The association of patient characteristics and surgical variables on symptoms of pain and function over 5-years following primary hip replacement surgery: population-based cohort study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002453
Article Type:	Research
Date Submitted by the Author:	05-Dec-2012
Complete List of Authors:	Judge, Andrew; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Arden, Nigel; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Batra, Rajbir; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Thomas, Geraint; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Beard, David; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Javaid, Kassim; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Cooper, Cyrus; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Murray, David; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Epidemiology
Keywords:	EPIDEMIOLOGY, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Bone diseases < ORTHOPAEDIC & TRAUMA SURGERY

The association of patient characteristics and surgical variables on symptoms of pain and function over 5-years following primary hip replacement surgery: population-based cohort study

Andy Judge[1,2], Senior Statistician

Nigel K Arden[1,2], Professor in Rheumatic Diseases and Consultant Rheumatologist

Rajbir Batra[1], Medical Statistician

Geraint Thomas[1], Clinical Fellow

David Beard[1], Professor of Musculoskeletal Sciences

M Kassim Javaid[1,2], Lecturer in Metabolic Bone Disease

Cyrus Cooper[1,2], Director and Professor of Rheumatology

David Murray[1], Professor of Orthopaedic Surgery and Consultant Orthopaedic Surgeon

Exeter Primary Outcomes Study (EPOS) group*

[1] Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Headington, Oxford, OX3 7LD, UK

[2] MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, SO16 6YD, UK

*The following people are the main investigators of the EPOS group: Professor D Murray, Mr G Andrew, Mr P Gibson, Mr J Nolan, Mr A Hamer, Mr M Fordyce and Mr K Tuson. Dr DJ Beard is a member of the EPOS steering group and scientific advisor. The following are study co-ordinators for the EPOS group: Anne Potter, Ann McGovern, Kathleen Reilly, Cathy Jenkins, Adele Cooper, Clare Darrah, Lynne Cawton, Praveen Inaparthy and Clare Pitchfork.

Corresponding Author: Andy Judge
NIHR Musculoskeletal Biomedical Research Unit,
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences,
University of Oxford,
Windmill Road, Headington,
Oxford, OX3 7LD, UK.
Fax: +44 (0) 1865 227 966
Tel: +44 (0) 1865 737 837

Email: Andrew.Judge@ndorms.ox.ac.uk

Words: Abstract 262; Text 3430

Tables: 2; **Figures:** 3.

Key words: Epidemiology, Osteoarthritis, Hip replacement, Patient reported outcome

Abstract

Objectives

To identify patient characteristics and surgical factors associated with patient reported outcomes over 5-years following primary total hip replacement (THR)

Design

Prospective population-based cohort study

Participants and Setting

The Exeter Primary Outcomes Study of 1,431 primary hip replacements for osteoarthritis

Main outcome measures

The Oxford Hip Score (OHS) collected pre-operatively and each year up to 5-years post-operatively. Repeated measures linear regression modelling is used to identify patient and surgical predictors of outcome and describe trends over time.

Results

The majority of patient's demonstrated substantial improvement in pain/function in the first year after surgery – between one and five-years follow-up there was neither further improvement nor decline. The strongest determinant of attained post-operative OHS was the *pre-operative OHS* – those with worse pre-operative pain/function had worse post-operative pain/function. Other predictors with small but significant effects included: *femoral component size* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50-60, younger (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* more *co-existing diseases* and worse *SF36 mental health*, was related to worse post-operative pain/function. Assessment of change in OHS between pre- and post-operative assessments revealed that patients achieved substantial and clinically relevant symptomatic improvement (change), regardless of variation in these patient and surgical factors.

Conclusions

Patients received substantial benefit from surgery, regardless of their pre-operative patient and surgical characteristics (baseline pain/function, age, BMI, comorbidities, mental health and femoral

component size). Further research is needed to identify other factors that can improve our ability to identify patients at risk of poor outcomes from THR surgery.

ARTICLE SUMMARY

Article focus

- Total hip replacement is a common and successful surgical intervention, providing substantial relief from pain and improvement in function in patients with hip arthritis
- An important minority of patients continue to experience some pain and functional disability following surgery
- Relatively little work has been done to establish the predictors of patient reported outcomes after hip replacement, in particular the role of intra-operative surgical factors and how symptoms change over time in the mid to long term

Key messages

- The majority of patients achieved large improvement in symptoms of pain and function in the first year following surgery – there was no further improvement nor decline between one and five years
- An new finding is that a larger femoral component size (offset) is associated with better outcomes of THR in women
- Small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, are greatly outweighed by the substantial *change* in OHS achieved by these patients

Strengths and limitations

- Strengths include the large sample size, repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR, with data collected prospectively over 5-years with a good rate of follow up.
- Further strengths include the use of multiple imputation and bootstrapping as an internal validation technique, ensuring the predictors identified are those most likely to be replicated in external validation studies.
- Limitations are that other potential predictive variables were not available in this study, such as radiographic grade, pattern of OA, patient expectations of surgery and the type and extent of joint damage.
- Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated.

INTRODUCTION

Total hip replacement surgery (THR) is a commonly performed and successful surgical intervention, providing substantial relief from pain and improvement in functional disability in patients with hip arthritis[1-4]. Attention has turned from looking at the technical outcomes of surgery, such as prosthesis survival, to the use of patient reported outcome measures (PROMs) to see whether surgery has been successful from the patient's perspective[3,5]. Through the use of PROMs it has emerged that whilst on average the majority of patients improve after surgery, an important minority of patients continue to experience some pain and functional disability after THR, where some have no improvement or get worse[6-10].

Relatively little work has been done to establish the predictors of good or bad patient reported outcomes after THR[11]. Several potential determinants of outcomes of THR have been identified within the literature including baseline levels of pain and function[8,12-16], severity of clinical disease[13,16], age[13,16,17], gender[13,15,18], radiographic grade[13,14], education[8,12,14,18], obesity[15,17], co-morbidities[8,15], living alone[15,19], mental health[16], and patients expectations of surgery[14,20]. Little is known about the role intra-operative surgical factors may have on patient reported outcomes. In addition the majority of prior research looks at short-term outcomes and few studies have examined how symptoms change over time in the longer term.

Using a large prospective cohort of patients receiving primary THR for osteoarthritis (OA) with repeated measures of patient reported outcomes (as measured by the Oxford Hip Score) at yearly intervals over a 5-year follow up period, the aim of this study was to identify patient characteristics and intra-operative surgical factors associated with differences in: a) attained post-operative levels of pain and function, b) change (temporal trends) in symptoms of pain and function over time between pre and post-operative assessments.

METHODS

We obtained information from the Exeter Primary Outcomes Study. Details of the study have previously been published elsewhere[21-23]. Patients were consecutively recruited between January 1999 and January 2002 at seven centres across England and Scotland. Patients underwent THR using a cemented Exeter femoral stem component (Stryker Howmedica Osteonics, Mahwah, New Jersey)[24]. A variety of cemented and uncemented acetabular components were used. Patients were included if they were undergoing primary hip replacement with an Exeter cemented femoral stem and were willing and able to give consent to participate in the study. North Western Multiple Centre

Research Ethics Committee and the local research ethics committees in all the participating centres gave ethical approval for conducting the study. All eligible patients were invited to participate in the study. Patient recruitment varied between the centres but was between 80%-90% of eligible patients. The geographical area covered by the participating hospitals was wide and included both university teaching and district general hospitals that included urban as well as rural locations and represented both affluent as well as inner city suburbs. The catchment area of the four combined units included over a million people. There were 1,375 patients (1,431 hips) with a primary diagnosis of OA. The unit of analysis was the implant rather than the patient, of whom 56 had bilateral procedures. We examined 1431 THRs performed by consultant and non-consultant surgeons and using anterolateral or posterior approaches.

Patient Variables

At the pre-operative assessment information was collected on age, gender, height and weight (from which body mass index (BMI) was calculated), primary diagnosis and current occupation. Patients were asked whether they were using concomitant therapies such as oral anticoagulants, corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDS) and other analgesics. An ordinal variable was created of the number of pre-operative co-existent diseases a patient had, which included deep venous thrombosis, pulmonary embolism, urinary tract infection, other musculoskeletal disease, neurological, respiratory, cardiovascular, renal, hepatic disease or treatment for other medical conditions. Fixed flexion range of motion recorded in degrees was obtained from the Charnley Modification of D'Aubigne-Postel Grade questionnaire[25,26]. Patients completed a pre-operative Short Form 36 (SF-36)[27], which measures Quality of Life generically through eight domains: physical function (PF), bodily pain (BP), general health (GH), role physical (RP), vitality (VT), social function (SF), role emotional (RE) and mental health (MH). The lowest score, 0, corresponds to the worst possible health and 100, to the best possible health.

Surgical variables

Detailed intra-operative information was collected for each patient. This included information on the grade of operator (consultant, registrar, senior house officer), surgical approach (anterolateral, posterior) and patient position (supine, lateral). Data was available on whether or not a lavage system was used for the acetabular component, whether there was cement pressurisation for both femoral and acetabular components, the type of cement used in both the socket (none, simplex, cmw1, palacos r, other) and the femur (simplex, cmw1, cmw3, palacos r, palacos lv), the type of polyethylene used (uhmwpe, cross-linked), whether the femoral head was stainless steel or ceramic, femoral head size (22, 26 or 28 millimetres), and the femoral component stem size (35.0, 37.5, 44.0, 50.0 millimetres offset). The duration of the operation was recorded in minutes.

Outcome variable

Prior to surgery, patients completed an Oxford Hip Score (OHS) questionnaire with follow-up questionnaires being filled in at 1, 2, 3, 4 and 5-years post-surgery. The Oxford hip score was introduced in 1996 predominantly for use in clinical trials[28]. The score is joint specific and has been assessed for reliability and validity[29]. The OHS consists of 12 questions asking patients to describe their hip pain and function during the past 4 weeks. Each question is on a Likert scale taking values from 0-4. An overall score is created by summing the responses to each of the 12 questions. A total score was created ranging from 0 to 48, where 0 is the worst possible score (most severe symptoms) and 48 the best score (least symptoms).

Statistical Methods

Stata version 11.1 (Stata, College Station, TX) was used for all statistical analyses. Potential prognostic variables included the patient and surgical variables described above. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. To overcome this bias we used multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[30] and 10 imputed datasets created. We included all predictor variables in the multiple imputation process (as listed earlier), together with the outcome variable as this carries information about missing values of the predictors. We fitted two models to describe the association of the patient and surgical variables on the following outcomes:

a) Attained post-operative OHS at 1, 2, 3, 4 and 5-years follow up

A repeated measures linear regression model was fitted where the outcomes were the OHS at 1, 2, 3, 4 and 5-years follow up, adjusting for the pre-operative OHS as a covariate in the model. Generalised Estimating Equations (GEE) was used to account for clustering within the data using an exchangeable correlation matrix. This model estimates the impact of predictors on the average OHS over the 5 follow up time points. Fractional polynomial regression modelling was used to explore evidence of non-linear relationships for continuous variables. Interaction terms were fitted between the predictor variable and time, to see if the association of the predictor on outcome changed between 1 and 5-years follow up.

b) Change in OHS between baseline and 1, 2, 3, 4 and 5-year follow ups

For variables identified as significant predictors of attained post-operative OHS, the repeated measures linear regression model is fitted, where the outcome is expanded to include the pre-operative and 1, 2, 3, 4 and 5-year post-operative OHS. Interaction terms are fitted between the predictor variable and time, to describe the change in OHS over time[31] across categories of the predictor variable, for example, in those who are obese versus not obese.

Model validation

The full regression model including all predictor variables is fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[30]. Given the extensive list of patient and surgical variables considered for inclusion in the model, we wanted to ensure that we minimised the possibility of making a type 1 error (rejecting the null hypothesis when it is true) – e.g. the chance that a variable identified as being 'significant' in this dataset may not be replicated in other samples of patients. For internal validation of the model we therefore used a combination of multiple imputation and bootstrapping[32,33] (see **supplementary file**). 200 bootstrap samples are randomly drawn with replacement. An automatic backward selection procedure is applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of $\alpha = 0.157$. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time. To assess model discrimination we use the R^2 statistic as a measure of explained variation[34].

RESULTS

Data is available on 1375 patients (1431 hips) receiving primary hip replacement surgery between January 1999 and January 2002. Of these patients 1281 (89.5%) completed a pre-operative OHS questionnaire and at least 1 of the follow up questionnaires, and were included in the analysis. 80% of patients completed the OHS at the 1-year follow up and this declined to a 70% response rate by 5-years. Baseline demographic details are described in **Table 1**. Comparing patients that did, and did not, respond to the 5-year follow up questionnaire, there were no differences in baseline pain and function as assessed by the OHS. Differences were observed where those that responded were younger, less likely to be unemployed/retired, and had better pre-operative SF36 mental health scores.

Histograms of the distribution of OHS at baseline, follow-up, and the absolute difference in scores (**Figure 1**) highlight that at the 1 and 5-year follow ups the score is negatively skewed to the left, suggesting the majority of patients achieve improvement in pain and function. The histograms of the difference in scores highlight that almost all patients get better with only a small minority getting worse or receiving no improvement (2.3% by 1-year and 1.2% by 5-years). The change in OHS over time from the repeated measures regression model is displayed in **Figure 2**. This demonstrates that regardless of the level of pre-operative OHS, patients achieved substantial improvement in pain and function following surgery. Those with the worst pre-operative scores achieved the greatest improvement (28.8 point change in those with pre-operative OHS < 5), however patients with the best pre-operative scores still achieved substantial improvement (change of 10.6 points in those with pre-operative OHS > 30). Interestingly, between 1 and 5-years follow up a steady state is reached where there is no further improvement nor decline, with a non-significant trend between OHS and time (p=0.88).

A number of variables were identified as important predictors of attained post-operative OHS (**Table 2**). The strongest determinant of outcomes was the baseline OHS. Increasing baseline OHS (better pre-operative pain/function) was associated with increasing follow-up OHS (better post-operative pain/function). The effect of age was non-linear, where compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcomes. Increasing BMI, patients with a greater number of co-existing diseases prior to surgery, and those with worse pre-operative SF36 mental health scores, also had worse outcomes. The surgical predictor we identified was femoral component size (offset), where patients with larger components (offset of 44 or more) had significantly better outcomes. We hypothesised that this may be explained by an interaction with gender, where a larger offset is used in men. A significant interaction was observed, where no association was observed in men, whilst in women those with larger components had better outcomes. The effect of surgical approach was significant at the 1-year follow-up, where the anterolateral approach had better outcomes than posterior (difference in 1-year OHS of 2.2 units 95%CI (1.1 to 3.3), however the effect size attenuated over time and became no longer significant between 3 to 5 years follow up.

Figure 3 describes the change in OHS over time stratified according to each of the predictor variables we identified in this study. The graphs highlight that whilst there are small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, this is greatly outweighed by the substantial *change* in OHS achieved by these patients, regardless of whether they are old or young, obese or not obese. These patient groups still receive great benefit from surgery.

Assessing the discriminatory ability of the final model, including the baseline OHS alone explained 10.3% of the variability in outcome. The final predictive model including the patient and surgical variables explained 16.6%. This suggests that although we have identified significant patient and surgical predictors of outcome, they have smaller effects, and explain little of the variability in attained post-operative OHS above that of the baseline score.

DISCUSSION

Main findings

Within a large prospective cohort of patients receiving primary THR in the UK, we identified a number of predictors of differences in *attained* post-operative pain and function. Determinants included: *pre-operative pain and function* – those with worse pre-operative pain/function had worse post-operative pain/function; *femoral component size (offset)* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* a greater number of *co-existing diseases* and worse *SF36 mental health* at the time of surgery, was related to worse post-operative pain and function. The strongest determinant of outcome was the baseline score with the patient and surgical variables contributing small but statistically significant effects.

Assessing the relationship of *change* in symptoms of pain and function between pre and post-operative assessments for predictor variables in the final model revealed that patients achieved large symptomatic improvement (change), regardless of differences in pre-operative patient and surgical factors. The change in symptoms greatly outweighs any differences in attained post-operative score – patients achieved great benefit from surgery regardless of factors such as their age and BMI at the time of surgery. Exploring temporal trends in symptoms of pain and function over time demonstrated that there was little further improvement nor decline in the short to mid-term (between 1 and 5-years follow-up) where a steady symptomatic state was reached.

Strengths and limitations

The strength of this study lies in its large sample size and repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR[28,29], with data collected prospectively over 5-years with a good rate of follow up. The use of multiple imputation and bootstrapping as an internal validation technique is a strength of this study, ensuring the predictors we identified are those most likely to be replicated in external validation studies, and not chance significant findings that are anomalies of our dataset. Limitations are that other potential predictive variables were not available in this study, predominantly radiographic factors such as x-ray grade and pattern of OA, and other factors including patient expectations of surgery and the type and extent of joint damage. Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated in this study.

What is already known

In our study an important new finding was that a larger femoral component size (offset) is associated with better outcomes of THR. We hypothesised that the effect of femoral component size may be explained by an interaction with gender, as men have larger offsets. A significant interaction was observed, where no association was observed in men, whilst in women those with larger offsets had better outcomes. The choice of offset can affect hip stability as well as abduction strength, potentially resulting in abnormal (trendelenburg) gait. Component offset may be preoperatively templated, although common offsets are often assumed (i.e. 37.5mm for females and 44.0mm males). There is greater potential to decrease offset in females, partly because of the above assumption, and it is sometimes difficult to use the larger offset components because a smaller femoral canal diameter precludes their use. The choice of offset for the femoral stem has not changed since this study was conducted and these finding are generalizable to clinical practice. We are not aware of any data in the literature describing the relationship of intra-operative surgical factors on patient reported outcomes. Within the literature, data exists on the relationship between head size and failure of a THR, whereby larger head size (40mm versus 28mm) with a ceramic-on-ceramic bearing surface was associated with lower 5-year revision rates[35]. This is though to be related to larger diameter heads increasing fluid-film lubrication, in turn reducing wear, and decreasing dislocation rates[35]. However, femoral stem offset and head size are independent factors that are unrelated to one another. Within this study we found no association between head size and patient reported outcomes – this is unsurprising since the head sizes used were 22, 26 and 28mm and common to orthopaedic practice at the time this study took place. Thinner liners are now manufactured allowing larger head sizes to be used in acetabular components of the same size.

