Trajectories of pain and function in the first 5 years after hip and knee joint replacement: an analysis of patient reported outcome data from the National Joint Registry of England, Wales, Northern Ireland and the Isle of Man

Authors & Affiliations:

Author	Affiliations	Email
Jack Dainty PhD	Norwich Medical School, UEA	Jack.dainty@uea.ac.uk
Toby O Smith PhD MCSP	Nuffield Department of Orthopaedics , Rheumatology and Musculoskeletal Sciences, University of Oxford	toby.smith@ndorms.ox.ac.uk
Emma M Clark PhD MRCP	Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, University of Bristol.	emma.clark@bristol.ac.uk
Michael R Whitehouse PhD FRCS (Tr & Orth)	Musculoskeletal Research Unit, Translational Health Sciences, Bristol Medical School, University of Bristol.	michael.whitehouse@bristol.ac.uk
Andrew Price DPhil FRCS (Tr & Orth)	Nuffield Department of Orthopaedics , Rheumatology and Musculoskeletal Sciences, University of Oxford	andrew.price@ndorms.ox.ac.uk
Alex MacGregor PhD MRCP	Norwich Medical School, UEA	a.macgregor@uea.ac.uk

Corresponding Author: Professor Alex MacGregor. Norwich Medical School, University of East Anglia, Norwich, NR4 7JT. Email: a.macgregor@uea.ac.uk

DECLARATIONS AND ACKNOWLEDGEMENTS

Acknowledgement: We thank the patients and staff of all the hospitals in England, Wales and Northern Ireland who have contributed data to the National Joint Registry. We are grateful to the Healthcare Quality Improvement Partnership (HQIP), the NJR Research Committee and staff at the

NJR Centre for facilitating this work. The authors have conformed to the NJR's standard protocol for data access and publication. The views expressed represent those of the authors and do not necessarily reflect those of the National Joint Registry Steering Committee or the Healthcare Quality Improvement Partnership (HQIP) who do not vouch for how the information is presented.
The Healthcare Quality Improvement Partnership ("HQIP") and/or the National Joint Registry ("NJR") take no responsibility for the accuracy, currency, reliability and correctness of any data used or referred to in this report, nor for the accuracy, currency, reliability and correctness of links or references to other information sources and disclaims all warranties in relation to such data, links and references to the maximum extent permitted by legislation.

HQIP and NJR shall have no liability (including but not limited to liability by reason of negligence) for any loss, damage, cost or expense incurred or arising by reason of any person using or relying on the data within this report and whether caused by reason of any error, omission or misrepresentation in the report or otherwise. This report is not to be taken as advice. Third parties using or relying on the data in this report do so at their own risk and will be responsible for making their own assessment and should verify all relevant representations, statements and information with their own professional advisers.

Funding: No funding was receive to support this study. This study was supported by the NIHR Oxford Biomedical Research Centre and the Bristol Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health and Social Care.

Sponsorship: This work was undertaken as part of the National Joint Registry strategic plan.
Conflict of Interest: All authors declare that they have no conflicts of interest in relation to this paper.

Ethical Approvals: Research ethics committee (REC) approval was granted by the NRES London – Central REC (10/H0718/89). The research project was internally approved by the NJR Research Committee.

Data sharing: Access to data is available from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man.

Patient and Public Involvement: Patients were involved in the reporting and dissemination plans of our research. Our PPI members provided valuable input in relation to the interpretation of our findings and how these may be most usefully presented to a wide number of stakeholders including patient and public members.

Contributorship Statement:
Conceptualisation: AM, TS, JD
Data curation: AM
Investigation:AM, JD, TS, EC, MB, AP
Methodology: AM, JD, TS, EC, MB, AP
Formal analysis: JD, AM, TS
Validation: AM, JD, TS, EC, MB, AP
Visualization: AM, JD, TS, EC, MB, AP
Writing – original draft: AM, JD, TS, EC, MB, AP
Writing – review and editing: AM, JD, TS, EC, MB, AP

ABSTRACT

AIMS: To determine the trajectories of patient reported pain and functional disability over five years following total hip replacement (THR) or total knee replacement (TKR).

METHODS: A prospective, longitudinal cohort sub-study within the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR) was undertaken. 20,089 patients who underwent primary THR and 22,489 who underwent primary TKR between 2009 and 2010 were sent Oxford Hip Score (OHS) and Oxford Knee Score (OKS) questionnaires at six months, one, three and five years post-operatively. OHS and OKS were disaggregated into pain and function subscales. A k-means clustering procedure assigned each patient to a longitudinal trajectory group for pain and function. Ordinal regression was used to predict trajectory group membership using baseline OHS and OKS score, age, BMI, IMD, gender, ethnicity, geographical location and ASA grade.

RESULTS: Data described two discrete trajectories for pain and function: 'Level 1' responders (70% of cases) in whom a high-level of improvement is sustained over five years, and a 'Level 2' responders who sustained improvement, but at a lower level. Baseline patient variables were only weak predictors of pain trajectory and modest predictors of function trajectory. Those with worse baseline pain and function tended to show a greater likelihood of following a 'Level 2' trajectory. Six-month PROMs data reliably predict the class of five-year outcome trajectory for both pain and function.

CONCLUSION: Reviewing patient outcomes at six months post-operatively is a reliable indicator of outcome at five years.

Keywords: arthroplasty; surveillance; monitoring; PROMs; national registry

Word Count: Abstract: 245; Manuscript: 2840

INTRODUCTION

An awareness of long-term outcomes following total hip (THR) and total knee replacement (TKR) is essential for planning the provision of care for patients with established hip and knee osteoarthritis. While registries have tended to focus on revision rates as the key measure with which to assess outcomes following joint replacement, revision itself has recognised shortcomings as an indicator of joint failure.[1,2]

Not all joints that fail are revised and not all joints are revised because of failure.[3,4] Patient reported outcome measures (PROMs) of pain and functional limitation provide additional insight[5] and can provide meaningful data on the patient's perception of the effect of their intervention. PROMs for THR and TKR have been collected routinely pre-operatively and at six months after surgery in all patients undergoing hip and knee joint replacement in England since 2009 as part of a national programme and have been widely adopted as quality metrics to evaluate performance variation.[6] The majority of hip and knee replacements will last for more than 25 years[7,8] so the six month follow-up time-point that is currently captured represents a relatively short window of observation and the value of these data is debated.[9] An evaluation at six months might not reflect the maximal improvement following surgery or adequately predict any later decline.

