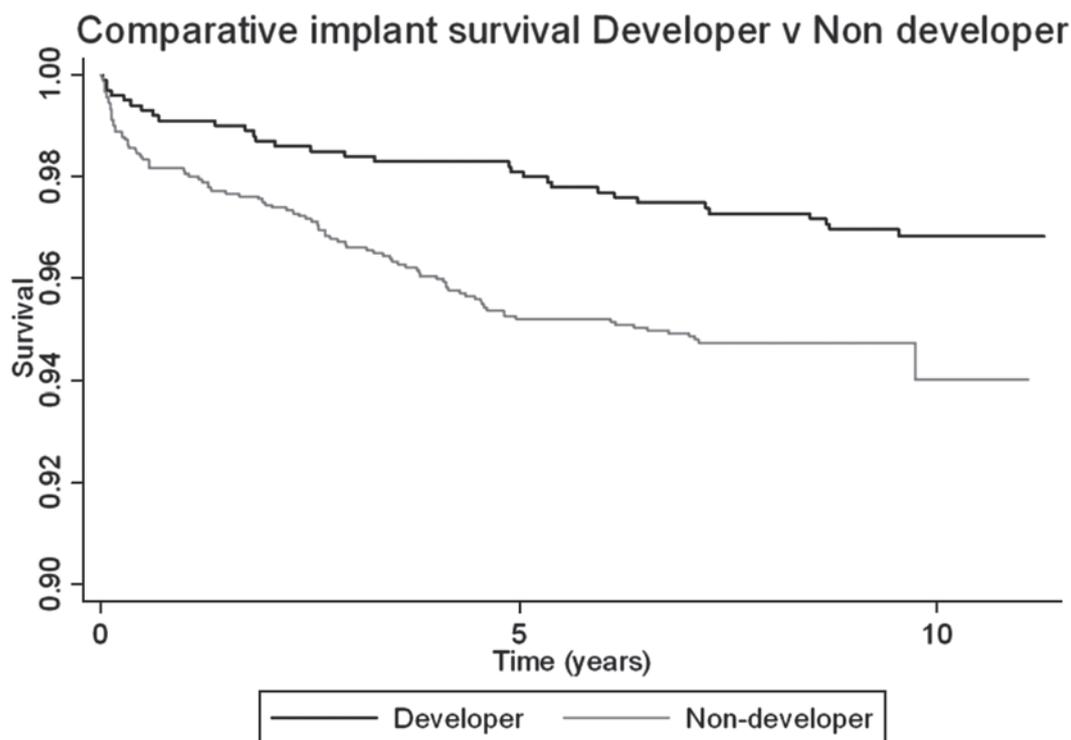
## RESULTS

2391 were operated by developer surgeons and 2144 by trainee surgeons. The mean age in developer group was 53.2 years (16.7-84.4) and 52.7 years (15.8-87.8) in trainee group. The mean follow-up was 8.5 years (5.7-10.8) and 8.1 years (5.5-10.8) in developer and trainee groups. The pre op HHS improved from 58 to 95 at 1 year in developer group and from 59 to 91 in trainee group. There was a significant difference between scores in both