Consistent with others in the literature we observed that worse pre-operative mental health was a predictor of poor outcome[16] as were greater numbers of pre-operative co-morbidities[8,15]. We found that older age and higher BMI was associated with worse patient reported outcomes. Within the literature whilst some authors conclude that older age and increasing BMI area associated with worse outcomes, others have found no evidence of an association[7-9,12-19,31,36]. This is in line with the conclusions of large literature reviews stating that such factors are not strong predictors of outcome[11]. The findings are important to decision making as physicians often advise patients they are too old or obese to receive THR[11,37]. We can conclude that in relation to patient reported outcomes of THR, even if some groups fare less well after THR, it does not mean these patients don't get benefit from surgery[11]. Expectations of the patients may also play a role, where for example, what a young person wants to achieve in functional rehabilitation is different to an older person, for whom a lower functional score may be perfectly acceptable. It also is well known within the literature that patients with better pre-operative pain and function achieve better *attained* post-operative pain and function, and that patients with worse pre-operative pain and function get the greatest *change* (symptomatic improvement) between baseline and follow up[12,30,38-40].

Whilst the effects of pre-operative patient characteristics including age, BMI, co-existing diseases and mental health may already be known within the existing literature, what is novel about this study is contrasting the effects of these factors on attained levels of post-operative scores with a graphical representation of change (improvement in scores). Our findings highlight that small but significant differences in attained scores are greatly outweighed by the fact these patient groups get great benefit from surgery (in terms of substantial change in symptoms of pain and function), regardless of differences in pre-operative patient characteristics.

What this study adds

Within this study we have identified a number of patient and surgical predictors of attained post-operative pain and function following THR surgery. These predictors remain related to outcome over the short to mid term. An important new finding was that larger femoral component size (offset) was associated with better post-operative pain and function in women. This finding implies that greater consideration should be given to measuring and deciding upon the choice of offset in women as there is potential to undersize. Whilst age, BMI, co-existing diseases, mental health and femoral component size were associated with small but significant differences in attained pain and function, analyses of change demonstrate that these patient still achieve substantial symptomatic benefit from surgery regardless of differences in these pre-operative factors. The findings will be important to inform patient and clinician decision regarding the likely outcomes of surgery for these patient groups. Although we have assessed a wide range of patient characteristics and intra-operative surgical factors there is still uncertainty as to the cause of variation in outcomes of hip replacement. There is a need to focus on issues that remain unclear such as the effect of soft tissue and the severity and pattern of OA. Further research is needed using more detailed measures of existing predictive variables, and identification of other factors beyond those observed in this study that explain a greater proportion of the variability in outcome to improve our ability to identify patients at risk of poor outcomes from THR surgery.

Funding statement

This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10064). The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This study has been supported by Stryker Howmedica Osteonics (Newbury, UK). Support was also received from the Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Orthopaedic Centre, University of Oxford and the UK Medical Research Council, MRC Lifecourse Epidemiology Unit, University of Southampton.

Competing interest declaration

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: AJ, RNB and GT have no conflicts of interest; MKJ, NKA, and CC have received honorariums, held advisory board positions (which involved receipt of fees), and received consortium research grants, respectively, from: Novartis and Alliance for Better Health and Lilly; Merck, Merck Sharp and Dohme, Roche, Novartis, Smith and Nephew, Q-MED, Nicox, Servier, GlaxoSmithKline, Schering-Plough, Pfizer, and Rottapharm; and Alliance for Better Bone Health, Amgen, Novartis, Merck Sharp and Dohme, Servier, Eli Lilly, and GlaxoSmithKline; DB has held an independent consultancy with ICNet and Stryker, and has been a grant holder on a Genzyme-funded study; DM has received royalties from Biomet and Wright Medical Technology, Inc, and receives research support from DePuy, A Johnson & Johnson Company, Stryker, Zimmer and Wright Medical Technology, Inc; they have no other relationships or activities that could appear to have influenced the submitted work.

Contributors

AJ NKA RB GT DB MKJ CC DM: (1) substantial contributions to conception and design, analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published.

Ethics

Ethical approval was obtained from the Salford and Trafford Research Ethics Committee (Project No: - 98105 – MREC 98/8/20 UK Multicentre Exeter Primary Outcome Study). Informed written consent was obtained from all participants.

Data Sharing

No additional data are available.

REFERENCES

1. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (1). *N. Engl. J. Med.* 1990;323(11):725-731.

2. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (2). *N. Engl. J. Med.* 1990;323(12):801-807.

3. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-Related Quality of Life in Total Hip and Total Knee Arthroplasty. A Qualitative and Systematic Review of the Literature. *J. Bone Joint Surg. Am.* 2004;86(5):963-974.

4. Cooper C, Arden NK. Excess mortality in osteoarthritis. *BMJ.* 2011;342.

5. Professor the Lord Darzi of Denham Kbe HonFREng FmedSci. *High Quality Care for All. NHS Next Stage Review Final Report - Summary*: Department of Health;2008.

6. Williams O, Fitzpatrick R, Hajat S, et al. Mortality, morbidity, and 1-year outcomes of primary elective total hip arthroplasty. *J. Arthroplasty.* 2002;17(2):165-171.

7. Nilsdotter AK, Petersson IF, Roos EM, Lohmander LS. Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study. *Ann. Rheum. Dis.* 2003;62(10):923-930.

8. MacWilliam CH, Yood MU, Verner JJ, McCarthy BD, Ward RE. Patient-related risk factors that predict poor outcome after total hip replacement. *Health Serv. Res.* 1996;31(5):623-638.

9. Judge A, Cooper C, Williams S, Dreinhoefer K, Dieppe P. Patient-Reported Outcomes One Year After Primary Hip Replacement in a European Collaborative Cohort. *Arthritis Care Res.* 2010;62(4):480-488.

10. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. *Curr. Opin. Rheumatol.* 2006;18(5):526-530.

11. Santaguida PL, Hawker GA, Hudak PL, et al. Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can. J. Surg.* 2008;51(6):428-436.

12. Fortin PR, Clarke AE, Joseph L, et al. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. *Arthritis Rheum.* 1999;42(8):1722-1728.

13. Cushnaghan J, Coggon D, Reading I, et al. Long-term outcome following total hip arthroplasty: a controlled longitudinal study. *Arthritis Rheum.* 2007;57(8):1375-1380.

14. Judge A, Cooper C, Arden NK, et al. Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres. *Osteoarthritis Cartilage.* 2011;19(6):659-667.

15. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. The Effect of Age on Pain, Function, and Quality of Life After Total Hip and Knee Arthroplasty. *Arch. Intern. Med.* 2001;161(3):454-460.

16. Quintana JM, Escobar A, Aguirre U, Lafuente I, Arenaza JC. Predictors of health-related quality-of-life change after total hip arthroplasty. *Clin. Orthop.* 2009;467(11):2886-2894.

17. Braeken AM, Lochhaas-Gerlach JA, Gollish JD, Myles JD, Mackenzie TA. Determinants of 6-12 month postoperative functional status and pain after elective total hip replacement. *Int. J. Qual. Health Care.* 1997;9(6):413-418.

18. Rissanen P, Aro S, Sintonen H, Slatis P, Paavolainen P. Quality of life and functional ability in hip and knee replacements: a prospective study. *Qual. Life Res.* 1996;5(1):56-64.

19. Hopman WM, Mantle M, Towheed TE, Mackenzie TA. Determinants of health-related quality of life following elective total hip replacement. *Am. J. Med. Qual.* 1999;14(3):110-116.

20. Gandhi R, Davey JR, Mahomed N. Patient expectations predict greater pain relief with joint arthroplasty. *J. Arthroplasty.* 2009;24(5):716-721.

21. Andrew JG, Palan J, Kurup HV, Gibson P, Murray DW, Beard DJ. Obesity in total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2008;90-B(4):424-429.

22. Palan J, Beard D, Murray D, Andrew J, Nolan J. Which Approach for Total Hip Arthroplasty: Anterolateral or Posterior? *Clin. Orthop.* 2009;467(2):473-477.

23. Palan J, Gulati A, Andrew JG, Murray DW, Beard DJ, the Esg. The trainer, the trainee and the surgeons' assistant: clinical outcomes following total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2009;91-B(7):928-934.

24. Hossain M, Parfitt D, Beard D, et al. Does pre-operative psychological distress affect patient satisfaction after primary total hip arthroplasty? *BMC Musculoskeletal Disorders.* 2011;12(1):122.

25. Charnley J. The Long-Term Results of Low-Friction Arthroplasty of the Hip Performed as a Primary Intervention. *J. Bone Joint Surg. Br.* February 1, 1972 1972;54-B(1):61-76.

26. D'Aubigne RM, Postel M. Functional Results of Hip Arthroplasty With Acrylic Prosthesis. *J. Bone Joint Surg. Am.* June 1, 1954 1954;36(3):451-475.

27. Ware JE. *How to score the revised MOS Short Form Health Scales*: Boston: The Health Institute, New England Medical Centre Hospital;1988.

28. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J. Bone Joint Surg. Br.* 2007;89-B(8):1010-1014.

29. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J. Bone Joint Surg. Br.* 1996;78-B(2):185-190.

30. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).

31. Fitzgerald JD, Orav EJ, Lee TH, et al. Patient quality of life during the 12 months following joint replacement surgery. *Arthritis Rheum.* 2004;51(1):100-109.

32. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.

33. Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology.* 2007;7(1):33.

34. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

35. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *The Lancet.*379(9822):1199-1204.

36. Harse JD, Holman CD. Charlson's Index was a poor predictor of quality of life outcomes in a study of patients following joint replacement surgery. *J. Clin. Epidemiol.* 2005;58(11):1142-1149.

37. Hawker G, Wright J, Coyte P, et al. Health-related quality of life after knee replacement. *J. Bone Joint Surg. Am.* 1998;80(2):163-173.

Table 1. Descriptive statistics and comparison of those who did, and did not, complete the 5-year follow up questionnaire

Variable	Missing	Baseline (n = 1431)	Non Responders at Year 5	Responders at Year 5	P-value
Oxford Hip Score					
Pre-operative	70 (4.9%)	16.4 (7.8)	16.1 (8.2)	16.5 (7.6)	0.35
1-year post-op	274 (19.1%)	43 (36, 46)	-	-	
2-year post-op	316 (22.1%)	43 (36, 47)	-	-	
3-year post-op	368 (25.7%)	43 (36, 47)	-	-	
4-year post-op	430 (30.0%)	44 (36, 47)	-	-	
5-year post-op	396 (27.7%)	43 (37, 47)	-	-	
Patient characteristics					
Age	9 (1%)	70.0 (63.9, 76.1)	73.6 (66.2, 79.3)	68.8 (62.7, 74.5)	< 0.001
BMI	95 (7%)	27.4 (4.9)	27.1 (4.9)	27.6 (4.8)	0.077
Gender	7 (0%)				0.46
Male		537 (38%)	171 (39%)	366 (37%)	
Female		887 (62%)	266 (61%)	621 (63%)	
Occupation	0 (0%)				0.005
Heavy manual		41 (3%)	8 (2%)	33 (3%)	
Light manual		89 (6%)	20 (5%)	69 (7%)	
Office / professional		107 (7%)	21 (5%)	86 (9%)	
Housewife		187 (13%)	67 (15%)	120 (12%)	
Unemployed / retired		1007 (70%)	325 (74%)	682 (69%)	
No. of Co-existing Diseases	0 (0%)				0.94
0		431 (30%)	136 (31%)	295 (30%)	
1		498 (35%)	147 (33%)	351 (35%)	
2		315 (22%)	99 (22%)	216 (22%)	
3		140 (10%)	43 (10%)	97 (10%)	
4		47 (3%)	16 (4%)	31 (3%)	
Concomitant therapy used	8 (1%)				0.84
No		104 (7%)	31 (7%)	73 (7%)	
Yes		1319 (93%)	406 (93%)	913 (93%)	
SF36 Mental Health Score	515 (36%)	74 (60, 88)	72 (52, 88)	76 (60, 88)	0.046

Cells represent either: number (percentage), mean (standard deviation), median (interquartile range)
† T-tests are used for continuous variables and chi-squared tests for categorical variables.
Where continuous variables were not normally distributed, a non-parametric t-test (Kruskal-Wallis) was used.
Fishers exact test is used where expected counts are less than 5.

Table 2. Repeated measures Analysis of Covariance (ANCOVA) models to identify predictors of the average OHS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Multivariable Δ Coef (95% CI)	P-value
Patient variables				
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	0.01 (-0.11, 0.13)	0.88
Age				
<50	-1.44 (-3.92, 1.03)	0.25		
50-60	-0.96 (-2.27, 0.35)	0.15	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	-1.54 (-2.45, -0.64)	0.001
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	-0.90 (-1.27, -0.54)	< 0.001
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	0.76 (0.46, 1.07)	< 0.001
Surgical variables				
Stem size (mm offset)	0.15 (0.04, 0.26)	0.01	0.17 (0.06, 0.28)	0.002
R^2				17.4%
Optimism				0.8%
Bias-Corrected R^2				16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Variables included in the final regression model are those that are retained in at least 70% of the 200 bootstrap backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OHS only

Figure 1. Distribution of OHS at baseline, follow-up and absolute difference in scores

Figure 2. Change in Oxford Hip Score over time, stratified by baseline score

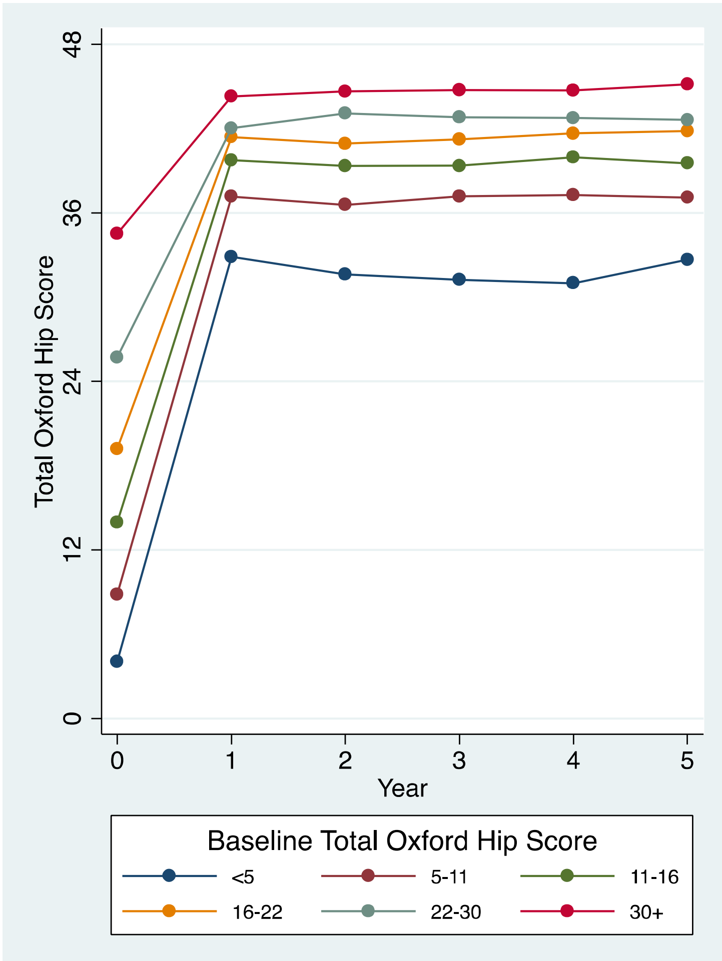
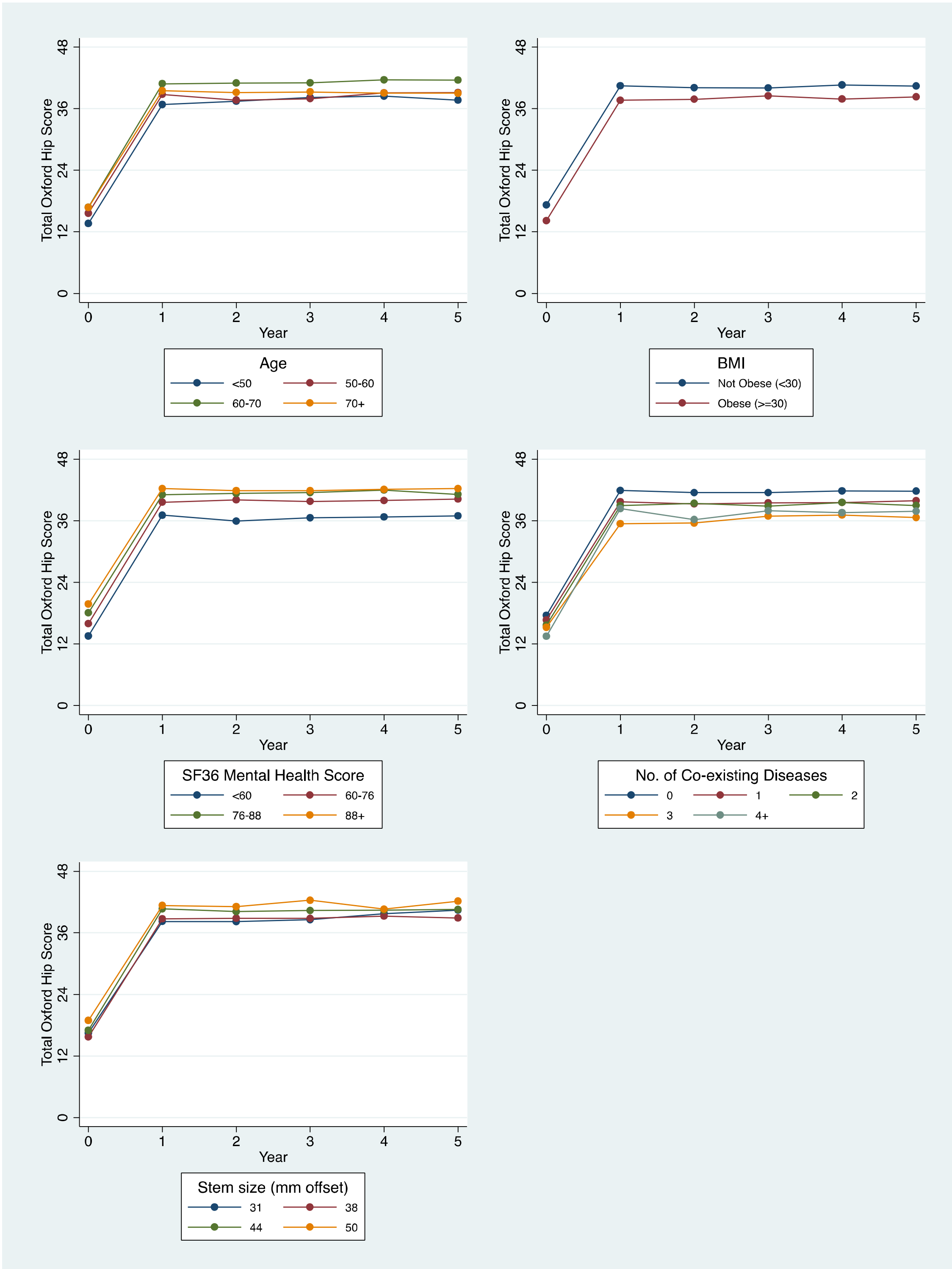


Figure 3. Change in Oxford Hip Score over time, stratified by predictive variables



SUPPLEMENTARY MATERIAL

MULTIPLE IMPUTATION METHODS

The results of complete case analyses can be biased[1]. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. Multiple imputation methods can be used to handle datasets with missing values. The risk of bias depends on the reasons why the data are missing. Missing data are seldom completely random. They are usually related, directly or indirectly, to other subject or disease characteristics, including the outcome under study[2]. If it is plausible the data are missing at random, but not completely at random, analyses on complete cases may be biased[3]. This bias can be overcome by using multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[4-6]. The first stage is to create multiple copies of the dataset with missing values replaced by imputed ones (we have created 10 copies). Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. In the second stage regression models are fitted to each of the imputed datasets and averaged together to give overall estimated associations. Standard errors are calculated using Rubins Rules. We have included all predictor variables in the multiple imputation process, together with the outcome variable as this carries information about missing values of the predictors.