In this analysis, we describe the trajectories of patient reported pain and functional disability reported prospectively over five years among patients in a large cohort of patients following THR and TKR enrolled in the National Joint Register of England, Wales, Northern Ireland and the Isle of Man (NJR). We assess the value of the six-month data in predicting future outcomes.

METHODS

The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines were followed in the reporting of this comparative prospective cohort study.[10]

Participants

The data included in this analysis were collected as part of a stand-alone prospective study commissioned by the NJR to collect long-term PROMs data on patients undergoing THR and TKR. The base sample included patients with a primary THR or TKR who had returned both preoperative and six-month questionnaire sent routinely in England as part of a national surveillance programme. The patients included in this cohort were identified from those who returned six-month questionnaires consecutively over a 12-month period in 2009 and 2010. All had data relating to their surgery entered onto the NJR and had completed preoperative questionnaires. In addition to routine six-month NJR questionnaires, these patients were sent further PROMs questionnaires at one, three and five years post-operatively. The baseline NJR data provided information on age at surgery, gender, body mass index and American Society of Anesthesiologists (ASA) grade. The records were anonymously linked to Hospital Episode Statistics data from 2010, providing information on ethnicity, index of multiple deprivation (IMD) and geographical location.

Patients who had further surgery involving the index joint at any point in the five-year postoperative period were excluded from the baseline sample. Following the guidance of Murray et al,[11]

questionnaires were excluded where data was missing from three or more items within the 12 OHS or OKS questions.

Outcome Measures

The PROMs questionnaires captured responses to variables included in the Oxford Hip and Knee Score (OHS);[12] Oxford Knee Score (OKS).[13] The scores are composed of the sum of 12 individual attributes which a factor analysis has indicated as comprising two underlying components (domains) reflecting degree of pain and functional disability.[14,15] In this analysis, we disaggregated the total OHS and OKS scores into a 'pain' domain (ranging from 0 to 24 for OHS and 0 to 28 for OKS) and the 'function' domain (ranging from 0 to 24 for OHS and 0 to 20 for OKS) with higher values on each domain representing less pain and better function. As an additional exclusion criterion, if there were two missing items within the pain or function domain, that questionnaire was excluded.

Statistical Analysis

To analyse the trajectories of patient outcome over the five-year period of follow-up, we conducted a longitudinal cluster-based analysis. A k-means procedure was used to allocate each patient to a longitudinal trajectory group for 'pain' and 'function' outcomes separately. The procedure was implemented in the 'kml' package in R[16] that is specifically designed for longitudinal data. K-means is an expectation-maximisation hill climbing algorithm that maximises between-cluster variance and minimises within-cluster variance by initially determining the 'centre' of each cluster during the expectation phase and assigning each observation (patient) to its nearest cluster during the maximisation phase. Only cases with data from at least four out of the five possible time points after baseline were included in the analysis.

Cluster sizes from two to six were examined. For each number of clusters, the k-means algorithm was run 20 times with varying starting conditions (randomly allocating a selection of patients to each cluster) to maximise the likelihood that the solution converged to global, rather than local, maxima. The optimal number of clusters was determined by a set of three non-parametric tests,[17-19] which use the ratio of between to within cluster variance and two likelihood-based parametric tests (AIC, BIC). The cluster solution for which the parametric and non-parametric tests converged was chosen to best reflect the data.

Ordinal regression methods were used to predict cluster group membership, initially using baseline patient data and patient OHS and OKS score separately, and in combination with age, gender, ethnicity, geographical location, IMD, preoperative body mass index (BMI), and ASA grade (as a marker for comorbidity). The model was extended to examine the additional impact of including data from OHS and OKS scores reported on the questionnaires at six months and one year.

The data were randomly divided into 'training' and 'test' sets (each set was 50% of the entire data). The test data were used to observe if it was possible to predict the 'unseen' trajectories (identified by the clusters) in the training data. This was repeated for 1000 iterations to create multiple, random splits of the data into training and test sets (Monte Carlo cross-validation) to generate mean predictions and 95% confidence intervals.

The accuracy of each model's binary classification was evaluated by comparing its performance against the Proportional Chance Criterion (PCC).[20] PCC is a measure of the proportion of subjects that would have been classified accurately by chance alone. The probability that the model's accuracy exceeded a threshold of 25% higher than the PCC (>1.25xPCC) was used to evaluate the significance of the model's predictive capacity.

The impact of missing data on the choice of number of clusters was explored by imputation methods, implemented within the kml package. Imputation methods were also used to explore the impact of missing covariate data on the outcomes of the regression analysis.

RESULTS

Patient Characteristics

One, three- and five-year follow-up questionnaires were sent to 20,089 patients who underwent a primary THR and 22,489 who underwent a primary TKR, and who also returned baseline preoperative and six month postoperative PROMs. After accounting for patients who died during the interval, the response rate to follow-up questionnaires sent at one, three- and five-years patients was 83%, 71% and 47% for THR and 84%, 70%, and 53% for TKR. A cohort flow chart is presented in **Supplementary Figures 2a** and **2b**.

For the trajectory analysis, we included patients who had complete data from a minimum of four out of the five time points. For patients undergoing THR, 5,754 patients from the pain domain and 5,723 from the function domain analysis were excluded. For patients who underwent TKR, 6,529 were excluded from the pain domain and 6,644 from the function domain analysis.

A total of 1,360 patients with THR and 2,308 patients with TKR had further surgery on the index joint at any point in the follow-up interval were excluded from the main analysis and analysed separately.

The characteristics of the THR and TKR cohorts are presented in **Supplementary Tables 1** and **2**. The OHS and OKS outcomes (total and subset data) for missing and analysed cohorts is presented in **Supplementary Table 3**.

Outcome Trajectories

The results of the k-mean clustering analysis suggested that the outcomes in both the THR and TKR patients could be described as following one of two trajectories (or clusters) for both the 'pain' and the 'function' domains as defined by the goodness of fit criteria. Both the non-parametric and parametric (AIC, BIC) quality criteria all suggested that two clusters were the best solution. The non-parametric criteria are shown in **Supplementary Figure 1**.