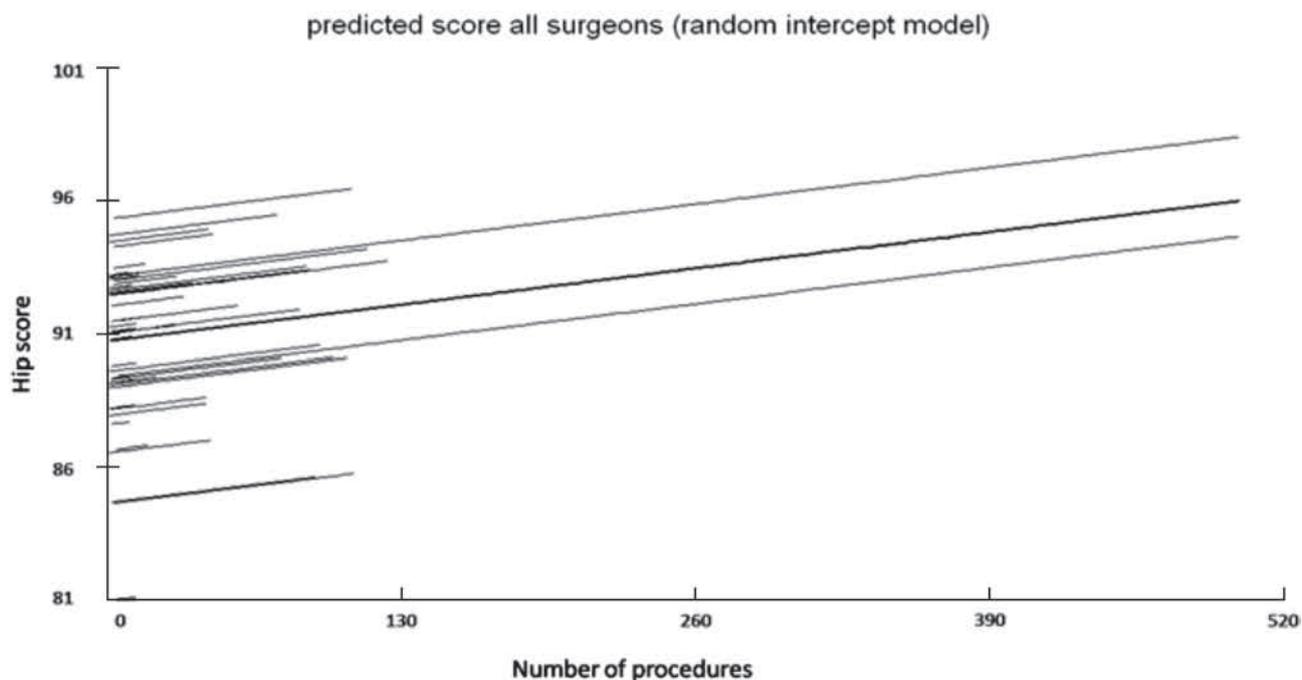
groups with developer group scoring 3.8 points higher than other group ( $p < 0.001$ ). There were 50 failures in developer group with 13 neck fractures and 3 aseptic loosening. The trainee group had 104 failures with 38 fractures. The implant survival was significantly different in both groups with 96.0% vs. 95% ( $p < 0.001$ ) (Fig. 1).

In the level one model we observed that the mean score at year one post-operative was 90.8 which had a significant positive increment of 0.011 after each procedure ( $p < 0.001$ ).

The next model showed a better fit as -2 Log likelihood changed from 16118.51 to 16035.38 and this change was significant ( $p < 0.001$ ). In this model the variance was separated into patient or procedure level and surgeon level and surgeon accounted for 8.5% variance in the score which was significant ( $p < 0.001$ ). The change in score after each procedure was still 0.010 and significant. Using the results from this analysis we plotted the predicted scores against the number of procedures (Fig. 2). The resulting plot showed an increase after each



*Fig. 1.* — Implant survival was 96.8% in developer surgeons as compared to 94% in non developer surgeons. This difference was significant ( $p = 0.0025$ ).



**Fig. 2.** — Random intercept model showing predicted scores for each surgeon following consecutive resurfacing procedures. The red line represents the overall mean for the whole cohort and blue lines correspond to individual surgeons. The slope of lines is positive which means that as the number of procedures increase, the hip score improves.

procedure which meant that after each procedure the patient had a higher score. This relates to a learning effect in surgical technique. In the next model (random coefficients) the slope for each surgeon varied. This model had a better fit as compared to random intercept model as log likelihood decreased to 16026.82, which was significant ( $p < 0.025$ ). In this model the variance matrix calculates a new value, the covariance between intercept variance and slope variance. This covariance was negative (-0.023) which means that a higher intercept will result in a low slope and vice versa. This means that the scores will tend to converge after number of procedures. The results from this model were used to plot predicted scores with a random slope for each surgeon (Fig. 3).

In the second plot the scores for each surgeon are seen to increase with increasing number of procedures and converge. This means that as the number of procedures increase the surgical technique is improving.