INTERNAL VALIDATION

Model building in prognostic studies is usually performed using automatic stepwise variable selection procedures. However, stepwise methods have a number of disadvantages[3,7,8], where their power to select true variables is limited and estimates of predictive validity and fit may be overly optimistic. It has been suggested that of the final significant variables included in a prognostic model using backwards selection only half may be true risk factors that would be replicated by other studies[8]. To overcome these limitations, it has been suggested that bootstrapping combined with automatic backward regression can be used to provide information on model stability[7,8].

For internal validation of the regression models we therefore use a combination of multiple imputation and bootstrapping. Firstly, missing data is imputed using the ICE procedure in Stata and 10 imputed datasets created. Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. Using the 'micombine' procedure in Stata the full regression models including all predictor variables are fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[6]. Second, 200 bootstrap samples are then randomly drawn with replacement (e.g. when a patient is randomly selected their data is taken from each of the 10 imputed datasets). An automatic backward selection procedure is then applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of α = 0.157. This conservation p-value is comparable with the more complex Akaike Information Criterion (AIC)[7]. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time.

REFERENCES

1. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338(jun29_1):b2393.
2. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.
3. Harrell FE, Jr. *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*: New York: Springer; 2001.
4. Royston P. Multiple imputation of missing values. *The Stata Journal.* 2004;4(3).
5. Royston P. Multiple imputation of missing values: update. *The Stata Journal.* 2005;5(2).
6. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).
7. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.
8. Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology.* 2007;7(1):33.

Repeated measures Analysis of covariance (ANCOVA) models to identify predictors of the average OKS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Percentage retained in model	Multivariable (70% cut-off) Δ Coef (95% CI)	P-value
Patient variables					
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	100.0%	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	100.0%	0.01 (-0.11, 0.13)	0.88
Age			-		
<50	-1.44 (-3.92, 1.03)	0.25	-		
50-60	-0.96 (-2.27, 0.35)	0.15	59.5%	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	92.5%	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	98.0%	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	100.0%	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	98.0%	-1.54 (-2.45, -0.64)	0.001
Gender			-		
Male	0.00 (0.00, 0.00)	-	-		
Female	-1.30 (-2.13, -0.48)	0.002	38.0%		
Hip For Surgery			-		
Left	0.00 (0.00, 0.00)	-	-		
Right	-0.16 (-0.99, 0.67)	0.71	15.0%		
Hip Indicated			-		
Unilateral	0.00 (0.00, 0.00)	-	-		
Bilateral	-0.48 (-3.01, 2.06)	0.71	16.5%		
Occupation			-		
Heavy manual	0.00 (0.00, 0.00)	-	-		
Light manual	-0.85 (-2.50, 0.80)	0.31	38.5%		
Office / professional	2.26 (0.96, 3.57)	0.001	58.0%		
Housewife	-0.92 (-2.18, 0.33)	0.15	54.5%		
Unemployed / Retired	-0.15 (-1.04, 0.73)	0.74	43.5%		
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	98.0%	-0.90 (-1.27, -0.54)	< 0.001
Concomitant therapy used			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.73 (-2.02, 0.56)	0.27	22.0%		
Centre number			-		
1	0.00 (0.00, 0.00)	-	-		
2	-0.93 (-2.23, 0.38)	0.16	38.0%		
3	0.31 (-0.69, 1.32)	0.54	34.5%		

4	1.48 (0.28, 2.68)	0.02	44.5%		
5	-1.76 (-3.27, -0.25)	0.02	69.0%		
6	-1.18 (-2.44, 0.07)	0.06	62.0%		
7	2.22 (1.33, 3.12)	<0.001	45.0%		
Fixed flexion	0.03 (-0.01, 0.06)	0.12	47.0%		
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	100.0%	0.76 (0.46, 1.07)	< 0.001
Surgical variables					
Grade of Operator			-		
Consultant, locum consultant, assoc. specialist/staff	0.00 (0.00, 0.00)	-	-		
Fellow, senior registrar, registrar, locum registrar	-1.22 (-2.13, -0.32)	0.008	19.5%		
Surgical Approach			-		
Anterolateral	0.00 (0.00, 0.00)	-	-		
Posterior	1.03 (0.07, 1.98)	0.04	17.5%		
Patient's position			-		
Supine	0.00 (0.00, 0.00)	-	-		
Lateral	-0.99 (-2.00, 0.03)	0.06	25.5%		
Lavage System (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.70 (-2.60, 1.21)	0.47	64.5%		
Cement Pressurisation (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.37 (-1.43, 0.68)	0.48	34.5%		
Type of cement (Socket)			-		
No Cement	0.00 (0.00, 0.00)	-	-		
simplex	0.09 (-0.73, 0.91)	0.84	43.0%		
cmw1	0.63 (-0.33, 1.58)	0.2	28.5%		
palacos r	-1.21 (-2.34, -0.09)	0.03	33.0%		
Cement pressurisations (Femur)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	0.58 (-3.85, 5.01)	0.8	20.0%		
Type of cement (Femur)			-		
simplex	0.00 (0.00, 0.00)	-	-		
cmw1	1.49 (0.20, 2.79)	0.02	48.0%		

cmw3	0.23 (-0.88, 1.34)	0.68	27.5%		
palacos r	-1.26 (-2.49, -0.03)	0.04	22.0%		
palacos lv	-1.44 (-4.89, 2.01)	0.41	37.0%		
Stem size (mm offset)	0.15 (0.04, 0.26)	0.01	84.0%	0.17 (0.06, 0.28)	0.002
Femoral Head			-		
Stainless Steel	0.00 (0.00, 0.00)	-	-		
Ceramic - Zirconia/ Alumina	0.76 (-0.66, 2.17)	0.29	21.0%		
Head size			-		
22	0.00 (0.00, 0.00)	-	-		
26	1.11 (0.28, 1.93)	0.009	32.0%		
28	-0.15 (-0.97, 0.68)	0.73	29.0%		
Type of Polythene			-		
uhmwpe	0.00 (0.00, 0.00)	-	-		
duration	1.64 (0.83, 2.46)	<0.001	52.0%		
Hip Dislocation			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-4.27 (-8.05, -0.50)	0.03	74.5%	-3.77 (-7.47, -0.07)	0.05
Acetabular cup inclination (10 degrees)	-0.11 (-0.81, 0.58)	0.75	6.0%		
Acetabular cup version (10 degrees)	0.82 (0.17, 1.47)	0.01	32.5%		
Duration of Operation (Log)	-1.90 (-3.10, -0.70)	0.002	57.5%		
R2		-			17.4%
Optimism		-			0.8%
Bias-Corrected R2		-			16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Percentage: the proportion of times the variable was retained in the backward selection regression models using a P-value of 0.157 (inclusion frequency)

70% cut-off: Variables included in the final regression model are those that are retained in at least 70% of the backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OKS only

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract In the title and in the abstract as "population-based cohort study"
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found This is provided in the abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported A brief background to the study is given in the introduction section
Objectives	3	State specific objectives, including any prespecified hypotheses The overall aim of the study is stated at the end of the introduction "the aim of this study was to identify patient characteristics and intra-operative surgical factors associated with differences in: a) attained post-operative levels of pain and function, b) change (temporal trends) in symptoms of pain and function over time between pre and post-operative assessments."
Methods		
Study design	4	Present key elements of study design early in the paper This is given at the start of the methods section
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection This is described in the methods
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up The eligibility criteria are described as: "Patients were included if they were undergoing primary hip replacement with an Exeter cemented femoral stem and were willing and able to give consent to participate in the study. All eligible patients were invited to participate in the study. Patient recruitment varied between the centres but was between 80%-90% of eligible patients. The geographical area covered by the participating hospitals was wide and included both university teaching and district general hospitals that included urban as well as rural locations and represented both affluent as well as inner city suburbs. The catchment area of the four combined units included over a million people. There were 1,375 patients (1,431 hips) with a primary diagnosis of OA. The unit of analysis was the implant rather than the patient, of whom 56 had bilateral procedures. We examined 1431 THRs performed by consultant and non-consultant surgeons and using anterolateral or posterior approaches." When describing the outcome variable we state the method of follow up "Prior to surgery, patients completed an Oxford Hip Score (OHS) questionnaire with follow-up questionnaires being filled in at 1, 2, 3, 4 and 5-years post-surgery."
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable The outcomes and predictor variables are all clearly described under separate subheadings in the methods section.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group We have briefly described and referenced methods of assessment such as the Charnley

		Modification of D'Aubigne-Postel Grade questionnaire, Short Form 36 (SF-36) and Oxford Hip Score (OHS) questionnaire.
Bias	9	Describe any efforts to address potential sources of bias Efforts to address bias are covered in the statistical methods section, where we describe methods to deal with missing data and internal validation methods.
Study size	10	Explain how the study size was arrived at This was an analysis of an existing cohort of patients previously collected. We have referenced previous papers describing the cohort in greater detail, and included all available patients in the analysis for this paper stating the sample size.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why For each variable we have been clear to describe whether it is continuous or categorical. All continuous variables have been kept as continuous with the exception of age as the effect on outcome was non-linear. We state the categories of each categorical variable in the results tables.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding Statistical methods are clearly described in the paper. Generalised Estimating Equations (GEE) was used to account for clustering within the data using an exchangeable correlation matrix. Using GEE, a repeated measures linear regression model was fitted where the outcomes were the OHS at 1, 2, 3, 4 and 5-years follow up, adjusting for the pre-operative OHS as a covariate in the model. A supplementary file is provided giving greater detail of the statistical methodology.
		(*b*) Describe any methods used to examine subgroups and interactions We state in the methods section where interactions have been pre-specified. We state that for significant predictor variables, interaction terms are fitted between the predictor variable and time, where the regression models are stratified by the variable of interest to describe the change in OHS over time, for example, in those who are obese versus not obese. In the results section we describe how that for a significant surgical predictor that was identified, we hypothesised that this may be explained by an interaction with gender, and described the interaction.
		(*c*) Explain how missing data were addressed We explain in the statistical methods how missing data were addressed by using multiple imputation methods
		(*d*) If applicable, explain how loss to follow-up was addressed We explain how we have used repeated measures regression modelling whereby patients are included in the analysis if they have responded to at least 1 of the yearly follow up assessments between 1 and 5-years follow up
		(*e*) Describe any sensitivity analyses No sensitivity analyses have been conducted.
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed At the start of the results section we say how many people received hip replacement surgery, and the number included in the study and analysed
		(b) Give reasons for non-participation at each stage The only reason patients were excluded was if they did not complete at least 1 of the follow up Oxford Hip Score outcome assessments

		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders Descriptive statistics of study participants are provided in table 1
		(b) Indicate number of participants with missing data for each variable of interest The number of patients with missing data for each variable is presented in table 1.
		(c) Summarise follow-up time (eg, average and total amount) We have explained that patients were followed up at yearly intervals up to 5-years following surgery.
Outcome data	15*	Report numbers of outcome events or summary measures over time The outcome is self reported pain and function as measured by the Oxford Hip Score and Table 1 and Figure 2 summarises how this changes over time.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included Table 2 provides both adjusted and unadjusted estimates with 95% confidence intervals and p-values. We explained which predictor variables were considered for inclusion in the model in the methods section and how bootstrapping with backward selection is used to identify those remaining in the final model.
		(*b*) Report category boundaries when continuous variables were categorized Category boundaries are reported in the tables such as for age where categorised due to non-linearity.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Only the regression coefficients are reported as a measure of effect size.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses Interactions showing how the Oxford hip score changes over time, by categories of significant predictor variables are displayed in Figures 2 and 3.
Discussion		
Key results	18	Summarise key results with reference to study objectives The main findings are described at the start of the discussion section.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias We have discussed both the strengths and limitations of the study and the potential role of response bias.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence We have described how the findings of our study relate to what is already known in the literature and what this study adds.
Generalisability	21	Discuss the generalisability (external validity) of the study results We have considered the issue of generalizability highlighting that this study has used data from seven high volume centres with skilled surgeons.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based The source of funding has been described.

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

The association of patient characteristics and surgical variables on symptoms of pain and function over 5-years following primary hip replacement surgery: prospective cohort study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-002453.R1
Article Type:	Research
Date Submitted by the Author:	20-Jan-2013
Complete List of Authors:	Judge, Andrew; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Arden, Nigel; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Batra, Rajbir; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Thomas, Geraint; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Beard, David; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit Javaid, Kassim; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Cooper, Cyrus; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit; University of Southampton, MRC Lifecourse Epidemiology Unit Murray, David; University of Oxford, Oxford NIHR Musculoskeletal Biomedical Research Unit
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Surgery, Rehabilitation medicine
Keywords:	EPIDEMIOLOGY, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Musculoskeletal disorders < ORTHOPAEDIC & TRAUMA SURGERY

The association of patient characteristics and surgical variables on symptoms of pain and function over 5-years following primary hip replacement surgery: prospective cohort study

Andy Judge[1,2], Senior Statistician

Nigel K Arden[1,2], Professor in Rheumatic Diseases and Consultant Rheumatologist

Rajbir N Batra[1], Medical Statistician

Geraint Thomas[1], Clinical Fellow

David Beard[1], Professor of Musculoskeletal Sciences

M Kassim Javaid[1,2], Lecturer in Metabolic Bone Disease

Cyrus Cooper[1,2], Director and Professor of Rheumatology

David Murray[1], Professor of Orthopaedic Surgery and Consultant Orthopaedic Surgeon

Exeter Primary Outcomes Study (EPOS) group*

[1] Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Headington, Oxford, OX3 7LD, UK

[2] MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, SO16 6YD, UK

*The following people are the main investigators of the EPOS group: Professor D Murray, Mr G Andrew, Mr P Gibson, Mr J Nolan, Mr A Hamer, Mr M Fordyce and Mr K Tuson. Dr DJ Beard is a member of the EPOS steering group and scientific advisor. The following are study co-ordinators for the EPOS group: Anne Potter, Ann McGovern, Kathleen Reilly, Cathy Jenkins, Adele Cooper, Clare Darrah, Lynne Cawton, Praveen Inaparthy and Clare Pitchfork.

Corresponding Author: Andy Judge
NIHR Musculoskeletal Biomedical Research Unit,
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences,
University of Oxford,
Windmill Road, Headington,
Oxford, OX3 7LD, UK.
Fax: +44 (0) 1865 227 966
Tel: +44 (0) 1865 737 837

Email: Andrew.Judge@ndorms.ox.ac.uk

Words: Abstract 261; Text 4082

Tables: 2; **Figures:** 3.

Key words: Epidemiology, Osteoarthritis, Hip replacement, Patient reported outcome

Abstract

Objectives

To identify patient characteristics and surgical factors associated with patient reported outcomes over 5-years following primary total hip replacement (THR)

Design

Prospective cohort study

Participants and Setting

The Exeter Primary Outcomes Study of 1,431 primary hip replacements for osteoarthritis

Main outcome measures

The Oxford Hip Score (OHS) collected pre-operatively and each year up to 5-years post-operatively. Repeated measures linear regression modelling is used to identify patient and surgical predictors of outcome and describe trends over time.

Results

The majority of patient's demonstrated substantial improvement in pain/function in the first year after surgery – between one and five-years follow-up there was neither further improvement nor decline. The strongest determinant of attained post-operative OHS was the *pre-operative OHS* – those with worse pre-operative pain/function had worse post-operative pain/function. Other predictors with small but significant effects included: *femoral component offset* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50-60, younger (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* more *co-existing diseases* and worse *SF36 mental health*, was related to worse post-operative pain/function. Assessment of change in OHS between pre- and post-operative assessments revealed that patients achieved substantial and clinically relevant symptomatic improvement (change), regardless of variation in these patient and surgical factors.

Conclusions

Patients received substantial benefit from surgery, regardless of their pre-operative patient and surgical characteristics (baseline pain/function, age, BMI, comorbidities, mental health and femoral

component offset). Further research is needed to identify other factors that can improve our ability to identify patients at risk of poor outcomes from THR surgery.

ARTICLE SUMMARY

Article focus

- Total hip replacement is a common and successful surgical intervention, providing substantial relief from pain and improvement in function in patients with hip arthritis
- An important minority of patients continue to experience some pain and functional disability following surgery
- Relatively little work has been done to establish the predictors of patient reported outcomes after hip replacement, in particular the role of intra-operative surgical factors and how symptoms change over time in the mid to long term

Key messages

- The majority of patients achieved large improvement in symptoms of pain and function in the first year following surgery – there was no further improvement nor decline between one and five years
- An new finding is that a larger femoral component offset is associated with better outcomes of THR in women – this finding requires confirmation in other large cohorts
- Small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, are greatly outweighed by the substantial *change* in OHS achieved by these patients

Strengths and limitations

- Strengths include the large sample size, repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR, with data collected prospectively over 5-years with a good rate of follow up.
- Further strengths include the use of multiple imputation and bootstrapping as an internal validation technique, ensuring the predictors identified are those most likely to be replicated in external validation studies.
- Limitations are that other potential predictive variables were not available in this study, such as radiographic grade, pattern of OA, patient expectations of surgery and the type and extent of joint damage.
- Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated.

INTRODUCTION

Total hip replacement surgery (THR) is a commonly performed and successful surgical intervention, providing substantial relief from pain and improvement in functional disability in patients with hip arthritis[1-4]. Attention has turned from looking at the technical outcomes of surgery, such as prosthesis survival, to the use of patient reported outcome measures (PROMs) to see whether surgery has been successful from the patient's perspective[3,5]. Through the use of PROMs it has emerged that whilst on average the majority of patients improve after surgery, an important minority of patients continue to experience some pain and functional disability after THR, where some have no improvement or get worse[6-10].