The mean pain and function domain scores for the two trajectories (termed 'Level 1' and 'Level 2') are presented in **Figure 1** and **Figure 2**. Both the Level 1 and Level 2 trajectories show a sustained improvement over baseline throughout the five year follow up interval; however, the improvement is greater at all timepoints in the Level 1 than in Level 2 groups. For both Level 1 and Level 2 groups the maximum improvement in pain and function was achieved by about 12 months for both THR and TKR and remained little changed over subsequent follow up.

Predictors of Pain and Function

The ability of baseline preoperative characteristics and subsequent outcomes to predict trajectory group membership is shown in **Table 1**.

For pain, following THR, baseline preoperative patient characteristics including OHS did not significantly predict trajectory group membership. For TKR, patient characteristics combined with baseline OKS scores were significant predictors of group membership, but only classified patients into Level 2 with an accuracy of 30%. By contrast, for both THR and TKR, data collected at six months

on OHS and OKS combined with patient characteristics was a highly significant predictor of outcome group membership, predicting Level 2 membership with an accuracy of 75% and 77% respectively. The accuracy improved further when data from one year were included in the model.

For function, both the THR and TKR baseline preoperative data combined with OHS and OKS were significant predictors of group membership. However, the accuracy of prediction of Level 2 membership was relatively low (36% for THR and for 55% TKR). Level 2 prediction improved to 79% for THR and 78% for TKR when data were included on function at six months, and there were further improvements when data were included from OHS and OKS at one year.

Sensitivity Analysis

Imputing missing values in the kml procedure led to no difference in the choice of the number of clusters that best described the data. Imputing values for the missing data in the ordinal regression analysis led to no important difference in the estimates of the predictive capacity of the individual models.

DISCUSSION

This analysis of data gathered prospectively from a large national cohort of patients undergoing THR and TKR shows that, on average, the maximal improvement in pain and function has occurred by 12 months postoperatively. The analysis indicates that PROMs for pain and function for the first five years following THR and TKR can be described as following two distinct trajectories: 'Level 1' responders (comprising around 70% of cases) in whom a high level of improvement is sustained over the five year period, and 'Level 2' responders who have sustained improvement, but at a lower level. In this cohort, baseline variables do not predict pain trajectory following THR and are only weak predictors of pain trajectory following TKR. Preoperative variables are only modest predictors of function trajectory for THR and TKR. By contrast, data collected on OHS and OKS at six months, reliably predict the class of five-year outcome trajectory for both pain and function.

Our data, which show that individual-level patient-related variables at baseline have relatively little impact on the PROMs trajectory after THR and TKR, appear to contrast with the conclusions from a number of previous studies.[21,22] Previous literature has explored pre-operative predictors of outcomes. These have suggested that factors such as preoperative pain and function, radiological status and body mass index[23,24] as well as age and mental health score[21] are important predictors. These analyses have used regression approaches to identify associations. While these possible associations are reflected in our data, our analysis emphasises that these variables are, at best, only weakly predictive of outcome trajectory up to five years. Our data show that, irrespective of baseline preoperative characteristics, the most important predictor of long-term outcome is the patient post-operative state, captured at six month on questionnaires.

These findings need to be considered in the context of a number of limitations. As in all real-world observational studies, the analysis relies on patients to complete and return questionnaires. Our sample was based on consecutive six month PROMs returns and while we are unable to estimate the level of response at that timepoint directly, typical responses rates for hip and knee surgery of six to 12 month PROMs are reported in the region of 75%.[25] We note that the characteristics of patients in the initial sample who returned preoperative questionnaires are similar to those reported

in patients undergoing primary hip and knee surgery in the 8th NJR annual reports,[6] giving confidence that our sample is representative.

The rate of questionnaires returned fell, over time. Approximately, 40-45% of the subjects had data at all-time points once exclusion criteria were applied. Data on patient related variables are missing in a number of instances, and there were insufficient resources to be able to undertake a more rigorous programme of reminders for non-responders. While the response rate to our follow-up questionnaires was comparatively good for a study of this type,[26,27] data were missing from a large number of individuals. We also excluded data where there was more than one missing time point after six months. The nature of the available data available did not allow us to separately identify cases who died among those who did not return questionnaires.

We note that the characteristics of the non-responders and those otherwise excluded from the analysis were similar to those who contributed to the trajectory analysis. This suggests no important bias in our estimate of the lack of influence of baseline characteristics on predicting outcome trajectory class. However, the data that are available from the group that returned questionnaires of between one and three timepoints, suggests that the patients who were not included because of insufficient serial data, are likely to have outcomes similar to those in the Level 2 response category. This might have led to an underestimate of the proportion of those with poorer long-term outcome trajectories.

We chose to investigate longitudinal trajectories through k-means clustering procedure because we felt that it was most appropriate technique for the type of data that had been collected. Although the serial questionnaires were returned in sequence over a five-year period post-operatively, it was not possible to validate with precision the exact dates at which questionnaires were completed. It was therefore felt that the data was not sufficiently well-specified to support more sophisticated approaches including, for example, multilevel modelling. K-means clustering is, in effect, a nonparametric approach, in which it is not necessary to specify or model individual-level trajectories explicitly. Instead, it provides an appropriate way to summarise sets of complex trajectories to allow predictive factors to be explored. The Level 1 and 2 trajectories will encompass a range of individual patterns of both improvement and decline in pain and function scores over the follow-up period. The procedure does not account for censoring (through subsequent surgery, death, or loss to follow-up).

We acknowledge that better or improving scores for pain and function may not necessarily equate to satisfaction with the procedure. We deliberately chose the terms ‘Level 1’ and ‘Level 2’ to describe the two trajectories identified in this analysis rather than categorise these as ‘poor’ or ‘good’ responders. However, we note that the level of satisfaction was lower among those whose trajectory was classified into level 2.

One consequence of the approach we have used is that the findings are only generalisable to patients who survive for five years with their implant in situ. It should be noted, given the typical five-year revision rate for THR and TKR implants of around three percent,[6] this represents the majority who undergo joint replacement surgery. K-means is an exploratory approach, and there is no statistical test to indicate the clusters are real. However, it is of interest that when all the available PROMs data were analysed separately from patients who subsequently underwent surgery, we found that the PROMs trajectories mirrored closely those of the Level 2 response category. This lends a degree of construct validity to the two response categories as providing plausible representations of five-year outcomes.