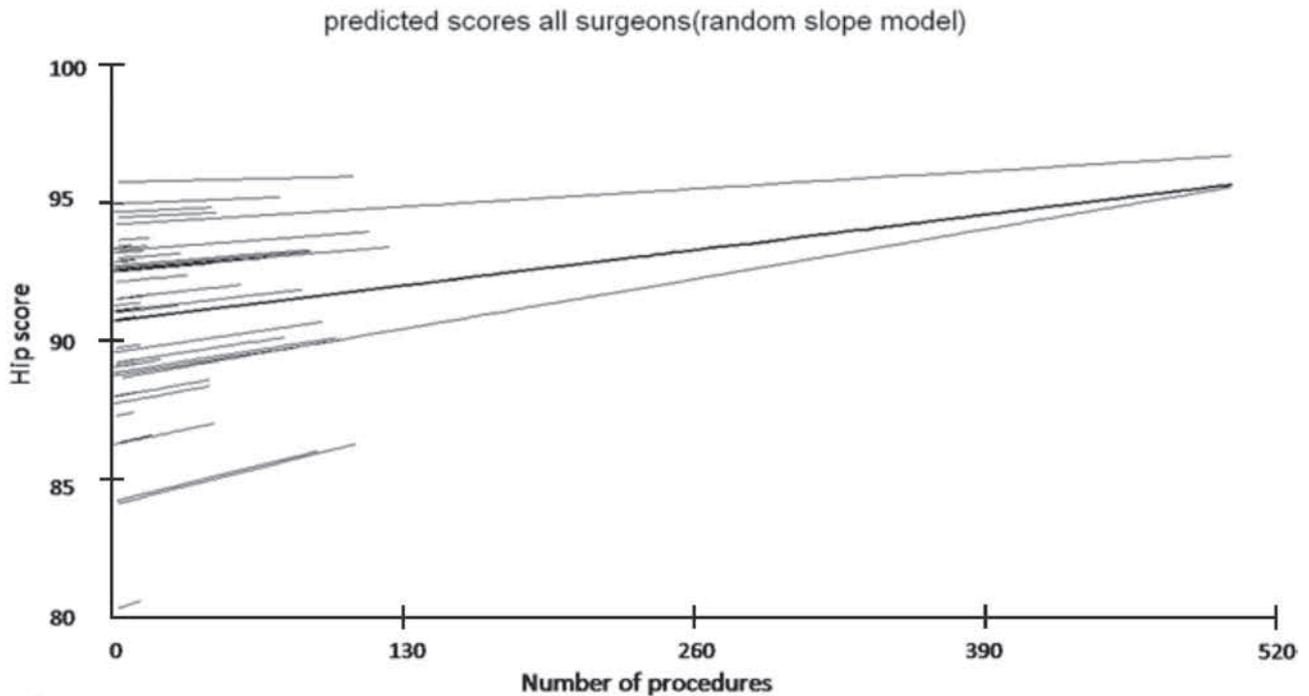
In this model we included all the surgeons irrespective of being a developer or non-developer.

Interestingly, the developer surgeons did have a high intercept but a low slope, which meant that even developer surgeons had a learning effect, although the slope was very low as compared to non-developer surgeons.

## DISCUSSION

The modified harris hip score is considered related to the success of the surgical technique, however it is still related to the initial pathology as it is a functional outcome score based around pain, mobility, range of movement, and satisfaction. Therefore, it is actually the improvement in the score that is of greatest importance as opposed to the solitary values of the score.

Our study has shown that function can be used to assess the level of surgical competence. The results from this multilevel analysis have helped to answer the questions posed in the introduction. We have seen that there is learning effect in surgeons undertaking hip resurfacing. This seems to be justified given that hip resurfacing is a new procedure. There



**Fig. 3.** — Random slope model showing predicted scores for each surgeon. The red line is the overall mean for whole cohort and blue lines represent individual surgeons. As the number of procedures increase, the lines tend to converge. This may point towards the number of procedures required to overcome the learning phase.

is a learning effect, not only in non-developer surgeons but also in the developer surgeons. Although, the gain in score after each procedure is very low in developer surgeons as compared to non-developer surgeons.

A recent study reported learning curve for hip resurfacing using pre and post op radiographs (14). The authors measured femoral and acetabular component position using the neck shaft angle and acetabular inclination respectively. They showed an improvement in component positioning with subsequent resurfacings. Similar results were also reported by Witjes *et al* (19) who compared cup angles between first 10 resurfacings and subsequent procedures. Using radiographs for assessment of surgical technique is strongly confounded by the presence of intra and inter-observer variability. If the components are not aligned biomechanically, the dynamics of the joint will be affected to a similar degree and will be apparent as symptomatic pain, limp or decreased ability to undertake a specific task. Patient

completed hip scores are designed to measure these subjective changes. Complications and subsequent revision rates have been used to assess learning effect in hip resurfacing arthroplasty. One commonly used method is Cumulative Summation analysis of failures following a surgical procedure (3,13,19). In our cohort that would involve separate analysis for each surgeon, which can be considered inefficient as compared to a multilevel analysis of hip scores which are reflective of the surgical technique. It is vital to differentiate between a complication and revision as all complications do not proceed to revision. In such a case, although the joint has not been revised, the complication will have a negative effect on the function. A functional score is resultant of the clinical and biomechanical state of a hip joint and can be used to assess surgical procedures and quantify trainee competence.