Relatively little work has been done to establish the predictors of good or bad patient reported outcomes after THR[11]. Several potential determinants of outcomes of THR have been identified within the literature including baseline levels of pain and function[8,12-16], severity of clinical disease[13,16], age[13,16,17], gender[13,15,18], radiographic grade[13,14], education[8,12,14,18], obesity[15,17], co-morbidities[8,15], living alone[15,19], mental health[16], and patients expectations of surgery[14,20]. Little is known about the role intra-operative surgical factors may have on patient reported outcomes. In addition the majority of prior research looks at short-term outcomes and few studies have examined how symptoms change over time in the longer term.

Using a large prospective cohort of patients receiving primary THR for osteoarthritis (OA) with repeated measures of patient reported outcomes (as measured by the Oxford Hip Score) at yearly intervals over a 5-year follow up period, the aim of this study was to: a) identify patient characteristics and intra-operative surgical factors associated with differences in attained post-operative levels of pain and function, b) for variables identified as significant predictors of attained post-operative score assess change (temporal trends) in symptoms of pain and function over time between pre and post-operative assessments.

METHODS

We obtained information from the Exeter Primary Outcomes Study. Details of the study have previously been published elsewhere[21-23]. Patients were consecutively recruited between January 1999 and January 2002 at seven centres across England and Scotland. Patients underwent THR using a cemented Exeter femoral stem component (Stryker Howmedica Osteonics, Mahwah, New Jersey)[24]. A variety of cemented and uncemented acetabular components were used. Patients were included if they were undergoing primary hip replacement with an Exeter cemented femoral stem and were willing and able to give consent to participate in the study. North Western Multiple Centre Research Ethics Committee and the local research ethics committees in all the participating centres gave ethical approval for conducting the study. All eligible patients were invited to participate in the study. Patient recruitment varied between the centres but was between 80%-90% of eligible patients. The geographical area covered by the participating hospitals was wide and included both university teaching and district general hospitals that included urban as well as rural locations and represented both affluent as well as inner city suburbs. The catchment area of the four combined units included over a million people. There were 1,375 patients (1,431 hips) with a primary diagnosis of OA. The unit of analysis was the implant rather than the patient, of whom 56 had bilateral procedures. We examined 1431 THRs performed by consultant and non-consultant surgeons and using anterolateral or posterior approaches.

An extensive range of patient and intra-operative surgical factors has been collected within the EPOS study. A-priori a reduced set of variables was selected for inclusion in the analysis, based on factors previously shown within the literature to be related to patient reported outcomes of hip replacement, in addition to further variables that were considered potentially relevant – in particular the intra-operative factors, as little is known within the literature on the possible role with patient outcomes. The final set of patient and surgical factors that were selected are described below.

Patient Variables

At the pre-operative assessment information was collected on age, gender, height and weight (from which body mass index (BMI) was calculated), primary diagnosis and current occupation. Patients were asked whether they were using concomitant therapies such as oral anticoagulants, corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDS) and other analgesics. Data was collected on co-existent diseases including whether the patient had ever had deep venous thrombosis and pulmonary embolism, whether there was any evidence of urinary tract infection in the 4-weeks priori to surgery, whether the patient had any other musculoskeletal disease, whether the patient suffers from neurological, respiratory, cardiovascular, renal, and hepatic disease, and whether the patient was currently receiving treatment for any other medical conditions. An ordinal variable was created of the number of co-existent diseases a patient had at the pre-operative assessment. Fixed flexion range of motion recorded in degrees was obtained from the Charnley Modification of D'Aubigne-Postel Grade questionnaire[25,26]. Patients completed a pre-operative Short Form 36 (SF-36)[27], which measures Quality of Life generically through eight domains: physical function (PF), bodily pain (BP), general health (GH), role physical (RP), vitality (VT), social function

(SF), role emotional (RE) and mental health (MH). The lowest score, 0, corresponds to the worst possible health and 100, to the best possible health.

Surgical variables

Detailed intra-operative information was collected for each patient. This included information on the grade of operator (consultant, registrar, senior house officer), surgical approach (anterolateral, posterior) and patient position (supine, lateral). Data was available on whether or not a lavage system was used for the acetabular component, whether there was cement pressurisation for both femoral and acetabular components, the type of cement used in both the socket (none, simplex, cmw1, palacos r, other) and the femur (simplex, cmw1, cmw3, palacos r, palacos lv), the type of polyethylene used (uhmwpe, cross-linked), whether the femoral head was stainless steel or ceramic, femoral head size (22, 26 or 28 millimetres), and the femoral component offset size (35.0, 37.5, 44.0, 50.0 millimetres offset). The duration of the operation was recorded in minutes.

Outcome variable

Prior to surgery, patients completed an Oxford Hip Score (OHS) questionnaire with follow-up questionnaires being filled in at 1, 2, 3, 4 and 5-years post-surgery. Pre- and post-operative scores were completed independently by the patient prior to clinical examination. The Oxford hip score was introduced in 1996 predominantly for use in clinical trials[28]. The score is joint specific and has been assessed for reliability and validity[29]. The OHS consists of 12 questions asking patients to describe their hip pain and function during the past 4 weeks. Each question is on a Likert scale taking values from 0-4. An overall score is created by summing the responses to each of the 12 questions. A total score was created ranging from 0 to 48, where 0 is the worst possible score (most severe symptoms) and 48 the best score (least symptoms).

Statistical Methods

Stata version 11.1 (Stata, College Station, TX) was used for all statistical analyses. Potential prognostic variables included the patient and surgical variables described above. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. To overcome this bias we used multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[30] and 10 imputed datasets created. We included all predictor variables in the multiple imputation process (as listed earlier), together with the outcome variable as this carries information about missing values of the predictors. We fitted two models to describe the association of the patient and surgical variables on the following outcomes:

a) Attained post-operative OHS at 1, 2, 3, 4 and 5-years follow up

A repeated measures linear regression model was fitted where the outcomes were the OHS at 1, 2, 3, 4 and 5-years follow up, adjusting for the pre-operative OHS as a covariate in the model. Generalised Estimating Equations (GEE) was used to account for clustering within the data using an exchangeable correlation matrix. This model estimates the impact of predictors on the average OHS over the 5 follow up time points. Fractional polynomial regression modelling was used to explore evidence of non-linear relationships for continuous variables. Interaction terms were fitted between the predictor variable and time, to see if the association of the predictor on outcome changed between 1 and 5-years follow up.

b) Change in OHS between baseline and 1, 2, 3, 4 and 5-year follow ups

For variables identified as significant predictors of attained post-operative OHS, the repeated measures linear regression model is fitted, where the outcome is expanded to include the pre-operative and 1, 2, 3, 4 and 5-year post-operative OHS. Interaction terms are fitted between the predictor variable and time, to describe the change in OHS over time[31] across categories of the predictor variable, for example, in those who are obese versus not obese.

Model validation

The full regression model including all predictor variables is fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[30]. Given the extensive list of patient and surgical variables considered for inclusion in the model, we wanted to ensure that we minimised the possibility of making a type 1 error (rejecting the null hypothesis when it is true) – e.g. the chance that a variable identified as being 'significant' in this dataset may not be replicated in other samples of patients. For internal validation of the model we therefore used a combination of multiple imputation and bootstrapping[32,33] (see **supplementary file**). 200 bootstrap samples are randomly drawn with replacement. An automatic backward selection procedure is applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of $\alpha = 0.157$. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time. To assess model discrimination we use the R^2 statistic as a measure of explained variation[34].

RESULTS

Data is available on 1375 patients (1431 hips) receiving primary hip replacement surgery between January 1999 and January 2002. Of these patients 1281 (89.5%) completed a pre-operative OHS questionnaire and at least 1 of the follow up questionnaires, and were included in the analysis. 80% of patients completed the OHS at the 1-year follow up and this declined to a 70% response rate by 5-years. Baseline demographic details are described in **Table 1**. Comparing patients that did, and did not, respond to the 5-year follow up questionnaire, there were no differences in baseline pain and function as assessed by the OHS. Differences were observed where those that responded were younger, less likely to be unemployed/retired, and had better pre-operative SF36 mental health scores.

Histograms of the distribution of OHS at baseline, follow-up, and the absolute difference in scores (**Figure 1**) highlight that at the 1 and 5-year follow ups the score is negatively skewed to the left, suggesting the majority of patients achieve improvement in pain and function. The histograms of the difference in scores highlight that almost all patients get better with only a small minority getting worse or receiving no improvement (2.3% by 1-year and 1.2% by 5-years). The change in OHS over time from the repeated measures regression model is displayed in **Figure 2**. This demonstrates that regardless of the level of pre-operative OHS, patients achieved substantial improvement in pain and function following surgery. Those with the worst pre-operative scores achieved the greatest improvement (28.8 point change in those with pre-operative OHS < 5), however patients with the best pre-operative scores still achieved substantial improvement (change of 10.6 points in those with pre-operative OHS > 30). Interestingly, between 1 and 5-years follow up a steady state is reached where there is no further improvement nor decline, with a non-significant trend between OHS and time (p=0.88).

A number of variables were identified as important predictors of attained post-operative OHS (**Table 2**). The strongest determinant of outcomes was the baseline OHS. Increasing baseline OHS (better pre-operative pain/function) was associated with increasing follow-up OHS (better post-operative pain/function). The effect of age was non-linear, where compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcomes. Increasing BMI, patients with a greater number of co-existing diseases prior to surgery, and those with worse pre-operative SF36 mental health scores, also had worse outcomes. The surgical predictor we identified was femoral component offset, where patients with larger offset size (offset of 44 or more) had significantly better outcomes. We hypothesised that this may be explained by an interaction with gender, where a larger offset is used in men. A significant interaction was observed, where no association was observed in men, whilst in women those with larger offsets had better outcomes. The effect of surgical approach was significant at the 1-year follow-up, where the anterolateral approach had better outcomes than posterior (difference in 1-year OHS of 2.2 units 95%CI (1.1 to 3.3), however the effect size attenuated over time and became no longer significant between 3 to 5 years follow up.

Figure 3 describes the change in OHS over time stratified according to each of the predictor variables we identified in this study. The graphs highlight that whilst there are small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, this is greatly outweighed by the substantial *change* in OHS achieved by these patients, regardless of whether they are old or young, obese or not obese. These patient groups still receive great benefit from surgery.

Assessing the discriminatory ability of the final model, including the baseline OHS alone explained 10.3% of the variability in outcome. The final predictive model including the patient and surgical variables explained 16.6%. This suggests that although we have identified significant patient and surgical predictors of outcome, they have smaller effects, and explain little of the variability in attained post-operative OHS above that of the baseline score.

DISCUSSION

Main findings

Within a large prospective cohort of patients receiving primary THR in the UK, we identified a number of predictors of differences in *attained* post-operative pain and function. Determinants included: *pre-operative pain and function* – those with worse pre-operative pain/function had worse post-operative pain/function; *femoral component offset* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* a greater number of *co-existing diseases* and worse *SF36 mental health* at the time of surgery, was related to worse post-operative pain and function. The strongest determinant of outcome was the baseline score with the patient and surgical variables contributing small but statistically significant effects.

Assessing the relationship of *change* in symptoms of pain and function between pre and post-operative assessments for predictor variables in the final model revealed that patients achieved large symptomatic improvement (change), regardless of differences in pre-operative patient and surgical factors. The change in symptoms greatly outweighs any differences in attained post-operative score – patients achieved great benefit from surgery regardless of factors such as their age and BMI at the time of surgery. Exploring temporal trends in symptoms of pain and function over time demonstrated that there was little further improvement nor decline in the short to mid-term (between 1 and 5-years follow-up) where a steady symptomatic state was reached.

Strengths and limitations

The strength of this study lies in its large sample size and repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR[28,29], with data collected prospectively over 5-years with a good rate of follow up. The use of multiple imputation and bootstrapping as an internal validation technique is a strength of this study, ensuring the predictors we identified are those most likely to be replicated in external validation studies, and not chance significant findings that are anomalies of our dataset. Within this study the aim was to identify predictors of differences in attained post-operative pain and function – it should be noted that predictors we identified of attained health state may not necessarily be the same as predictors of improvement (change) in symptoms. It should be noted that it remains unclear which measure of the two outcomes of change in status or attained status is most appropriate for judging the value of surgery[35] – THR could either be viewed as intended to preserve the highest levels of pain function, or alternatively to maximise the potential for symptomatic improvement. Limitations are that other potential predictive variables were not available in this study, predominantly radiographic factors such as x-ray grade and pattern of OA, and other factors including patient expectations of surgery and the type and extent of joint damage. Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated in this study.

Within this study histograms of the absolute difference (change) in scores between pre and post-operative assessments highlight that almost all patients get better with only a small minority getting worse or receiving no improvement (2.3% by 1-year and 1.2% by 5-years). This is set in the context that data in the study come from seven high volume centres with skilled surgeons across England and Scotland and raise the possibility that outcomes in this study may be better than expected. Comparing this finding with data from other cohorts, in the EUROHIP study consisting of THR patients across 20 European orthopaedic centres, based on the change in WOMAC score at 12-months 58 (6.9%) of 845 patients had no change or worsening of symptoms – although this varied by region and in patients from the UK, only 4 (3.5%) of 111 patients had no change or got worse. Data from the Elective Orthopaedic Centre (EOC) database that include primary hip replacements performed in four acute NHS trusts in the UK in South West London, 88 (5.1%) of 1711 patients symptoms got worse by 6-months based on the OHS, and in data from St. Helier district general hospital, Carshalton, UK, 14 (2.3%) of 619 patients had no change or worsening in OHS at 12-months. Hence, whilst it is plausible that outcomes in the EPOS study may be better than expected, the findings are consistent with other cohorts.

What is already known

In our study an important new finding was that a larger femoral component offset is associated with better outcomes of THR. We hypothesised that the effect of offset size may be explained by an interaction with gender, as men have larger offsets. A significant interaction was observed, where no association was observed in men, whilst in women those with larger offsets had better outcomes. The choice of offset can affect hip stability as well as abduction strength, potentially resulting in abnormal (trendelenburg) gait. Component offset may be preoperatively templated, although common offsets are often assumed (i.e. 37.5mm for females and 44.0mm males). There is greater potential to decrease offset in females, partly because of the above assumption, and it is sometimes difficult to use the larger offset components because a smaller femoral canal diameter precludes their use. The choice of offset for the femoral component has not changed since this study was conducted and these finding are generalizable to clinical practice. We are not aware of any data in the literature describing the relationship of intra-operative surgical factors on patient reported outcomes. Within the literature, data exists on the relationship between head size and failure of a THR, whereby larger head size (40mm versus 28mm) with a ceramic-on-ceramic bearing surface was associated with lower 5-year revision rates[36]. This is though to be related to larger diameter heads increasing fluid-film lubrication, in turn reducing wear, and decreasing dislocation rates[36]. However, femoral component offset and head size are independent factors that are unrelated to one another. Within this study we found no association between head size and patient reported outcomes – this is unsurprising since the head sizes used were 22, 26 and 28mm and common to orthopaedic practice at the time this study took place. Thinner liners are now manufactured allowing larger head sizes to be used in acetabular components of the same size.

Consistent with others in the literature we observed that worse pre-operative mental health was a predictor of poor outcome[16] as were greater numbers of pre-operative co-morbidities[8,15]. We found that older age and higher BMI was associated with worse patient reported outcomes. Within the literature whilst some authors conclude that older age and increasing BMI area associated with worse outcomes, others have found no evidence of an association[7-9,12-19,31,37]. This is in line with the conclusions of large literature reviews stating that such factors are not strong predictors of outcome[11]. The findings are important to decision making as physicians often advise patients they are too old or obese to receive THR[11,38]. We can conclude that in relation to patient reported outcomes of THR, even if some groups fare less well after THR, it does not mean these patients don't get benefit from surgery[11]. Expectations of the patients may also play a role, where for example, what a young person wants to achieve in functional rehabilitation is different to an older person, for whom a lower functional score may be perfectly acceptable. It also is well known within the literature that patients with better pre-operative pain and function achieve better *attained* post-operative pain and function, and that patients with worse pre-operative pain and function get the greatest *change* (symptomatic improvement) between baseline and follow up[12,30,38-40].

Whilst the effects of pre-operative patient characteristics including age, BMI, co-existing diseases and mental health may already be known within the existing literature, what is novel about this study is contrasting the effects of these factors on attained levels of post-operative scores with a graphical representation of change (improvement in scores). Our findings highlight that small but significant differences in attained scores are greatly outweighed by the fact these patient groups get great benefit from surgery (in terms of substantial change in symptoms of pain and function), regardless of differences in pre-operative patient characteristics.

Within this study, although we examined a wide range of patient and intra-operative surgical factors we were only able to explain around 17% of the variability in patient reported outcomes of THR. Data were not available to us on other factors known within the literature to be predictive of outcomes of THR such as radiographic information and the presence of musculoskeletal disease in other joints. It has previously been shown that patients with pain in other joints[39] have worse outcomes. Patients with worse pre-operative x-rays are more likely to improve[13,14]. Worse outcome have been observed in patients who live alone[15,19,] those with less social support[31] and patients with lower educational attainment[8,12,14,18]. Patients with greater pre-operative expectations of surgery have been observed to have better outcomes[14,20]. Whilst inclusion of such factors would help to improve the predictive ability on the model, as individual factors of interest, it is likely we would see the same pattern observed whereby small differences are seen in attained score but patients achieve benefit.

What this study adds

Within this study we have identified a number of patient and surgical predictors of attained post-operative pain and function following THR surgery. These predictors remain related to outcome over the short to mid term. An important new finding was that larger femoral component offset was associated with better post-operative pain and function in women. This finding implies that greater consideration should be given to measuring and deciding upon the choice of offset in women as there is potential to undersize. Whilst age, BMI, co-existing diseases, mental health and femoral component offset were associated with small but significant differences in attained pain and function, analyses of change demonstrate that these patient still achieve substantial symptomatic benefit from surgery regardless of differences in these pre-operative factors. The findings will be important to inform patient and clinician decision regarding the likely outcomes of surgery for these patient groups. Although we have assessed a wide range of patient characteristics and intra-operative surgical factors there is still uncertainty as to the cause of variation in outcomes of hip replacement. There is a need to focus on issues that remain unclear such as the effect of soft tissue and the severity and pattern of OA. Further research is needed using more detailed measures of existing predictive variables, and identification of other factors beyond those observed in this study that explain a greater proportion of the variability in outcome to improve our ability to identify patients at risk of poor outcomes from THR surgery.