These findings suggest that for both THR and TKR, while average post-operative PROMS continue to improve of the first post-operative year, the six month assessment of pain and function is a reliable indicator of the prognosis for the following five years and appears to be an optimal time-point for assessing post-operative patient outcomes. This information should help to inform the design for routine monitoring pathways for people following THR and TKR where the current evidence-base is still limited. Our data also show that baseline patient characteristics alone are poor predictors of future outcome and emphasise that these should not be used as a basis for rationing treatment.

DECLARATIONS AND ACKNOWLEDGEMENTS

Acknowledgement: We thank the patients and staff of all the hospitals in England, Wales and Northern Ireland who have contributed data to the National Joint Registry. We are grateful to the Healthcare Quality Improvement Partnership (HQIP), the NJR Research Committee and staff at the NJR Centre for facilitating this work. The authors have conformed to the NJR's standard protocol for data access and publication. The views expressed represent those of the authors and do not necessarily reflect those of the National Joint Registry Steering Committee or the Healthcare Quality Improvement Partnership (HQIP) who do not vouch for how the information is presented.

Funding: No funding was received to support this study. [redacted].

Sponsorship: This work was undertaken as part of the National Joint Registry strategic plan.

Conflict of Interest: All authors declare that they have no conflicts of interest in relation to this paper.

Ethical Approvals: Research ethics committee (REC) approval was granted by the NRES London – Central REC ([redacted]). The research project was internally approved by the NJR Research Committee.

Data sharing: Access to data is available from the National Joint Registry for England, Wales, Northern Ireland and the Isle of Man.

Patient and Public Involvement: Patients were involved in the reporting and dissemination plans of our research. Our PPI members provided valuable input in relation to the interpretation of our findings and how these may be most usefully presented to a wide number of stakeholders including patient and public members.

Contributorship Statement:
[redacted]

FIGURE AND TABLE LEGENDS

Figure 1: Trajectory of pain domain for people following THR and TKR from pre-operative to 5-year follow-up.

Figure 2: Trajectory of function domain for people following THR and TKR from pre-operative to 5-year follow-up.

Table 1: Percentage accuracy (with 95% Confidence Interval) of model classification for patients in each trajectory group

Supplementary Figure 1: Line graph to illustrate the optimal number of clusters was determined by a set of three non-parametric tests.

Supplementary Figure 2: Flow-chart of participants to illustrate identification of analysed cohort.

Supplementary Table 1: Characteristics for patients following THR (n=20,089) tabulated by trajectory group according to 'pain' and 'function' domains.

Supplementary Table 2: Characteristics for patients following TKR (n=22,489) tabulated by trajectory group according to 'pain' and 'function' domains.

Supplementary Table 3: Table to illustrate the baseline, 6-month, 1-year, 3-year and 5-year OHR and OKS scores (overall) and pain and function subsets for both missing and analysed THR and TKR cohorts.

REFERENCES

1. Konan S, Haddad FS. Joint Registries: a Ptolemaic model of data interpretation? Bone Joint J. 2013;95-B(12):1585-6.

2. Wylde V, Blom AW. The failure of survivorship. J Bone Joint Surg Br. 2011 Apr 21;93-B:569–70.

3. Pietrzak J, Common H, Migaud H, Pasquier G, Girard J, Putman S. Have the frequency of and reasons for revision total knee arthroplasty changed since 2000? Comparison of two cohorts from the same hospital: 255 cases (2013-2016) and 68 cases (1991-1998). Orthop Traumatol Surg Res. 2019;105(4):639-645.

4. Niinimäki TT. The reasons for knee arthroplasty revisions are incomparable in the different arthroplasty registries. Knee. 2015;22(2):142-4.

5. Rolfson O, Eresian Chenok K, Bohm E, Lübbeke A, Denissen G, Dunn J, Lyman S, Franklin P, Dunbar M, Overgaard S, Garellick G, Dawson J; Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries. Patient-reported outcome measures in arthroplasty registries. Acta Orthop. 2016;87 Suppl 1:3-8.

6. NJR. National Joint Registry. Available at: www.njr.org.uk. Accessed: 18 December 2019.

7. Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A. How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. Lancet. 2019 Feb 16;393(10172):647–54.

8. Evans JT, Walker RW, Evans JP, Blom AW, Sayers A, Whitehouse MR. How long does a knee replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. Lancet. 2019 Feb 16;393(10172):655–63.

9. Clohisy JC, Kamath GV, Byrd GD, Steger-May K, Wright RW. Patient compliance with clinical follow-up after total joint arthroplasty. J Bone Joint Surg Am. 2008;90:1848-54.

10. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol 2008;61:344-9

11. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, Dawson J. The use of Oxford hip and knee scores. J Bone Joint Surg. 2007;89:1010-1014.

12. 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(2):185-90.

13. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. J Bone Joint Surg Br. 1998;80(1):63-9.

14. Harris KK, Dawson J, Doll H, Field RE, Murray DW, Fitzpatrick R, Jenkinson C, , Price AJ, Beard DJ. Can pain and function be distinguished in the Oxford Hip Score in a meaningful way? An exploratory and confirmatory factor analysis. Qual Life Res. 2013; 22: 2561-2568.

15. Harris KK, Price AJ, Beard DJ, Fitzpatrick R, Jenkinson C, Dawson J. Can pain and function be distinguished in the Oxford Hip Score in a meaningful way? An exploratory and confirmatory factor analysis. Bone Joint Res. 2014; 3: 305-9.

16. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R Packages to Cluster Longitudinal Data. Journal of Statistical Software 2015; 65: 1-34.

17. Calinski T, Harabasz J. A dendrite method for cluster analysis. Communicat Statistics Theor Method 1974; 1–27

18. Ray S, Turi RH. Determination of number of clusters in K-means clustering and application in colour image segmentation." In Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques 1999; 137–43.

19. Davies DL, Bouldin DW (1979). "A Cluster Separation Measure." IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), 224–227.