In summary, hip resurfacing is a surgical procedure with a learning phase and this learning effect is more pronounced in non-developer surgeons as

compared to developer surgeons. Hip scores can be used to assess proficiency and competence of surgeons undertaking hip resurfacing arthroplasty.

### REFERENCES

1. **Amstutz HC, Beaulé PE, Dorey FJ, Le Duff MJ, Campbell PA, Gruen TA.** Metal-on-metal hybrid surface arthroplasty : Two to six year follow-up study. *J Bone Joint Surg (Am)* 2004 ; 86 : 28-39.
2. **Back DL, Dalziel R, Young D, Shimmin A.** Early results of primary Birmingham hip resurfacing ; an independent prospective study of the first 230 hips. *J Bone Joint Surg (Br)* 2005 ; 87 : 324-9
3. **Colquhoun PHD.** CUSUM analysis of J-pouch surgery reflects no learning curve after board certification. *J Can Chir* 2008 ; 15 (4) : 296-99.
4. **Gilthorpe MS, Cunningham SJ.** The application of multi-level, multivariate modeling to orthodontic research data. *Community Dent Health* 2000 ; 17 (4) : 236-42.
5. **Goldstein H, Browne W, Rasbash J.** Multilevel modeling of medical data. *Stat Med* 2002 ; 21 (21) : 3291-315.
6. **Harris WH.** Traumatic arthritis of the hip after dislocation and acetabular fractures : Treatment by mold arthroplasty. *J Bone Joint Surg (Am)* 1969 ; 51 (4) : 737-55.
7. **Heilpern GNA, Shah NN, Fordyce MJF.** Birmingham hip resurfacing arthroplasty ; A series of 110 consecutive hips with a minimum five-year clinical and radiological follow-up. *J Bone Joint Surg (Br)* 2008 ; 90 : 1137-42.
8. **Hing CB, Back DL, Bailey M, Young DA, Dalziel RE, Shimmin AJ.** The results of primary Birmingham hip resurfacing at a mean of five years. *J Bone Joint Surg (Br)* 2007 ; 89 : 1431-8.
9. **Kaplan E, Meier P.** Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958 ; 53 : 457-81.
10. **Leyland AH, Groenewegen PP.** Multilevel modeling and public health policy. *Scand J Public Health* 2003 ; 31 (4) : 267-74.
11. **Lim TO, Soraya A, Ding LM, Morad Z.** Assessing doctors' competence : application of CUSUM technique in monitoring doctors' performance. *Int J Qual Health Care* 2002 ; 14 (3) : 251-8.
12. **Marker DR, Seyler TM, Jinnah RH, Delanois RE, Ulrich SD, Mont MA.** Femoral neck fractures after metal-on-metal total hip resurfacing : A prospective cohort study. *J Arthroplasty* 2007 ; 22 (7 Suppl 3) : 66-71.
13. **Novick RJ, Fox SA, Stitt LW, Kiaii BB, Swinamer SA, Rayman R et al.** Assessing the learning curve in off-pump coronary artery surgery via CUSUM failure analysis. *Ann Thorac Surg* 2002 ; 73 : 358-62.
14. **Nunley RM, Zhu J, Brooks PJ, Engh CA, Raterman SJ, Rogerson JS et al.** The learning curve for adopting hip resurfacing among hip specialists. *Clin Orthop Relat Res* 2010 ; 468 : 382-391.
15. **Nunley RM, della Valle CJ, Barrack RL.** Is patient selection important for hip resurfacing ? *Clin Orthop Relat Res* 2009 ; 467 : 56-65.
16. **O'Connell AA and McCoach DB.** Applications of hierarchical linear models for evaluations of health interventions : demystifying the methods and interpretations of multilevel models. *Eval Health Prof* 2004 ; 27 (2) : 119-51.
17. **Steiner SH, Cook RJ, Farewell VT, Treasure T.** Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000 ; 1 (4) : 441-52.
18. **Treacy RBC, McBryde CW, Pynsent PB.** Birmingham hip resurfacing arthroplasty : A minimum follow-up of five years. *J Bone Joint Surg (Br)* 2005 ; 87 : 167-70.
19. **Witjes S, Smolders JM, Beaulé PE, Pasker P, Van Susante JL.** Learning from the learning curve in total hip resurfacing : a radiographic analysis. *Arch Orthop Trauma Surg* 2009 ; 129 : 1293-99.