Funding statement

This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10064). The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This study has been supported by Stryker Howmedica Osteonics (Newbury, UK). Support was also received from the Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Orthopaedic Centre, University of Oxford and the UK Medical Research Council, MRC Lifecourse Epidemiology Unit, University of Southampton.

Competing interest declaration

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: AJ, RNB and GT have no conflicts of interest; MKJ, NKA, and CC have received honorariums, held advisory board positions (which involved receipt of fees), and received consortium research grants, respectively, from: Novartis and Alliance for Better Health and Lilly; Merck, Merck Sharp and

Dohme, Roche, Novartis, Smith and Nephew, Q-MED, Nicox, Servier, GlaxoSmithKline, Schering-Plough, Pfizer, and Rottapharm; and Alliance for Better Bone Health, Amgen, Novartis, Merck Sharp and Dohme, Servier, Eli Lilly, and GlaxoSmithKline; DB has held an independent consultancy with ICNet and Stryker, and has been a grant holder on a Genzyme-funded study; DM has received royalties from Biomet and Wright Medical Technology, Inc, and receives research support from DePuy, A Johnson & Johnson Company, Stryker, Zimmer and Wright Medical Technology, Inc; they have no other relationships or activities that could appear to have influenced the submitted work.

Contributors

AJ NKA RB GT DB MKJ CC DM: (1) substantial contributions to conception and design, analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published.

Ethics

Ethical approval was obtained from the Salford and Trafford Research Ethics Committee (Project No: - 98105 – MREC 98/8/20 UK Multicentre Exeter Primary Outcome Study). Informed written consent was obtained from all participants.

Data Sharing

No additional data available.

REFERENCES

1. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (1). *N. Engl. J. Med.* 1990;323(11):725-731.

2. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (2). *N. Engl. J. Med.* 1990;323(12):801-807.

3. Ethgen O, Bruyere O, Richy F, et al. Health-Related Quality of Life in Total Hip and Total Knee Arthroplasty. A Qualitative and Systematic Review of the Literature. *J. Bone Joint Surg. Am.* 2004;86(5):963-974.

4. Cooper C, Arden NK. Excess mortality in osteoarthritis. *BMJ.* 2011;342.

5. Professor the Lord Darzi of Denham Kbe HonFREng FmedSci. *High Quality Care for All. NHS Next Stage Review Final Report - Summary*: Department of Health;2008.

6. Williams O, Fitzpatrick R, Hajat S, et al. Mortality, morbidity, and 1-year outcomes of primary elective total hip arthroplasty. *J. Arthroplasty.* 2002;17(2):165-171.

7. Nilsdotter AK, Petersson IF, Roos EM, et al. Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study. *Ann. Rheum. Dis.* 2003;62(10):923-930.

8. MacWilliam CH, Yood MU, Verner JJ, et al. Patient-related risk factors that predict poor outcome after total hip replacement. *Health Serv. Res.* 1996;31(5):623-638.

9. Judge A, Cooper C, Williams S, Dreinhoefer K, Dieppe P. Patient-Reported Outcomes One Year After Primary Hip Replacement in a European Collaborative Cohort. *Arthritis Care Res.* 2010;62(4):480-488.

10. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. *Curr. Opin. Rheumatol.* 2006;18(5):526-530.

11. Santaguida PL, Hawker GA, Hudak PL, et al. Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can. J. Surg.* 2008;51(6):428-436.

12. Fortin PR, Clarke AE, Joseph L, et al. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. *Arthritis Rheum.* 1999;42(8):1722-1728.

13. Cushnaghan J, Coggon D, Reading I, et al. Long-term outcome following total hip arthroplasty: a controlled longitudinal study. *Arthritis Rheum.* 2007;57(8):1375-1380.

14. Judge A, Cooper C, Arden NK, et al. Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres. *Osteoarthritis Cartilage.* 2011;19(6):659-667.

15. Jones CA, Voaklander DC, Johnston DW, et al. The Effect of Age on Pain, Function, and Quality of Life After Total Hip and Knee Arthroplasty. *Arch. Intern. Med.* 2001;161(3):454-460.

16. Quintana JM, Escobar A, Aguirre U, et al. Predictors of health-related quality-of-life change after total hip arthroplasty. *Clin. Orthop.* 2009;467(11):2886-2894.

17. Braeken AM, Lochhaas-Gerlach JA, Gollish JD, et al. Determinants of 6-12 month postoperative functional status and pain after elective total hip replacement. *Int. J. Qual. Health Care.* 1997;9(6):413-418.

18. Rissanen P, Aro S, Sintonen H, et al. Quality of life and functional ability in hip and knee replacements: a prospective study. *Qual. Life Res.* 1996;5(1):56-64.

19. Hopman WM, Mantle M, Towheed TE, et al. Determinants of health-related quality of life following elective total hip replacement. *Am. J. Med. Qual.* 1999;14(3):110-116.

20. Gandhi R, Davey JR, Mahomed N. Patient expectations predict greater pain relief with joint arthroplasty. *J. Arthroplasty.* 2009;24(5):716-721.

21. Andrew JG, Palan J, Kurup HV, et al. Obesity in total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2008;90-B(4):424-429.

22. Palan J, Beard D, Murray D, et al. Which Approach for Total Hip Arthroplasty: Anterolateral or Posterior? *Clin. Orthop.* 2009;467(2):473-477.

23. Palan J, Gulati A, Andrew JG, et al. The trainer, the trainee and the surgeons' assistant: clinical outcomes following total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2009;91-B(7):928-934.

24. Hossain M, Parfitt D, Beard D, et al. Does pre-operative psychological distress affect patient satisfaction after primary total hip arthroplasty? *BMC Musculoskeletal Disorders.* 2011;12(1):122.

25. Charnley J. The Long-Term Results of Low-Friction Arthroplasty of the Hip Performed as a Primary Intervention. *J. Bone Joint Surg. Br.* February 1, 1972 1972;54-B(1):61-76.

26. D'Aubigne RM, Postel M. Functional Results of Hip Arthroplasty With Acrylic Prosthesis. *J. Bone Joint Surg. Am.* June 1, 1954 1954;36(3):451-475.

27. Ware JE. *How to score the revised MOS Short Form Health Scales*: Boston: The Health Institute, New England Medical Centre Hospital;1988.

28. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J. Bone Joint Surg. Br.* 2007;89-B(8):1010-1014.

29. Dawson J, Fitzpatrick R, Carr A, et al. Questionnaire on the perceptions of patients about total hip replacement. *J. Bone Joint Surg. Br.* 1996;78-B(2):185-190.

30. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).

31. Fitzgerald JD, Orav EJ, Lee TH, et al. Patient quality of life during the 12 months following joint replacement surgery. *Arthritis Rheum.* 2004;51(1):100-109.

32. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.

33. Heymans M, van Buuren S, Knol D, et al. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology.* 2007;7(1):33.

34. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

35. Losina E, Katz JN. Total knee replacement: pursuit of the paramount result. *Rheumatology (Oxford).* 2012.

36. Smith AJ, Dieppe P, Vernon K, et al. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *The Lancet.*379(9822):1199-1204.

37. Harse JD, Holman CD. Charlson's Index was a poor predictor of quality of life outcomes in a study of patients following joint replacement surgery. *J. Clin. Epidemiol.* 2005;58(11):1142-1149.

38. Hawker G, Wright J, Coyte P, et al. Health-related quality of life after knee replacement. *J. Bone Joint Surg. Am.* 1998;80(2):163-173.

39. Judge A, Javaid MK, Arden NK, et al. A clinical tool to identify patients who are most likely to receive long term improvement in physical function after total hip arthroplasty. *Arthritis Care Res.* 2012:n/a-n/a.

Table 1. Descriptive statistics and comparison of those who did, and did not, complete the 5-year follow up questionnaire

Variable	Missing	Baseline (n = 1431)	Non Responders at Year 5	Responders at Year 5	P-value
Oxford Hip Score					
Pre-operative	70 (4.9%)	16.4 (7.8)	16.1 (8.2)	16.5 (7.6)	0.35
1-year post-op	274 (19.1%)	43 (36, 46)	-	-	
2-year post-op	316 (22.1%)	43 (36, 47)	-	-	
3-year post-op	368 (25.7%)	43 (36, 47)	-	-	
4-year post-op	430 (30.0%)	44 (36, 47)	-	-	
5-year post-op	396 (27.7%)	43 (37, 47)	-	-	
Patient characteristics					
Age	9 (1%)	70.0 (63.9, 76.1)	73.6 (66.2, 79.3)	68.8 (62.7, 74.5)	< 0.001
BMI	95 (7%)	27.4 (4.9)	27.1 (4.9)	27.6 (4.8)	0.077
Gender	7 (0%)				0.46
Male		537 (38%)	171 (39%)	366 (37%)	
Female		887 (62%)	266 (61%)	621 (63%)	
Occupation	0 (0%)				0.005
Heavy manual		41 (3%)	8 (2%)	33 (3%)	
Light manual		89 (6%)	20 (5%)	69 (7%)	
Office / professional		107 (7%)	21 (5%)	86 (9%)	
Housewife		187 (13%)	67 (15%)	120 (12%)	
Unemployed / retired		1007 (70%)	325 (74%)	682 (69%)	
No. of Co-existing Diseases	0 (0%)				0.94
0		431 (30%)	136 (31%)	295 (30%)	
1		498 (35%)	147 (33%)	351 (35%)	
2		315 (22%)	99 (22%)	216 (22%)	
3		140 (10%)	43 (10%)	97 (10%)	
4		47 (3%)	16 (4%)	31 (3%)	
Concomitant therapy used	8 (1%)				0.84
No		104 (7%)	31 (7%)	73 (7%)	
Yes		1319 (93%)	406 (93%)	913 (93%)	
SF36 Mental Health Score	515 (36%)	74 (60, 88)	72 (52, 88)	76 (60, 88)	0.046

Cells represent either: number (percentage), mean (standard deviation), median (interquartile range)
† T-tests are used for continuous variables and chi-squared tests for categorical variables.
Where continuous variables were not normally distributed, a non-parametric t-test (Kruskal-Wallis) was used.
Fishers exact test is used where expected counts are less than 5.

Table 2. Repeated measures Analysis of Covariance (ANCOVA) models to identify predictors of the average OHS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Multivariable Δ Coef (95% CI)	P-value
Patient variables				
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	0.01 (-0.11, 0.13)	0.88
Age				
<50	-1.44 (-3.92, 1.03)	0.25		
50-60	-0.96 (-2.27, 0.35)	0.15	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	-1.54 (-2.45, -0.64)	0.001
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	-0.90 (-1.27, -0.54)	< 0.001
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	0.76 (0.46, 1.07)	< 0.001
Surgical variables				
Femoral component offset size (mm offset)	0.15 (0.04, 0.26)	0.01	0.17 (0.06, 0.28)	0.002
R^2				17.4%
Optimism				0.8%
Bias-Corrected R^2				16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Variables included in the final regression model are those that are retained in at least 70% of the 200 bootstrap backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OHS only

FIGURE LEGENDS

Figure 1. Distribution of OHS at baseline, follow-up and absolute difference in scores

Figure 2. Change in Oxford Hip Score over time, stratified by baseline score

Figure 3. Change in Oxford Hip Score over time, stratified by predictive variables

The association of patient characteristics and surgical variables on symptoms of pain and function over 5-years following primary hip replacement surgery: ~~population-based~~prospective cohort study

Andy Judge[1,2], Senior Statistician

Nigel K Arden[1,2], Professor in Rheumatic Diseases and Consultant Rheumatologist

Rajbir N Batra[1], Medical Statistician

Geraint Thomas[1], Clinical Fellow

David Beard[1], Professor of Musculoskeletal Sciences

M Kassim Javaid[1,2], Lecturer in Metabolic Bone Disease

Cyrus Cooper[1,2], Director and Professor of Rheumatology

David Murray[1], Professor of Orthopaedic Surgery and Consultant Orthopaedic Surgeon

Exeter Primary Outcomes Study (EPOS) group*

[1] Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Windmill Road, Headington, Oxford, OX3 7LD, UK

[2] MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, SO16 6YD, UK

*The following people are the main investigators of the EPOS group: Professor D Murray, Mr G Andrew, Mr P Gibson, Mr J Nolan, Mr A Hamer, Mr M Fordyce and Mr K Tuson. Dr DJ Beard is a member of the EPOS steering group and scientific advisor. The following are study co-ordinators for the EPOS group: Anne Potter, Ann McGovern, Kathleen Reilly, Cathy Jenkins, Adele Cooper, Clare Darrah, Lynne Cawton, Praveen Inaparthy and Clare Pitchfork.

Corresponding Author: Andy Judge
NIHR Musculoskeletal Biomedical Research Unit,
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences,
University of Oxford,
Windmill Road, Headington,
Oxford, OX3 7LD, UK.
Fax: +44 (0) 1865 227 966

Tel: +44 (0) 1865 737 837

Email: Andrew.Judge@ndorms.ox.ac.uk

Words: Abstract ~~262~~261; Text ~~3430~~ 4082

Tables: 2; **Figures:** 3.

Key words: Epidemiology, Osteoarthritis, Hip replacement, Patient reported outcome

Abstract

Objectives

To identify patient characteristics and surgical factors associated with patient reported outcomes over 5-years following primary total hip replacement (THR)

Design

Prospective ~~population-based~~ cohort study

Participants and Setting

The Exeter Primary Outcomes Study of 1,431 primary hip replacements for osteoarthritis

Main outcome measures

The Oxford Hip Score (OHS) collected pre-operatively and each year up to 5-years post-operatively. Repeated measures linear regression modelling is used to identify patient and surgical predictors of outcome and describe trends over time.

Results

The majority of patient's demonstrated substantial improvement in pain/function in the first year after surgery – between one and five-years follow-up there was neither further improvement nor decline. The strongest determinant of attained post-operative OHS was the *pre-operative OHS* – those with worse pre-operative pain/function had worse post-operative pain/function. Other predictors with small but significant effects included: *femoral component* ~~size~~ *offset* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50-60, younger (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* more *co-existing diseases* and worse *SF36 mental health*, was related to worse post-operative pain/function. Assessment of change in OHS between pre- and post-operative assessments revealed that patients achieved substantial and clinically relevant symptomatic improvement (change), regardless of variation in these patient and surgical factors.

Conclusions

Patients received substantial benefit from surgery, regardless of their pre-operative patient and surgical characteristics (baseline pain/function, age, BMI, comorbidities, mental health and femoral

component ~~size~~offset). Further research is needed to identify other factors that can improve our ability to identify patients at risk of poor outcomes from THR surgery.

ARTICLE SUMMARY

Article focus

- Total hip replacement is a common and successful surgical intervention, providing substantial relief from pain and improvement in function in patients with hip arthritis
- An important minority of patients continue to experience some pain and functional disability following surgery
- Relatively little work has been done to establish the predictors of patient reported outcomes after hip replacement, in particular the role of intra-operative surgical factors and how symptoms change over time in the mid to long term

Key messages

- The majority of patients achieved large improvement in symptoms of pain and function in the first year following surgery – there was no further improvement nor decline between one and five years
- An new finding is that a larger femoral component ~~size (~~offset~~)~~ is associated with better outcomes of THR in women – this finding requires confirmation in other large cohorts
- Small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, are greatly outweighed by the substantial *change* in OHS achieved by these patients

Strengths and limitations

- Strengths include the large sample size, repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR, with data collected prospectively over 5-years with a good rate of follow up.
- Further strengths include the use of multiple imputation and bootstrapping as an internal validation technique, ensuring the predictors identified are those most likely to be replicated in external validation studies.
- Limitations are that other potential predictive variables were not available in this study, such as radiographic grade, pattern of OA, patient expectations of surgery and the type and extent of joint damage.
- Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated.

INTRODUCTION

Total hip replacement surgery (THR) is a commonly performed and successful surgical intervention, providing substantial relief from pain and improvement in functional disability in patients with hip arthritis[1-4]. Attention has turned from looking at the technical outcomes of surgery, such as prosthesis survival, to the use of patient reported outcome measures (PROMs) to see whether surgery has been successful from the patient's perspective[3,5]. Through the use of PROMs it has emerged that whilst on average the majority of patients improve after surgery, an important minority of patients continue to experience some pain and functional disability after THR, where some have no improvement or get worse[6-10].

Relatively little work has been done to establish the predictors of good or bad patient reported outcomes after THR[11]. Several potential determinants of outcomes of THR have been identified within the literature including baseline levels of pain and function[8,12-16], severity of clinical disease[13,16], age[13,16,17], gender[13,15,18], radiographic grade[13,14], education[8,12,14,18], obesity[15,17], co-morbidities[8,15], living alone[15,19], mental health[16], and patients expectations of surgery[14,20]. Little is known about the role intra-operative surgical factors may have on patient reported outcomes. In addition the majority of prior research looks at short-term outcomes and few studies have examined how symptoms change over time in the longer term.

Using a large prospective cohort of patients receiving primary THR for osteoarthritis (OA) with repeated measures of patient reported outcomes (as measured by the Oxford Hip Score) at yearly intervals over a 5-year follow up period, the aim of this study was to: a) identify patient characteristics and intra-operative surgical factors associated with differences in: a) attained post-operative levels of pain and function, b) for variables identified as significant predictors of attained post-operative score assess change (temporal trends) in symptoms of pain and function over time between pre and post-operative assessments.

METHODS

We obtained information from the Exeter Primary Outcomes Study. Details of the study have previously been published elsewhere[21-23]. Patients were consecutively recruited between January 1999 and January 2002 at seven centres across England and Scotland. Patients underwent THR using a cemented Exeter femoral stem component (Stryker Howmedica Osteonics, Mahwah, New Jersey)[24]. A variety of cemented and uncemented acetabular components were used. Patients were included if they were undergoing primary hip replacement with an Exeter cemented femoral stem and were willing and able to give consent to participate in the study. North Western Multiple Centre Research Ethics Committee and the local research ethics committees in all the participating centres gave ethical approval for conducting the study. All eligible patients were invited to participate in the study. Patient recruitment varied between the centres but was between 80%-90% of eligible patients. The geographical area covered by the participating hospitals was wide and included both university teaching and district general hospitals that included urban as well as rural locations and represented both affluent as well as inner city suburbs. The catchment area of the four combined units included over a million people. There were 1,375 patients (1,431 hips) with a primary diagnosis of OA. The unit of analysis was the implant rather than the patient, of whom 56 had bilateral procedures. We examined 1431 THRs performed by consultant and non-consultant surgeons and using anterolateral or posterior approaches.

An extensive range of patient and intra-operative surgical factors has been collected within the EPOS study. A-priori a reduced set of variables was selected for inclusion in the analysis, based on factors previously shown within the literature to be related to patient reported outcomes of hip replacement, in addition to further variables that were considered potentially relevant – in particular the intra-operative factors, as little is known within the literature on the possible role with patient outcomes. The final set of patient and surgical factors that were selected are described below.