20. Morrison DG. On the interpretation of discriminant analysis. Journal of Marketing Research. 1969;6(2):156-163.

21. Arden N, Altman D, Beard D, Carr A, Clarke N, Collins G, Cooper C, Culliford D, Delmestri A, Garden S, Griffin T, Javaid K, Judge A, Latham J, Mullee M, Murray D, Ogundimu E, Pinedo-Villanueva R, Price A, Prieto-Alhambra D, Raftery J. Lower limb arthroplasty: can we produce a tool to predict outcome and failure, and is it cost-effective? An epidemiological study. Southampton (UK): NIHR Journals Library - Programme Grants for Applied Research; 2017 Jun.

22. Judge A, Arden NK, Price A, Glyn-Jones S, Beard D, Carr AJ, Dawson J, Fitzpatrick R, Field RE. Assessing patients for joint replacement: can pre-operative Oxford hip and knee scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection? J Bone Joint Surg Br. 2011;93(12):1660-4.

23. Hofstede SN, Gademan MG, Vliet Vlieland TP, Nelissen RG, Marang-van de Mheen PJ. Preoperative predictors for outcome after total hip replacement in patients with osteoarthritis a systematic review. BMC Musculoskeletal Disord 2016; 17: 212.

24. Bido J, Yang YH, Collins JE, Dong Y, Driscoll DA, Alcantara LA, Thornhill TS, Katz JN. Predictors of patient-reported outcomes of total joint arthroplasty in a developing country. J Arthroplasty 2017; 32: 1756-1762.

25. Wang K, Eftang CN, Jakobsen RB, Årøen A. Review of response rates over time in registry-based studies using patient-reported outcome measures. BMJ Open 2020;10: e030808.

26. Reinholdsson J, Kraus-Schmitz J, Forssblad M, Edman G, Byttner M, Stålman A. A non-response analysis of 2-year data in the Swedish Knee Ligament Register. Knee Surg Sports Traumatol Arthrosc. 2017;25(8):2481-2487.

27. Imam MA, Barke S, Stafford GH, Parkin D, Field RE. Loss to follow-up after total hip replacement: a source of bias in patient reported outcome measures and registry datasets? Hip Int. 2014;24(5):465-72.

T

Table 1: Percentage accuracy* (with 95% Confidence Interval) of model classification for patients in each trajectory group

Explanatory variable(s)	**Group**	**Outcome variable (THR)**				**Outcome variable (TKR)**			
		Pain domain	**PCC****, p-value	**Function domain**	**PCC****, p-value	**Pain domain**	PCC**, p-value	**Function domain**	PCC**, p-value
Pre-operative variables***	Level 1	99 (96, 100)	1.000	94 (91, 96)	0.971	97 (95, 99)	0.993	87 (81, 92)	0.110
	Level 2	3 (0, 11)		21 (13, 27)		11 (7, 16)		34 (25, 44)	
Baseline OHS/OKS	Level 1	98 (96, 100)	1.000	93 (92, 95)	0.299	92 (86, 95)	0.361	85 (82, 87)	**<0.0001**
	Level 2	8 (0, 13)		28 (24, 31)		27 (19, 40)		48 (45, 54)	
Pre-operative variables + Baseline OHS/OKS	Level 1	95 (93, 98)	1.000	92 (89, 95)	**0.0002**	92 (87, 95)	**0.014**	84 (80, 88)	**<0.0001**
	Level 2	18 (9, 25)		36 (27, 45)		30 (21, 40)		55 (46, 61)	
6-month OHS/OKS	Level 1	95 (95, 96)	**<0.0001**	94 (93, 95)	**<0.0001**	93 (93, 95)	**<0.0001**	90 (88, 91)	**<0.0001**
	Level 2	73 (70, 76)		74 (71, 76)		77 (73, 79)		76 (74, 79)	
Baseline + 6 month OHS/OKS	Level 1	95 (94, 96)	**<0.0001**	93 (92, 94)	**<0.0001**	94 (91, 95)	**<0.0001**	91 (88, 93)	**<0.0001**
	Level 2	75 (73, 77)		79 (76, 82)		77 (72, 82)		78 (76, 84)	
6 month + 1-year OHS/OKS	Level 1	95 (95, 97)	**<0.0001**	95 (93, 96)	**<0.0001**	95 (94, 96)	**<0.0001**	92 (91, 93)	**<0.0001**
	Level 2	82 (78, 85)		81 (77, 85)		83 (80, 86)		81 (79, 84)	
Baseline + 6 month + 1-year OHS/OKS	Level 1	95 (94, 97)	**<0.0001**	95 (94, 96)	**<0.0001**	95 (93, 96)	**<0.0001**	93 (92, 94)	**<0.0001**
	Level 2	82 (78, 86)		84 (81, 86)		84 (82, 87)		84 (82, 86)	

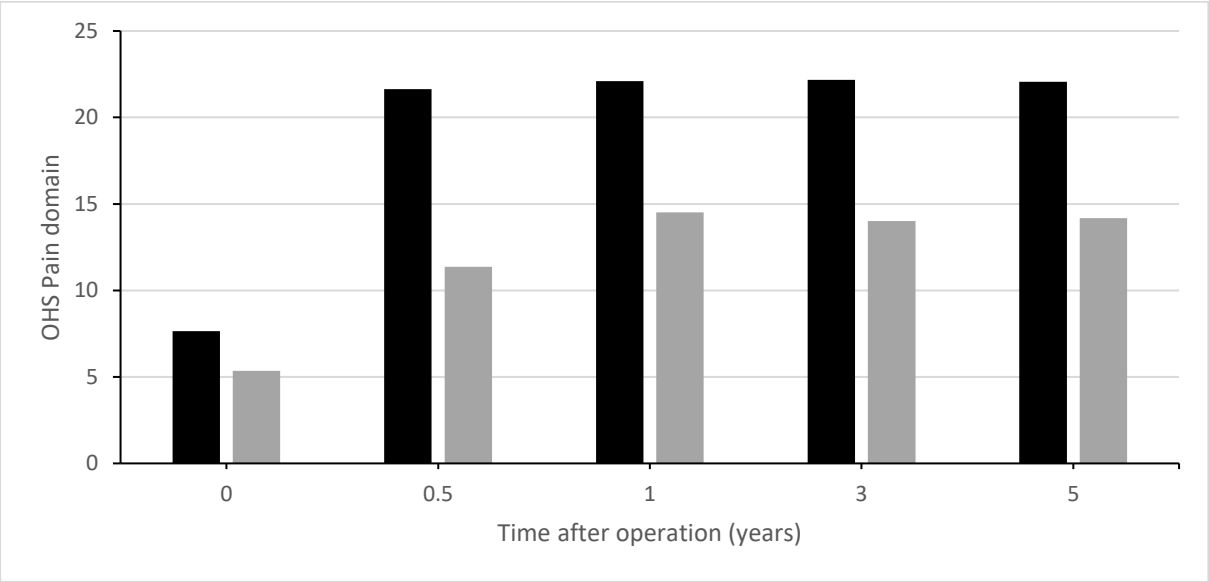
OHS – Oxford Hip Score; OKS – Oxford Knee Score; SD – standard deviation; THR – total hip replacement; TKR – total knee replacement