Patient Variables

At the pre-operative assessment information was collected on age, gender, height and weight (from which body mass index (BMI) was calculated), primary diagnosis and current occupation. Patients were asked whether they were using concomitant therapies such as oral anticoagulants, corticosteroids, non-steroidal anti-inflammatory drugs (NSAIDS) and other analgesics. Data was collected on co-existent diseases including whether the patient had ever had deep venous thrombosis and pulmonary embolism, whether there was any evidence of urinary tract infection in the 4-weeks priori to surgery, whether the patient had any other musculoskeletal disease, whether the patient suffers from neurological, respiratory, cardiovascular, renal, and hepatic disease, and whether the patient was currently receiving treatment for any other medical conditions. An ordinal variable was created of the number of ~~pre-operative~~ co-existent diseases a patient had at the pre-operative assessment~~, which included deep venous thrombosis, pulmonary embolism, urinary tract infection, other musculoskeletal disease, neurological, respiratory, cardiovascular, renal, hepatic disease or treatment for other medical conditions~~. Fixed flexion range of motion recorded in degrees was obtained from the Charnley Modification of D'Aubigne-Postel Grade questionnaire[25,26].

Patients completed a pre-operative Short Form 36 (SF-36)[27], which measures Quality of Life generically through eight domains: physical function (PF), bodily pain (BP), general health (GH), role physical (RP), vitality (VT), social function (SF), role emotional (RE) and mental health (MH). The lowest score, 0, corresponds to the worst possible health and 100, to the best possible health.

Surgical variables

Detailed intra-operative information was collected for each patient. This included information on the grade of operator (consultant, registrar, senior house officer), surgical approach (anterolateral, posterior) and patient position (supine, lateral). Data was available on whether or not a lavage system was used for the acetabular component, whether there was cement pressurisation for both femoral and acetabular components, the type of cement used in both the socket (none, simplex, cmw1, palacos r, other) and the femur (simplex, cmw1, cmw3, palacos r, palacos lv), the type of polyethylene used (uhmwpe, cross-linked), whether the femoral head was stainless steel or ceramic, femoral head size (22, 26 or 28 millimetres), and the femoral component ~~stem~~ offset size (35.0, 37.5, 44.0, 50.0 millimetres offset). The duration of the operation was recorded in minutes.

Outcome variable

Prior to surgery, patients completed an Oxford Hip Score (OHS) questionnaire with follow-up questionnaires being filled in at 1, 2, 3, 4 and 5-years post-surgery. Pre- and post-operative scores were completed independently by the patient prior to clinical examination. The Oxford hip score was introduced in 1996 predominantly for use in clinical trials[28]. The score is joint specific and has been assessed for reliability and validity[29]. The OHS consists of 12 questions asking patients to describe their hip pain and function during the past 4 weeks. Each question is on a Likert scale taking values from 0-4. An overall score is created by summing the responses to each of the 12 questions. A total score was created ranging from 0 to 48, where 0 is the worst possible score (most severe symptoms) and 48 the best score (least symptoms).

Formatted: Font: Not Italic

Statistical Methods

Stata version 11.1 (Stata, College Station, TX) was used for all statistical analyses. Potential prognostic variables included the patient and surgical variables described above. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. To overcome this bias we used multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[30] and 10 imputed datasets created. We included all predictor variables in the multiple imputation process (as listed earlier), together with the outcome variable as this carries information about missing values of the predictors. We fitted two models to describe the association of the patient and surgical variables on the following outcomes:

a) Attained post-operative OHS at 1, 2, 3, 4 and 5-years follow up

A repeated measures linear regression model was fitted where the outcomes were the OHS at 1, 2, 3, 4 and 5-years follow up, adjusting for the pre-operative OHS as a covariate in the model. Generalised Estimating Equations (GEE) was used to account for clustering within the data using an exchangeable correlation matrix. This model estimates the impact of predictors on the average OHS over the 5 follow up time points. Fractional polynomial regression modelling was used to explore evidence of non-linear relationships for continuous variables. Interaction terms were fitted between the predictor variable and time, to see if the association of the predictor on outcome changed between 1 and 5-years follow up.

b) Change in OHS between baseline and 1, 2, 3, 4 and 5-year follow ups

For variables identified as significant predictors of attained post-operative OHS, the repeated measures linear regression model is fitted, where the outcome is expanded to include the pre-operative and 1, 2, 3, 4 and 5-year post-operative OHS. Interaction terms are fitted between the predictor variable and time, to describe the change in OHS over time[31] across categories of the predictor variable, for example, in those who are obese versus not obese.

Model validation

The full regression model including all predictor variables is fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[30]. Given the extensive list of patient and surgical variables considered for inclusion in the model, we wanted to ensure that we minimised the possibility of making a type 1 error (rejecting the null hypothesis when it is true) – e.g. the chance that a variable identified as being 'significant' in this dataset may not be replicated in other samples of patients. For internal validation of the model we therefore used a combination of multiple imputation and bootstrapping[32,33] (see **supplementary file**). 200 bootstrap samples are randomly drawn with replacement. An automatic backward selection procedure is applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of $\alpha = 0.157$. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time. To assess model discrimination we use the R^2 statistic as a measure of explained variation[34].

RESULTS

Data is available on 1375 patients (1431 hips) receiving primary hip replacement surgery between January 1999 and January 2002. Of these patients 1281 (89.5%) completed a pre-operative OHS questionnaire and at least 1 of the follow up questionnaires, and were included in the analysis. 80% of patients completed the OHS at the 1-year follow up and this declined to a 70% response rate by 5-years. Baseline demographic details are described in **Table 1**. Comparing patients that did, and did not, respond to the 5-year follow up questionnaire, there were no differences in baseline pain and function as assessed by the OHS. Differences were observed where those that responded were younger, less likely to be unemployed/retired, and had better pre-operative SF36 mental health scores.

Histograms of the distribution of OHS at baseline, follow-up, and the absolute difference in scores (**Figure 1**) highlight that at the 1 and 5-year follow ups the score is negatively skewed to the left, suggesting the majority of patients achieve improvement in pain and function. The histograms of the difference in scores highlight that almost all patients get better with only a small minority getting worse or receiving no improvement (2.3% by 1-year and 1.2% by 5-years). The change in OHS over time from the repeated measures regression model is displayed in **Figure 2**. This demonstrates that regardless of the level of pre-operative OHS, patients achieved substantial improvement in pain and function following surgery. Those with the worst pre-operative scores achieved the greatest improvement (28.8 point change in those with pre-operative OHS < 5), however patients with the best pre-operative scores still achieved substantial improvement (change of 10.6 points in those with pre-operative OHS > 30). Interestingly, between 1 and 5-years follow up a steady state is reached where there is no further improvement nor decline, with a non-significant trend between OHS and time (p=0.88).

A number of variables were identified as important predictors of attained post-operative OHS (**Table 2**). The strongest determinant of outcomes was the baseline OHS. Increasing baseline OHS (better pre-operative pain/function) was associated with increasing follow-up OHS (better post-operative pain/function). The effect of age was non-linear, where compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcomes. Increasing BMI, patients with a greater number of co-existing diseases prior to surgery, and those with worse pre-operative SF36 mental health scores, also had worse outcomes. The surgical predictor we identified was femoral component ~~size (~~offset~~)~~, where patients with larger ~~components~~ offset size (offset of 44 or more) had significantly better outcomes. We hypothesised that this may be explained by an interaction with gender, where a larger offset is used in men. A significant interaction was observed, where no association was observed in men, whilst in women those with larger ~~components~~ offsets had better outcomes. The effect of surgical approach was significant at the 1-year follow-up, where the anterolateral approach had better outcomes than posterior (difference in 1-year OHS of 2.2 units 95%CI (1.1 to 3.3), however the effect size attenuated over time and became no longer significant between 3 to 5 years follow up.

Figure 3 describes the change in OHS over time stratified according to each of the predictor variables we identified in this study. The graphs highlight that whilst there are small statistically significant differences in *attained* post-operative OHS relating to patient (age, BMI, co-morbidities, mental health) and surgical (femoral component offset) characteristics at the time of surgery, this is greatly outweighed by the substantial *change* in OHS achieved by these patients, regardless of whether they are old or young, obese or not obese. These patient groups still receive great benefit from surgery.

Assessing the discriminatory ability of the final model, including the baseline OHS alone explained 10.3% of the variability in outcome. The final predictive model including the patient and surgical variables explained 16.6%. This suggests that although we have identified significant patient and surgical predictors of outcome, they have smaller effects, and explain little of the variability in attained post-operative OHS above that of the baseline score.

DISCUSSION

Main findings

Within a large prospective cohort of patients receiving primary THR in the UK, we identified a number of predictors of differences in *attained* post-operative pain and function. Determinants included: *pre-operative pain and function* – those with worse pre-operative pain/function had worse post-operative pain/function; *femoral component ~~size (~~offset~~)~~* – women with an offset of 44 or more had better outcomes; *age* - compared to those aged 50 to 60, younger patients (age <50) and older patients (age >60) had worse outcome; increasing *BMI,* a greater number of *co-existing diseases* and worse *SF36 mental health* at the time of surgery, was related to worse post-operative pain and function. The strongest determinant of outcome was the baseline score with the patient and surgical variables contributing small but statistically significant effects.

Assessing the relationship of *change* in symptoms of pain and function between pre and post-operative assessments for predictor variables in the final model revealed that patients achieved large symptomatic improvement (change), regardless of differences in pre-operative patient and surgical factors. The change in symptoms greatly outweighs any differences in attained post-operative score – patients achieved great benefit from surgery regardless of factors such as their age and BMI at the time of surgery. Exploring temporal trends in symptoms of pain and function over time demonstrated that there was little further improvement nor decline in the short to mid-term (between 1 and 5-years follow-up) where a steady symptomatic state was reached.

Strengths and limitations

The strength of this study lies in its large sample size and repeated measures of a reliable, valid and responsive instrument for assessing outcomes of THR[28,29], with data collected prospectively over 5-years with a good rate of follow up. The use of multiple imputation and bootstrapping as an internal validation technique is a strength of this study, ensuring the predictors we identified are those most likely to be replicated in external validation studies, and not chance significant findings that are anomalies of our dataset. Within this study the aim was to identify predictors of differences in attained post-operative pain and function – it should be noted that predictors we identified of attained health state may not necessarily be the same as predictors of improvement (change) in symptoms. It should be noted that it remains unclear which measure of the two outcomes of change in status or attained status is most appropriate for judging the value of surgery[35] – THR could either be viewed as intended to preserve the highest levels of pain function, or alternatively to maximise the potential for symptomatic improvement. Limitations are that other potential predictive variables were not available in this study, predominantly radiographic factors such as x-ray grade and pattern of OA, and other factors including patient expectations of surgery and the type and extent of joint damage. Response bias may play a role, as responders were younger and had better pre-operative SF36 mental health scores, hence the true effects of these predictors may be underestimated in this study.

Within this study histograms of the absolute difference (change) in scores between pre and post-operative assessments highlight that almost all patients get better with only a small minority getting worse or receiving no improvement (2.3% by 1-year and 1.2% by 5-years). This is set in the context that data in the study come from seven high volume centres with skilled surgeons across England and Scotland and raise the possibility that outcomes in this study may be better than expected.

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Comparing this finding with data from other cohorts, in the EUROHIP study consisting of THR patients across 20 European orthopaedic centres, based on the change in WOMAC score at 12-months 58 (6.9%) of 845 patients had no change or worsening of symptoms – although this varied by region and in patients from the UK, only 4 (3.5%) of 111 patients had no change or got worse. Data from the Elective Orthopaedic Centre (EOC) database that include primary hip replacements performed in four acute NHS trusts in the UK in South West London, 88 (5.1%) of 1711 patients symptoms got worse by 6-months based on the OHS, and in data from St. Helier district general hospital, Carshalton, UK, 14 (2.3%) of 619 patients had no change or worsening in OHS at 12-months. Hence, whilst it is plausible that outcomes in the EPOS study may be better than expected, the findings are consistent with other cohorts.

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Bold

Formatted: Font: Not Bold, Not Italic

Formatted: Font: Not Italic

What is already known

In our study an important new finding was that a larger femoral component ~~size (~~offset~~)~~ is associated with better outcomes of THR. We hypothesised that the effect of ~~femoral component~~offset size may be explained by an interaction with gender, as men have larger offsets. A significant interaction was observed, where no association was observed in men, whilst in women those with larger offsets had better outcomes. The choice of offset can affect hip stability as well as abduction strength, potentially resulting in abnormal (trendelenburg) gait. Component offset may be preoperatively templated, although common offsets are often assumed (i.e. 37.5mm for females and 44.0mm males). There is greater potential to decrease offset in females, partly because of the above assumption, and it is sometimes difficult to use the larger offset components because a smaller femoral canal diameter precludes their use. The choice of offset for the femoral ~~stem~~ component has not changed since this study was conducted and these finding are generalizable to clinical practice. We are not aware of any data in the literature describing the relationship of intra-operative surgical factors on patient reported outcomes. Within the literature, data exists on the relationship between head size and failure of a THR, whereby larger head size (40mm versus 28mm) with a ceramic-on-ceramic bearing surface was associated with lower 5-year revision rates[36]. This is though to be related to larger diameter heads increasing fluid-film lubrication, in turn reducing wear, and decreasing dislocation rates[36]. However, femoral ~~stem~~ component offset and head size are independent factors that are unrelated to one another. Within this study we found no association between head size and patient reported outcomes – this is unsurprising since the head sizes used were 22, 26 and 28mm and common to orthopaedic practice at the time this study took place.

Thinner liners are now manufactured allowing larger head sizes to be used in acetabular components of the same size.

Consistent with others in the literature we observed that worse pre-operative mental health was a predictor of poor outcome[16] as were greater numbers of pre-operative co-morbidities[8,15]. We found that older age and higher BMI was associated with worse patient reported outcomes. Within the literature whilst some authors conclude that older age and increasing BMI area associated with worse outcomes, others have found no evidence of an association[7-9,12-19,31,37]. This is in line with the conclusions of large literature reviews stating that such factors are not strong predictors of outcome[11]. The findings are important to decision making as physicians often advise patients they are too old or obese to receive THR[11,38]. We can conclude that in relation to patient reported outcomes of THR, even if some groups fare less well after THR, it does not mean these patients don't get benefit from surgery[11]. Expectations of the patients may also play a role, where for example, what a young person wants to achieve in functional rehabilitation is different to an older person, for whom a lower functional score may be perfectly acceptable. It also is well known within the literature that patients with better pre-operative pain and function achieve better *attained* post-operative pain and function, and that patients with worse pre-operative pain and function get the greatest *change* (symptomatic improvement) between baseline and follow up[12,30,38-40].

Whilst the effects of pre-operative patient characteristics including age, BMI, co-existing diseases and mental health may already be known within the existing literature, what is novel about this study is contrasting the effects of these factors on attained levels of post-operative scores with a graphical representation of change (improvement in scores). Our findings highlight that small but significant differences in attained scores are greatly outweighed by the fact these patient groups get great benefit from surgery (in terms of substantial change in symptoms of pain and function), regardless of differences in pre-operative patient characteristics.

Within this study, although we examined a wide range of patient and intra-operative surgical factors we were only able to explain around 17% of the variability in patient reported outcomes of THR. Data were not available to us on other factors known within the literature to be predictive of outcomes of THR such as radiographic information and the presence of musculoskeletal disease in other joints. It has previously been shown that patients with pain in other joints[39] [40] have worse outcomes. Patients with worse pre-operative x-rays are more likely to improve[13,14]. Worse outcome have been observed in patients who live alone[15,19,] those with less social support[31] and patients with

Formatted: Font: Font color: Auto, Check spelling and grammar, Not Superscript/ Subscript lower educational attainment[8,12,14,18]. Patients with greater pre-operative expectations of surgery have been observed to have better outcomes[14,20]. Whilst inclusion of such factors would help to improve the predictive ability on the model, as individual factors of interest, it is likely we would see the same pattern observed whereby small differences are seen in attained score but patients achieve benefit.

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: 11 pt

Formatted: Font: Not Italic, Font color: Black

What this study adds

Within this study we have identified a number of patient and surgical predictors of attained post-operative pain and function following THR surgery. These predictors remain related to outcome over the short to mid term. An important new finding was that larger femoral component size (offset) was associated with better post-operative pain and function in women. This finding implies that greater consideration should be given to measuring and deciding upon the choice of offset in women as there is potential to undersize. Whilst age, BMI, co-existing diseases, mental health and femoral component offsetsize were associated with small but significant differences in attained pain and function, analyses of change demonstrate that these patient still achieve substantial symptomatic benefit from surgery regardless of differences in these pre-operative factors. The findings will be important to inform patient and clinician decision regarding the likely outcomes of surgery for these patient groups. Although we have assessed a wide range of patient characteristics and intra-operative surgical factors there is still uncertainty as to the cause of variation in outcomes of hip replacement. There is a need to focus on issues that remain unclear such as the effect of soft tissue and the severity and pattern of OA. Further research is needed using more detailed measures of existing predictive variables, and identification of other factors beyond those observed in this study that explain a greater proportion of the variability in outcome to improve our ability to identify patients at risk of poor outcomes from THR surgery.

Funding statement

This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10064). The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This study has been supported by Stryker Howmedica Osteonics (Newbury, UK). Support was also received from the Oxford NIHR Musculoskeletal Biomedical Research Unit, Nuffield Orthopaedic Centre, University of Oxford and the UK Medical Research Council, MRC Lifecourse Epidemiology Unit, University of Southampton.

Competing interest declaration

"All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: AJ, RNB and GT have no conflicts of interest; MKJ, NKA, and CC have received honorariums, held advisory board positions (which involved receipt of fees), and received consortium research grants, respectively, from: Novartis and Alliance for Better Health and Lilly; Merck, Merck Sharp and Dohme, Roche, Novartis, Smith and Nephew, Q-MED, Nicox, Servier, GlaxoSmithKline, Schering-Plough, Pfizer, and Rottapharm; and Alliance for Better Bone Health, Amgen, Novartis, Merck Sharp and Dohme, Servier, Eli Lilly, and GlaxoSmithKline; DB has held an independent consultancy with ICNet and Stryker, and has been a grant holder on a Genzyme-funded study; DM has received royalties from Biomet and Wright Medical Technology, Inc, and receives research support from DePuy, A Johnson & Johnson Company, Stryker, Zimmer and Wright Medical Technology, Inc; they have no other relationships or activities that could appear to have influenced the submitted work.

Contributors

AJ NKA RB GT DB MKJ CC DM: (1) substantial contributions to conception and design, analysis and interpretation of data; (2) drafting the article or revising it critically for important intellectual content; and (3) final approval of the version to be published.

Ethics

Ethical approval was obtained from the Salford and Trafford Research Ethics Committee (Project No: - 98105 – MREC 98/8/20 UK Multicentre Exeter Primary Outcome Study). Informed written consent was obtained from all participants.

REFERENCES

1. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (1). *N. Engl. J. Med.* 1990;323(11):725-731.

2. Harris WH, Sledge CB. Total Hip and Total Knee Replacement (2). *N. Engl. J. Med.* 1990;323(12):801-807.

3. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-Related Quality of Life in Total Hip and Total Knee Arthroplasty. A Qualitative and Systematic Review of the Literature. *J. Bone Joint Surg. Am.* 2004;86(5):963-974.

4. Cooper C, Arden NK. Excess mortality in osteoarthritis. *BMJ.* 2011;342.

5. Professor the Lord Darzi of Denham Kbe HonFREng FmedSci. *High Quality Care for All. NHS Next Stage Review Final Report - Summary*: Department of Health;2008.

6. Williams O, Fitzpatrick R, Hajat S, et al. Mortality, morbidity, and 1-year outcomes of primary elective total hip arthroplasty. *J. Arthroplasty.* 2002;17(2):165-171.

7. Nilsdotter AK, Petersson IF, Roos EM, Lohmander LS. Predictors of patient relevant outcome after total hip replacement for osteoarthritis: a prospective study. *Ann. Rheum. Dis.* 2003;62(10):923-930.

8. MacWilliam CH, Yood MU, Verner JJ, McCarthy BD, Ward RE. Patient-related risk factors that predict poor outcome after total hip replacement. *Health Serv. Res.* 1996;31(5):623-638.

9. Judge A, Cooper C, Williams S, Dreinhoefer K, Dieppe P. Patient-Reported Outcomes One Year After Primary Hip Replacement in a European Collaborative Cohort. *Arthritis Care Res.* 2010;62(4):480-488.

10. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. *Curr. Opin. Rheumatol.* 2006;18(5):526-530.

11. Santaguida PL, Hawker GA, Hudak PL, et al. Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can. J. Surg.* 2008;51(6):428-436.

12. Fortin PR, Clarke AE, Joseph L, et al. Outcomes of total hip and knee replacement: preoperative functional status predicts outcomes at six months after surgery. *Arthritis Rheum.* 1999;42(8):1722-1728.

13. Cushnaghan J, Coggon D, Reading I, et al. Long-term outcome following total hip arthroplasty: a controlled longitudinal study. *Arthritis Rheum.* 2007;57(8):1375-1380.

14. Judge A, Cooper C, Arden NK, et al. Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres. *Osteoarthritis Cartilage.* 2011;19(6):659-667.

15. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. The Effect of Age on Pain, Function, and Quality of Life After Total Hip and Knee Arthroplasty. *Arch. Intern. Med.* 2001;161(3):454-460.

16. Quintana JM, Escobar A, Aguirre U, Lafuente I, Arenaza JC. Predictors of health-related quality-of-life change after total hip arthroplasty. *Clin. Orthop.* 2009;467(11):2886-2894.

17. Braeken AM, Lochhaas-Gerlach JA, Gollish JD, Myles JD, Mackenzie TA. Determinants of 6-12 month postoperative functional status and pain after elective total hip replacement. *Int. J. Qual. Health Care.* 1997;9(6):413-418.

18. Rissanen P, Aro S, Sintonen H, Slatis P, Paavolainen P. Quality of life and functional ability in hip and knee replacements: a prospective study. *Qual. Life Res.* 1996;5(1):56-64.

19. Hopman WM, Mantle M, Towheed TE, Mackenzie TA. Determinants of health-related quality of life following elective total hip replacement. *Am. J. Med. Qual.* 1999;14(3):110-116.

20. Gandhi R, Davey JR, Mahomed N. Patient expectations predict greater pain relief with joint arthroplasty. *J. Arthroplasty.* 2009;24(5):716-721.

21. Andrew JG, Palan J, Kurup HV, Gibson P, Murray DW, Beard DJ. Obesity in total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2008;90-B(4):424-429.

22. Palan J, Beard D, Murray D, Andrew J, Nolan J. Which Approach for Total Hip Arthroplasty: Anterolateral or Posterior? *Clin. Orthop.* 2009;467(2):473-477.

23. Palan J, Gulati A, Andrew JG, Murray DW, Beard DJ, the Esg. The trainer, the trainee and the surgeons' assistant: clinical outcomes following total hip replacement. *Journal of Bone and Joint Surgery - British Volume.* 2009;91-B(7):928-934.

24. Hossain M, Parfitt D, Beard D, et al. Does pre-operative psychological distress affect patient satisfaction after primary total hip arthroplasty? *BMC Musculoskeletal Disorders.* 2011;12(1):122.

25. Charnley J. The Long-Term Results of Low-Friction Arthroplasty of the Hip Performed as a Primary Intervention. *J. Bone Joint Surg. Br.* February 1, 1972 1972;54-B(1):61-76.

26. D'Aubigne RM, Postel M. Functional Results of Hip Arthroplasty With Acrylic Prosthesis. *J. Bone Joint Surg. Am.* June 1, 1954 1954;36(3):451-475.

27. Ware JE. *How to score the revised MOS Short Form Health Scales*: Boston: The Health Institute, New England Medical Centre Hospital;1988.

28. Murray DW, Fitzpatrick R, Rogers K, et al. The use of the Oxford hip and knee scores. *J. Bone Joint Surg. Br.* 2007;89-B(8):1010-1014.

29. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J. Bone Joint Surg. Br.* 1996;78-B(2):185-190.

30. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).

31. Fitzgerald JD, Orav EJ, Lee TH, et al. Patient quality of life during the 12 months following joint replacement surgery. *Arthritis Rheum.* 2004;51(1):100-109.

32. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.

33. Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology.* 2007;7(1):33.

34. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

35. Losina E, Katz JN. Total knee replacement: pursuit of the paramount result. *Rheumatology (Oxford).* 2012.

36. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *The Lancet.*379(9822):1199-1204.

37. Harse JD, Holman CD. Charlson's Index was a poor predictor of quality of life outcomes in a study of patients following joint replacement surgery. *J. Clin. Epidemiol.* 2005;58(11):1142-1149.

38. Hawker G, Wright J, Coyte P, et al. Health-related quality of life after knee replacement. *J. Bone Joint Surg. Am.* 1998;80(2):163-173.

39. Judge A, Javaid MK, Arden NK, et al. A clinical tool to identify patients who are most likely to receive long term improvement in physical function after total hip arthroplasty. *Arthritis Care Res.* 2012:n/a-n/a.

Table 1. Descriptive statistics and comparison of those who did, and did not, complete the 5-year follow up questionnaire

Variable	Missing	Baseline (n = 1431)	Non Responders at Year 5	Responders at Year 5	P-value
Oxford Hip Score					
Pre-operative	70 (4.9%)	16.4 (7.8)	16.1 (8.2)	16.5 (7.6)	0.35
1-year post-op	274 (19.1%)	43 (36, 46)	-	-	
2-year post-op	316 (22.1%)	43 (36, 47)	-	-	
3-year post-op	368 (25.7%)	43 (36, 47)	-	-	
4-year post-op	430 (30.0%)	44 (36, 47)	-	-	
5-year post-op	396 (27.7%)	43 (37, 47)	-	-	
Patient characteristics					
Age	9 (1%)	70.0 (63.9, 76.1)	73.6 (66.2, 79.3)	68.8 (62.7, 74.5)	< 0.001
BMI	95 (7%)	27.4 (4.9)	27.1 (4.9)	27.6 (4.8)	0.077
Gender	7 (0%)				0.46
Male		537 (38%)	171 (39%)	366 (37%)	
Female		887 (62%)	266 (61%)	621 (63%)	
Occupation	0 (0%)				0.005
Heavy manual		41 (3%)	8 (2%)	33 (3%)	
Light manual		89 (6%)	20 (5%)	69 (7%)	
Office / professional		107 (7%)	21 (5%)	86 (9%)	
Housewife		187 (13%)	67 (15%)	120 (12%)	
Unemployed / retired		1007 (70%)	325 (74%)	682 (69%)	
No. of Co-existing Diseases	0 (0%)				0.94
0		431 (30%)	136 (31%)	295 (30%)	
1		498 (35%)	147 (33%)	351 (35%)	
2		315 (22%)	99 (22%)	216 (22%)	
3		140 (10%)	43 (10%)	97 (10%)	
4		47 (3%)	16 (4%)	31 (3%)	
Concomitant therapy used	8 (1%)				0.84
No		104 (7%)	31 (7%)	73 (7%)	
Yes		1319 (93%)	406 (93%)	913 (93%)	
SF36 Mental Health Score	515 (36%)	74 (60, 88)	72 (52, 88)	76 (60, 88)	0.046

Cells represent either: number (percentage), mean (standard deviation), median (interquartile range)
† T-tests are used for continuous variables and chi-squared tests for categorical variables.
Where continuous variables were not normally distributed, a non-parametric t-test (Kruskal-Wallis) was used.
Fishers exact test is used where expected counts are less than 5.

Table 2. Repeated measures Analysis of Covariance (ANCOVA) models to identify predictors of the average OHS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Multivariable Δ Coef (95% CI)	P-value
Patient variables				
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	0.01 (-0.11, 0.13)	0.88
Age				
<50	-1.44 (-3.92, 1.03)	0.25		
50-60	-0.96 (-2.27, 0.35)	0.15	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	-1.54 (-2.45, -0.64)	0.001
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	-0.90 (-1.27, -0.54)	< 0.001
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	0.76 (0.46, 1.07)	< 0.001
Surgical variables				
~~Stem~~ Femoral component offset size (mm offset)	0.15 (0.04, 0.26)	0.01	0.17 (0.06, 0.28)	0.002
R^2				17.4%
Optimism				0.8%
Bias-Corrected R^2				16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Variables included in the final regression model are those that are retained in at least 70% of the 200 bootstrap backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OHS only

Figure 1. Distribution of OHS at baseline, follow-up and absolute difference in scores

Figure 2. Change in Oxford Hip Score over time, stratified by baseline score

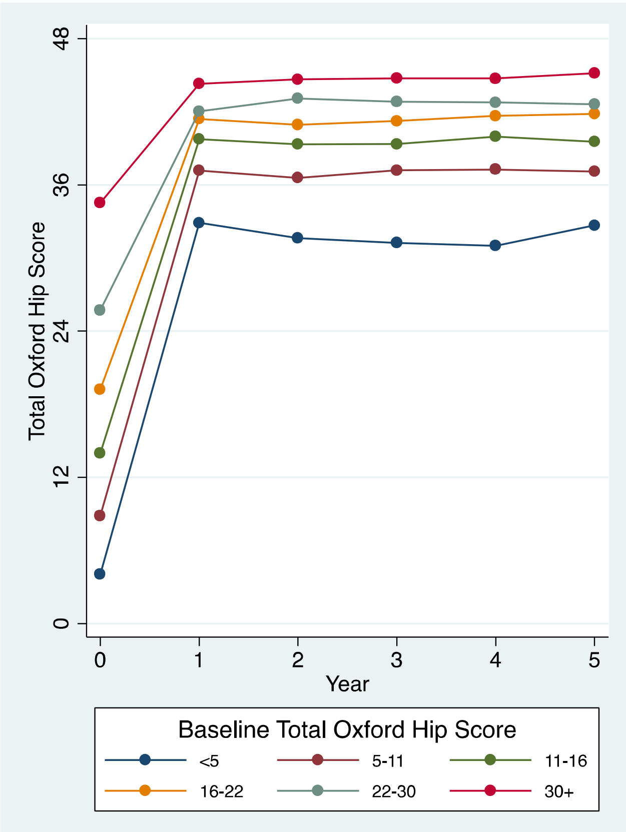
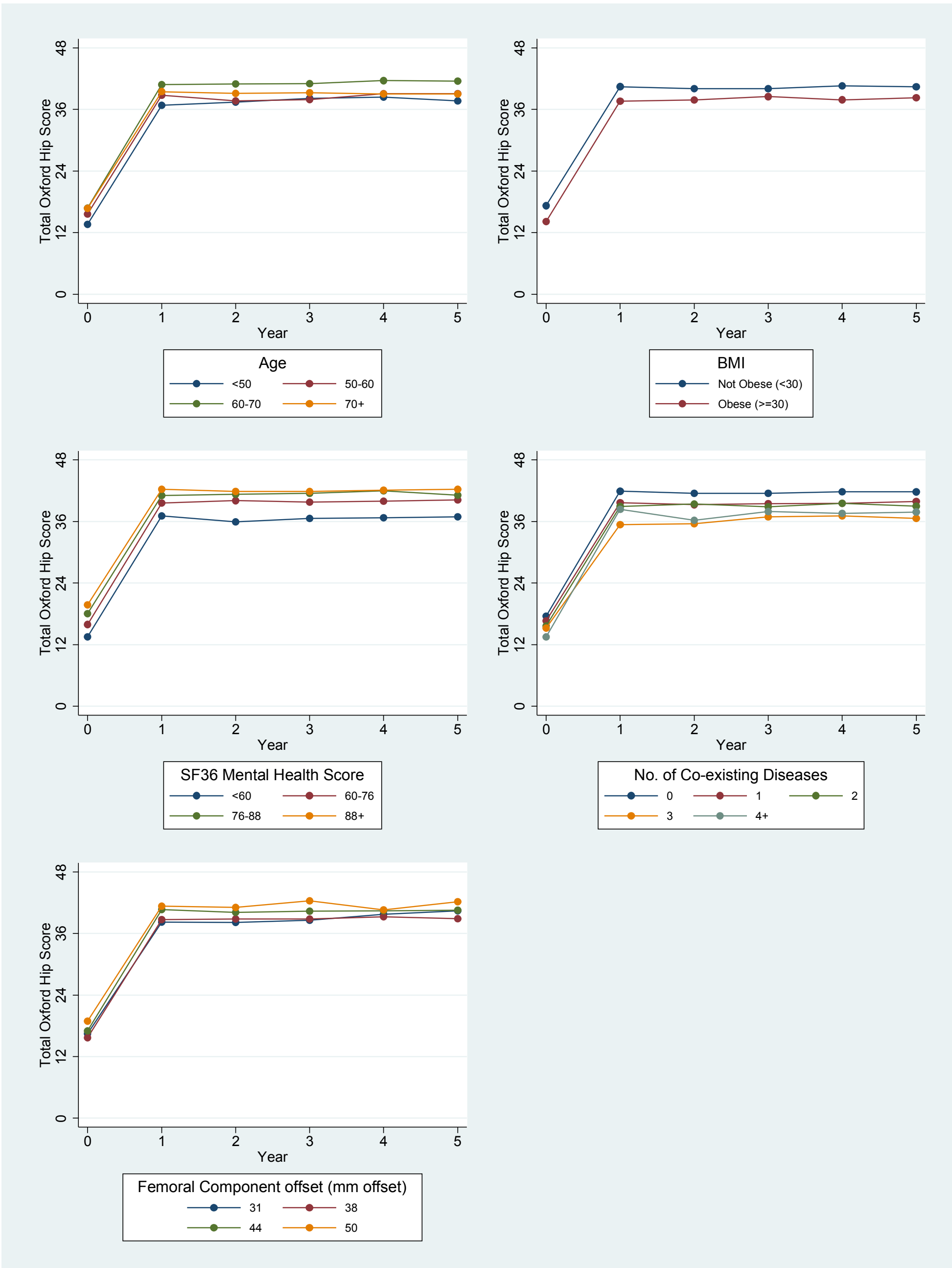


Figure 3. Change in Oxford Hip Score over time, stratified by predictive variables



Formatted: Font: 9 pt

SUPPLEMENTARY MATERIAL

MULTIPLE IMPUTATION METHODS

The results of complete case analyses can be biased[1]. The cumulative effect of missing data in several variables often leads to exclusion of a substantial proportion of the original sample, causing a loss of precision and power. Multiple imputation methods can be used to handle datasets with missing values. The risk of bias depends on the reasons why the data are missing. Missing data are seldom completely random. They are usually related, directly or indirectly, to other subject or disease characteristics, including the outcome under study[2]. If it is plausible the data are missing at random, but not completely at random, analyses on complete cases may be biased[3]. This bias can be overcome by using multiple imputation, which allows for the uncertainty about missing data by creating several plausible imputed datasets and appropriately combining their results. We have done this using the ICE procedure in Stata[4-6]. The first stage is to create multiple copies of the dataset with missing values replaced by imputed ones (we have created 10 copies). Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. In the second stage regression models are fitted to each of the imputed datasets and averaged together to give overall estimated associations. Standard errors are calculated using Rubins Rules. We have included all predictor variables in the multiple imputation process, together with the outcome variable as this carries information about missing values of the predictors.

INTERNAL VALIDATION

Model building in prognostic studies is usually performed using automatic stepwise variable selection procedures. However, stepwise methods have a number of disadvantages[3,7,8], where their power to select true variables is limited and estimates of predictive validity and fit may be overly optimistic. It has been suggested that of the final significant variables included in a prognostic model using backwards selection only half may be true risk factors that would be replicated by other studies[8]. To overcome these limitations, it has been suggested that bootstrapping combined with automatic backward regression can be used to provide information on model stability[7,8].

For internal validation of the regression models we therefore use a combination of multiple imputation and bootstrapping. Firstly, missing data is imputed using the ICE procedure in Stata and 10 imputed datasets created. Missing values are sampled from their predictive distribution based on the observed data. The imputation procedure accounts for uncertainty in predicting missing values by injecting appropriate variability into the multiple imputed values. Using the 'micombine' procedure in Stata the full regression models including all predictor variables are fitted to each of the imputed datasets and averaged together to give overall estimated associations with standard errors calculated using Rubins Rules[6]. Second, 200 bootstrap samples are then randomly drawn with replacement (e.g. when a patient is randomly selected their data is taken from each of the 10 imputed datasets). An automatic backward selection procedure is then applied to each of the 200 bootstrap samples of 10 imputed datasets using a Wald test with a stopping rule of α = 0.157. This conservation p-value is comparable with the more complex Akaike Information Criterion (AIC)[7]. Variables retained in the final regression model are those consistently selected across the re-samples at least 70% of the time.

REFERENCES

1. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338(jun29_1):b2393.

2. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

3. Harrell FE, Jr. *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis*: New York: Springer; 2001.

4. Royston P. Multiple imputation of missing values. *The Stata Journal.* 2004;4(3).

5. Royston P. Multiple imputation of missing values: update. *The Stata Journal.* 2005;5(2).

6. Royston P. Multiple imputation of missing values: Update of ice. *The Stata Journal.* 2005;5(4).

7. Vergouw D, Heymans M, Peat G, et al. The search for stable prognostic models in multiple imputed data sets. *BMC Medical Research Methodology.* 2010;10(1):81.

8. Heymans M, van Buuren S, Knol D, van Mechelen W, de Vet H. Variable selection under multiple imputation using the bootstrap in a prognostic study. *BMC Medical Research Methodology.* 2007;7(1):33.

Repeated measures Analysis of covariance (ANCOVA) models to identify predictors of the average OHS between 1 and 5-years follow up

Variable	Univariable Δ Coef (95% CI)	P-value	Percentage retained in model	Multivariable (70% cut-off) Δ Coef (95% CI)	P-value
Patient variables					
Baseline Total Oxford Hip Score (10 units)	3.68 (3.16, 4.20)	<0.001	100.0%	2.68 (2.16, 3.21)	< 0.001
Year	0.02 (-0.10, 0.13)	0.77	100.0%	0.01 (-0.11, 0.13)	0.88
Age			-		
<50	-1.44 (-3.92, 1.03)	0.25	-		
50-60	-0.96 (-2.27, 0.35)	0.15	59.5%	-1.87 (-3.22, -0.53)	0.006
60-70	0.00 (0.00, 0.00)	-	92.5%	0.00 (0.00, 0.00)	-
70-80	-0.37 (-1.20, 0.46)	0.38	98.0%	-1.49 (-2.37, -0.61)	0.001
80+	-2.29 (-3.69, -0.88)	0.001	100.0%	-3.81 (-5.29, -2.33)	< 0.001
BMI (10 units)	-1.14 (-2.05, -0.22)	0.02	98.0%	-1.54 (-2.45, -0.64)	0.001
Gender			-		
Male	0.00 (0.00, 0.00)	-	-		
Female	-1.30 (-2.13, -0.48)	0.002	38.0%		
Hip For Surgery			-		
Left	0.00 (0.00, 0.00)	-	-		
Right	-0.16 (-0.99, 0.67)	0.71	15.0%		
Hip Indicated			-		
Unilateral	0.00 (0.00, 0.00)	-	-		
Bilateral	-0.48 (-3.01, 2.06)	0.71	16.5%		
Occupation			-		
Heavy manual	0.00 (0.00, 0.00)	-	-		
Light manual	-0.85 (-2.50, 0.80)	0.31	38.5%		
Office / professional	2.26 (0.96, 3.57)	0.001	58.0%		
Housewife	-0.92 (-2.18, 0.33)	0.15	54.5%		
Unemployed / Retired	-0.15 (-1.04, 0.73)	0.74	43.5%		
No. of Co-existing Diseases	-1.06 (-1.43, -0.69)	<0.001	98.0%	-0.90 (-1.27, -0.54)	< 0.001
Concomitant therapy used			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.73 (-2.02, 0.56)	0.27	22.0%		
Centre number			-		
1	0.00 (0.00, 0.00)	-	-		
2	-0.93 (-2.23, 0.38)	0.16	38.0%		
3	0.31 (-0.69, 1.32)	0.54	34.5%		

4	1.48 (0.28, 2.68)	0.02	44.5%		
5	-1.76 (-3.27, -0.25)	0.02	69.0%		
6	-1.18 (-2.44, 0.07)	0.06	62.0%		
7	2.22 (1.33, 3.12)	<0.001	45.0%		
Fixed flexion	0.03 (-0.01, 0.06)	0.12	47.0%		
SF36 Mental Health Score (10 units)	0.80 (0.50, 1.11)	<0.001	100.0%	0.76 (0.46, 1.07)	< 0.001
Surgical variables					
Grade of Operator			-		
Consultant, locum consultant, assoc. specialist/staff	0.00 (0.00, 0.00)	-	-		
Fellow, senior registrar, registrar, locum registrar	-1.22 (-2.13, -0.32)	0.008	19.5%		
Surgical Approach			-		
Anterolateral	0.00 (0.00, 0.00)	-	-		
Posterior	1.03 (0.07, 1.98)	0.04	17.5%		
Patient's position			-		
Supine	0.00 (0.00, 0.00)	-	-		
Lateral	-0.99 (-2.00, 0.03)	0.06	25.5%		
Lavage System (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.70 (-2.60, 1.21)	0.47	64.5%		
Cement Pressurisation (Acetabular)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-0.37 (-1.43, 0.68)	0.48	34.5%		
Type of cement (Socket)			-		
No Cement	0.00 (0.00, 0.00)	-	-		
simplex	0.09 (-0.73, 0.91)	0.84	43.0%		
cmw1	0.63 (-0.33, 1.58)	0.2	28.5%		
palacos r	-1.21 (-2.34, -0.09)	0.03	33.0%		
Cement pressurisations (Femur)			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	0.58 (-3.85, 5.01)	0.8	20.0%		
Type of cement (Femur)			-		
simplex	0.00 (0.00, 0.00)	-	-		
cmw1	1.49 (0.20, 2.79)	0.02	48.0%		

cmw3	0.23 (-0.88, 1.34)	0.68	27.5%		
palacos r	-1.26 (-2.49, -0.03)	0.04	22.0%		
palacos lv	-1.44 (-4.89, 2.01)	0.41	37.0%		
Stem size (mm offset)	0.15 (0.04, 0.26)	0.01	84.0%	0.17 (0.06, 0.28)	0.002
Femoral Head			-		
Stainless Steel	0.00 (0.00, 0.00)	-	-		
Ceramic - Zirconia/ Alumina	0.76 (-0.66, 2.17)	0.29	21.0%		
Head size			-		
22	0.00 (0.00, 0.00)	-	-		
26	1.11 (0.28, 1.93)	0.009	32.0%		
28	-0.15 (-0.97, 0.68)	0.73	29.0%		
Type of Polythene			-		
uhmwpe	0.00 (0.00, 0.00)	-	-		
duration	1.64 (0.83, 2.46)	<0.001	52.0%		
Hip Dislocation			-		
No	0.00 (0.00, 0.00)	-	-		
Yes	-4.27 (-8.05, -0.50)	0.03	74.5%	-3.77 (-7.47, -0.07)	0.05
Acetabular cup inclination (10 degrees)	-0.11 (-0.81, 0.58)	0.75	6.0%		
Acetabular cup version (10 degrees)	0.82 (0.17, 1.47)	0.01	32.5%		
Duration of Operation (Log)	-1.90 (-3.10, -0.70)	0.002	57.5%		
R2		-			17.4%
Optimism		-			0.8%
Bias-Corrected R2		-			16.6%

Δ: Represents the average follow up OHS between 1, 2, 3, 4 and 5-years follow up.

Percentage: the proportion of times the variable was retained in the backward selection regression models using a P-value of 0.157 (inclusion frequency)

70% cut-off: Variables included in the final regression model are those that are retained in at least 70% of the backward selection regression models

Univariable – Each predictor in the model is adjusted for Baseline OKS only

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract In the title and in the abstract as "population-based cohort study"
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found This is provided in the abstract
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported A brief background to the study is given in the introduction section
Objectives	3	State specific objectives, including any prespecified hypotheses The overall aim of the study is stated at the end of the introduction "the aim of this study was to identify patient characteristics and intra-operative surgical factors associated with differences in: a) attained post-operative levels of pain and function, b) change (temporal trends) in symptoms of pain and function over time between pre and post-operative assessments."
Methods		
Study design	4	Present key elements of study design early in the paper This is given at the start of the methods section
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection This is described in the methods
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up The eligibility criteria are described as: "Patients were included if they were undergoing primary hip replacement with an Exeter cemented femoral stem and were willing and able to give consent to participate in the study. All eligible patients were invited to participate in the study. Patient recruitment varied between the centres but was between 80%-90% of eligible patients. The geographical area covered by the participating hospitals was wide and included both university teaching and district general hospitals that included urban as well as rural locations and represented both affluent as well as inner city suburbs. The catchment area of the four combined units included over a million people. There were 1,375 patients (1,431 hips) with a primary diagnosis of OA. The unit of analysis was the implant rather than the patient, of whom 56 had bilateral procedures. We examined 1431 THRs performed by consultant and non-consultant surgeons and using anterolateral or posterior approaches." When describing the outcome variable we state the method of follow up "Prior to surgery, patients completed an Oxford Hip Score (OHS) questionnaire with follow-up questionnaires being filled in at 1, 2, 3, 4 and 5-years post-surgery."
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable The outcomes and predictor variables are all clearly described under separate subheadings in the methods section.
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group We have briefly described and referenced methods of assessment such as the Charnley

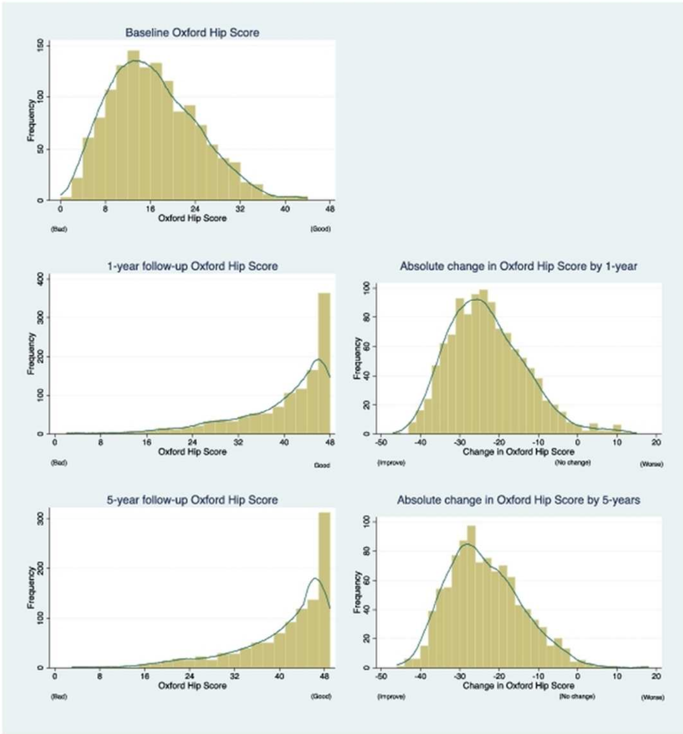
		Modification of D'Aubigne-Postel Grade questionnaire, Short Form 36 (SF-36) and Oxford Hip Score (OHS) questionnaire.
Bias	9	Describe any efforts to address potential sources of bias Efforts to address bias are covered in the statistical methods section, where we describe methods to deal with missing data and internal validation methods.
Study size	10	Explain how the study size was arrived at This was an analysis of an existing cohort of patients previously collected. We have referenced previous papers describing the cohort in greater detail, and included all available patients in the analysis for this paper stating the sample size.
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why For each variable we have been clear to describe whether it is continuous or categorical. All continuous variables have been kept as continuous with the exception of age as the effect on outcome was non-linear. We state the categories of each categorical variable in the results tables.
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding Statistical methods are clearly described in the paper. Generalised Estimating Equations (GEE) was used to account for clustering within the data using an exchangeable correlation matrix. Using GEE, a repeated measures linear regression model was fitted where the outcomes were the OHS at 1, 2, 3, 4 and 5-years follow up, adjusting for the pre-operative OHS as a covariate in the model. A supplementary file is provided giving greater detail of the statistical methodology.
		(*b*) Describe any methods used to examine subgroups and interactions We state in the methods section where interactions have been pre-specified. We state that for significant predictor variables, interaction terms are fitted between the predictor variable and time, where the regression models are stratified by the variable of interest to describe the change in OHS over time, for example, in those who are obese versus not obese. In the results section we describe how that for a significant surgical predictor that was identified, we hypothesised that this may be explained by an interaction with gender, and described the interaction.
		(*c*) Explain how missing data were addressed We explain in the statistical methods how missing data were addressed by using multiple imputation methods
		(*d*) If applicable, explain how loss to follow-up was addressed We explain how we have used repeated measures regression modelling whereby patients are included in the analysis if they have responded to at least 1 of the yearly follow up assessments between 1 and 5-years follow up
		(*e*) Describe any sensitivity analyses No sensitivity analyses have been conducted.
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed At the start of the results section we say how many people received hip replacement surgery, and the number included in the study and analysed
		(b) Give reasons for non-participation at each stage The only reason patients were excluded was if they did not complete at least 1 of the follow up Oxford Hip Score outcome assessments

		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders Descriptive statistics of study participants are provided in table 1
		(b) Indicate number of participants with missing data for each variable of interest The number of patients with missing data for each variable is presented in table 1.
		(c) Summarise follow-up time (eg, average and total amount) We have explained that patients were followed up at yearly intervals up to 5-years following surgery.
Outcome data	15*	Report numbers of outcome events or summary measures over time The outcome is self reported pain and function as measured by the Oxford Hip Score and Table 1 and Figure 2 summarises how this changes over time.
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included Table 2 provides both adjusted and unadjusted estimates with 95% confidence intervals and p-values. We explained which predictor variables were considered for inclusion in the model in the methods section and how bootstrapping with backward selection is used to identify those remaining in the final model.
		(*b*) Report category boundaries when continuous variables were categorized Category boundaries are reported in the tables such as for age where categorised due to non-linearity.
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Only the regression coefficients are reported as a measure of effect size.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses Interactions showing how the Oxford hip score changes over time, by categories of significant predictor variables are displayed in Figures 2 and 3.
Discussion		
Key results	18	Summarise key results with reference to study objectives The main findings are described at the start of the discussion section.
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias We have discussed both the strengths and limitations of the study and the potential role of response bias.
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence We have described how the findings of our study relate to what is already known in the literature and what this study adds.
Generalisability	21	Discuss the generalisability (external validity) of the study results We have considered the issue of generalizability highlighting that this study has used data from seven high volume centres with skilled surgeons.
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based The source of funding has been described.

*Give information separately for exposed and unexposed groups.

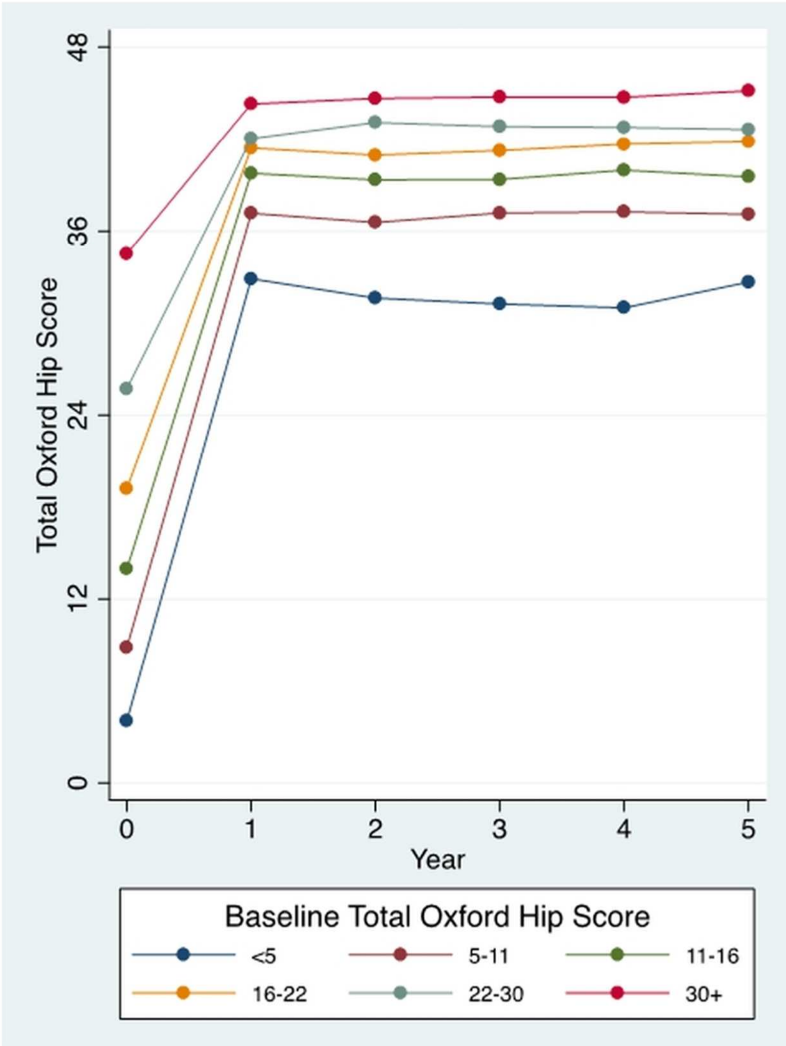
Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Figure 1. Distribution of OHS at baseline, follow-up and absolute difference in scores



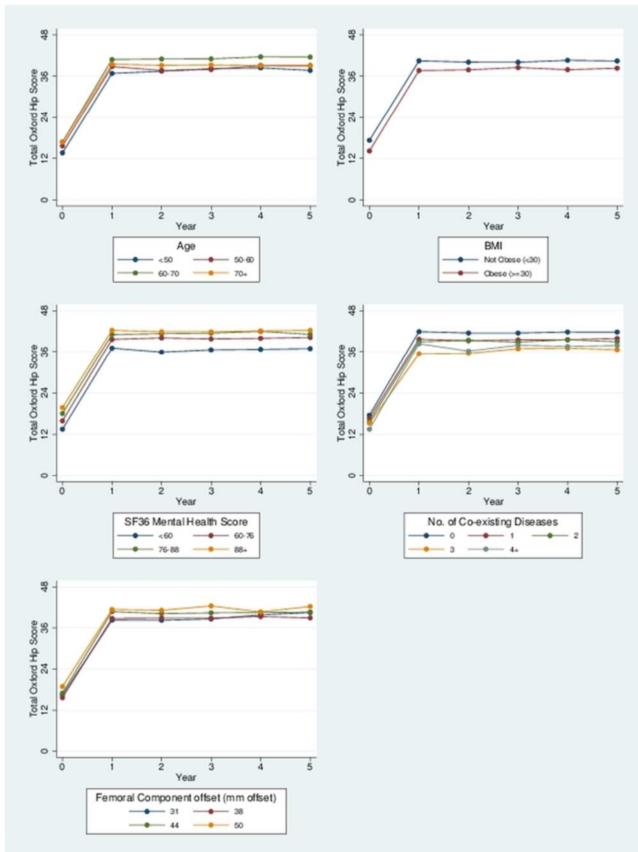
90x116mm (300 x 300 DPI)

Figure 2. Change in Oxford Hip Score over time, stratified by baseline score



90x103mm (300 x 300 DPI)

Figure 3. Change in Oxford Hip Score over time, stratified by predictive variables



1

90x127mm (300 x 300 DPI)