* Percentage accuracy is defined as the proportion classified into the correct trajectory group in the test data set using the model developed from the training data set

** Proportional Chance Criterion – this is used to test the predictive performance of each model. A result is considered significant if P<0.05 and indicates that the classification achieved by the model was significantly better than could reasonably expected by chance

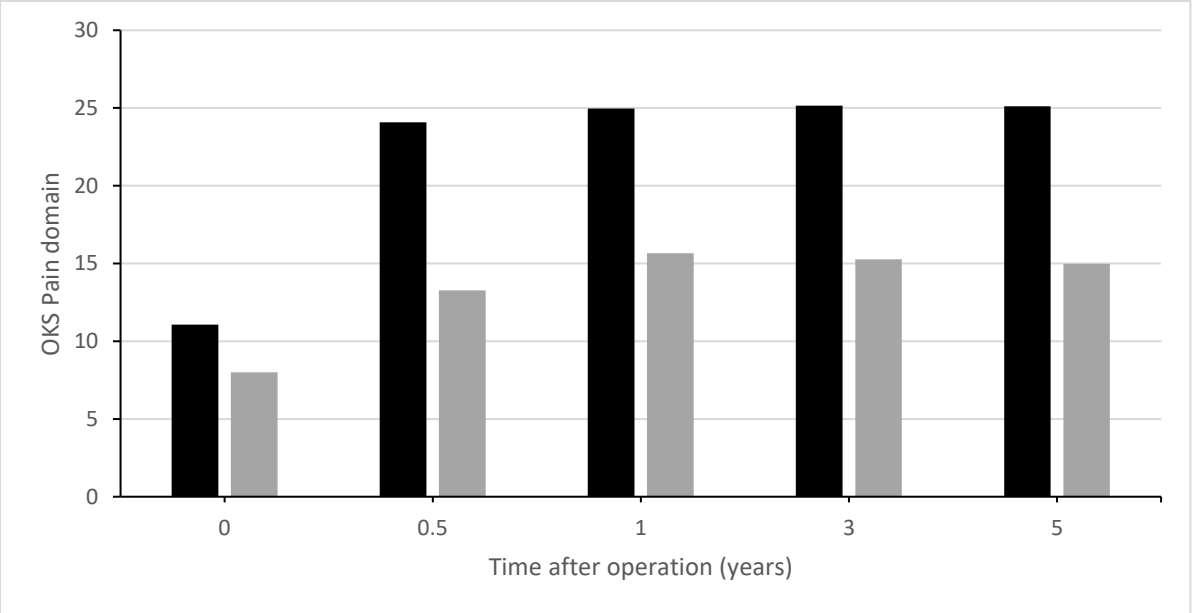
*** Age, BMI, gender, social deprivation (IMD), ethnicity, rural/urban location, ASA grade

Figure 1: Trajectory of pain domain for people following THR and TKR from pre-operative to 5-year follow-up. Black bars represent Level 1 patients; grey bars represent Level 2 patients.



Time post-operative (yrs)	0	0.5	1	3	5
Level 1					
N	10,675	10,649	10,628	10,236	7,086
Mean (95% CI)	7.64 (7.56, 7.72)	21.64 (21.59, 21.69)	22.09 (22.05,22.14)	22.17 (22.12, 22.22)	22.07 (22.00, 22.13)
Level 2					
N	3,633	3,610	3,608	3,345	2,274
Mean (95% CI)	5.34 (5.23, 5.46)	11.36 (11.15, 11.58)	14.51 (14.34, 14.68)	14.01 (13.83, 14.2)	14.18 (13.94, 14.42)

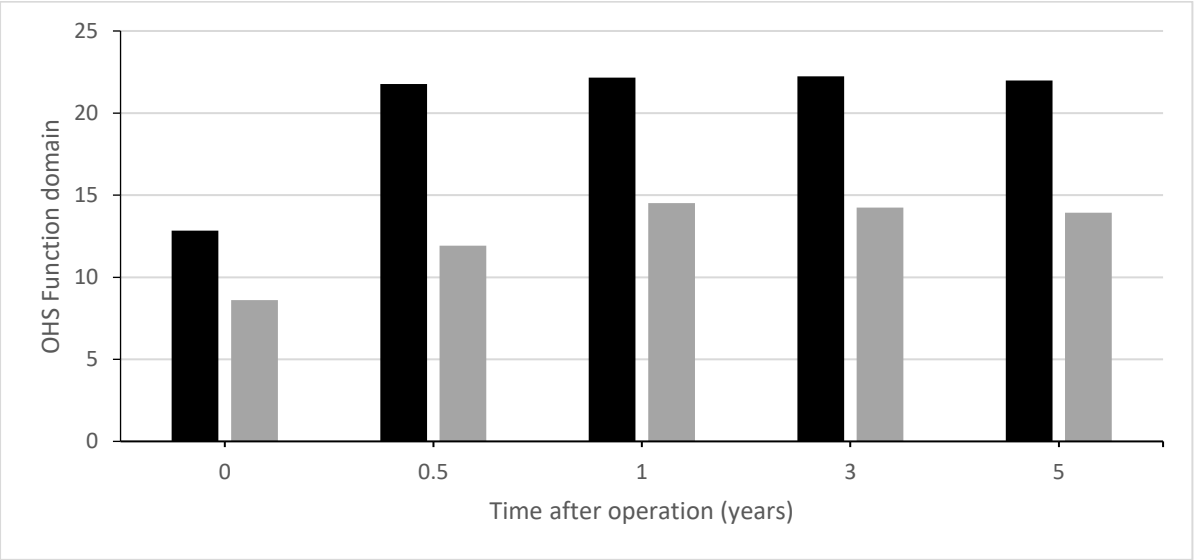
CI – confidence intervals; OHS - Oxford Hip Score; yrs - years



Time post-operative (yrs)	0	0.5	1	3	5
Level 1					
N	11,055	11,021	11,005	10,578	7,369
Mean (95% CI)	11.06 (10.98,11.15)	24.08 (24.01, 24.14)	24.96 (24.90, 25.01)	25.15 (25.09, 25.21)	25.10 (25.02, 25.18)
Level 2					
N	4,862	4,847	4,847	4,463	2,978
Mean (95% CI)	8.01 (7.90, 8.13)	13.27 (13.07, 13.46)	15.66 (15.50, 15.81)	15.27 (15.10, 15.43)	14.97 (14.76, 15.19)

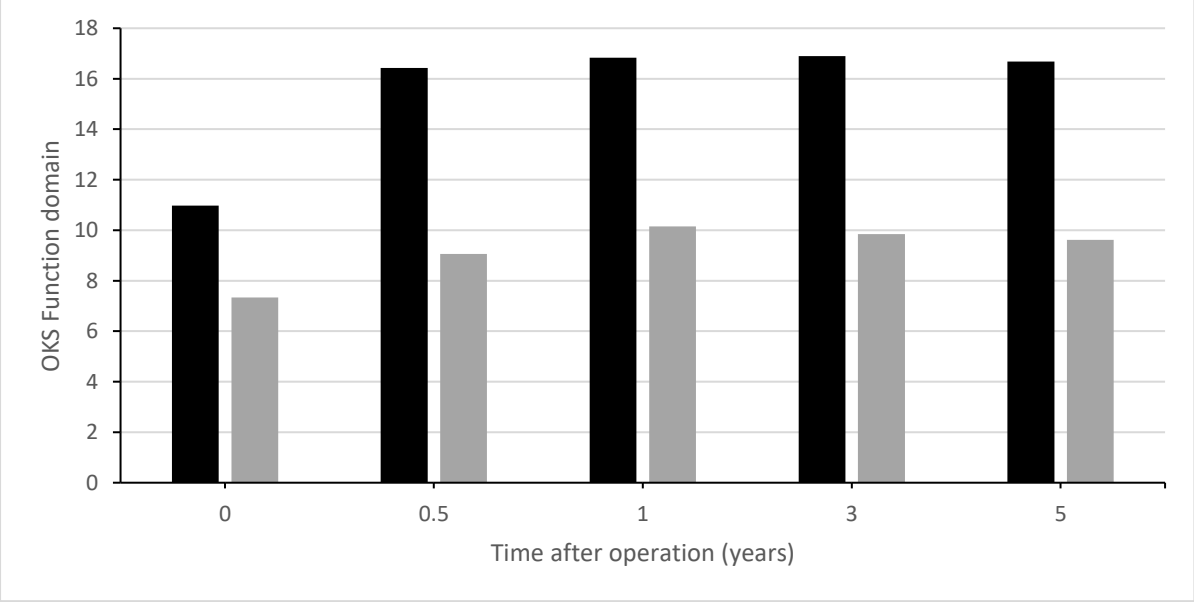
CI – confidence intervals; OKS - Oxford Knee Score; yrs - years

Figure 2. Trajectory of function domain for patients following THR and TKR from pre-operative to 5-year follow-up. Black bars represent Level 1 patients; grey bars represent Level 2 patients.



Time post-operative (yrs)	0	0.5	1	3	5
Level 1					
N	10,293	10,262	10,254	9,935	6,910
Mean (95% CI)	12.84 (12.76,12.93)	21.76 (21.72, 21.81)	22.16 (22.12, 22.2)	22.24 (22.19, 22.29)	21.99 (21.93, 22.05)
Level 2					
N	4,051	4,016	4,008	3,735	2,459
Mean (95% CI)	8.60 (8.48, 8.72)	11.93 (11.73, 12.12)	14.51 (14.37, 14.66)	14.25 (14.09, 14.4)	13.94 (13.74, 14.13)

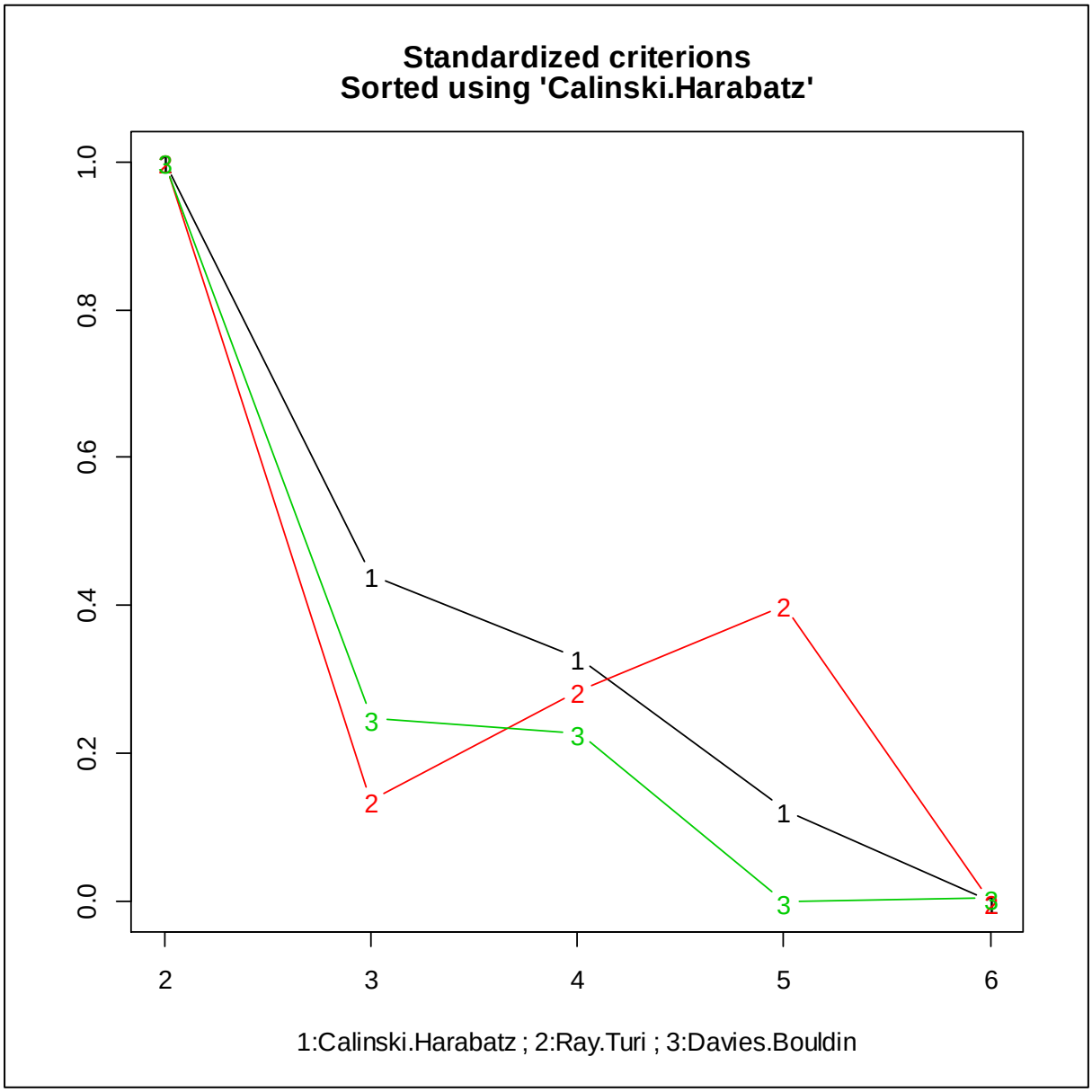
CI – confidence intervals; OHS - Oxford Hip Score; yrs - years



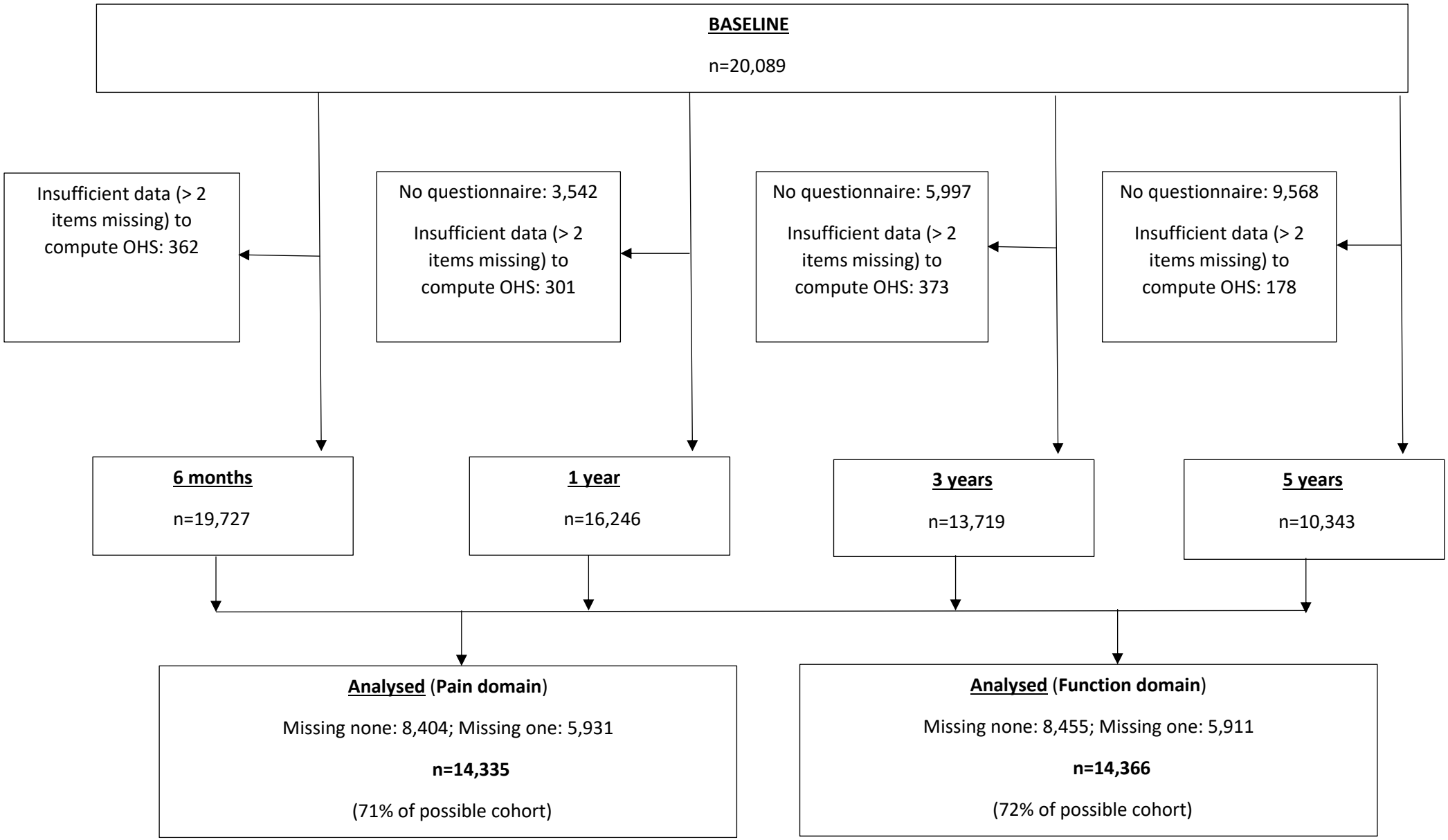
Time post-operative (yrs)	0	0.5	1	3	5
Level 1					
N	8,912	8,769	8,562	8,325	6,284
Mean (95% CI)	10.96 (10.90,11.03)	16.15 (16.08, 16.22)	16.69 (16.64, 16.74)	16.97 (16.92, 17.02)	16.70 (16.63, 16.76)
Level 2					
N	5,221	5,132	4,987	4,638	3,365
Mean (95% CI)	7.21 (7.13, 7.29)	10.03 (9.93, 10.13)	10.23 (10.14, 10.33)	9.58 (9.49, 9.68)	9.42 (9.30, 9.54)

CI – confidence intervals; OKS - Oxford Knee Score; yrs - years

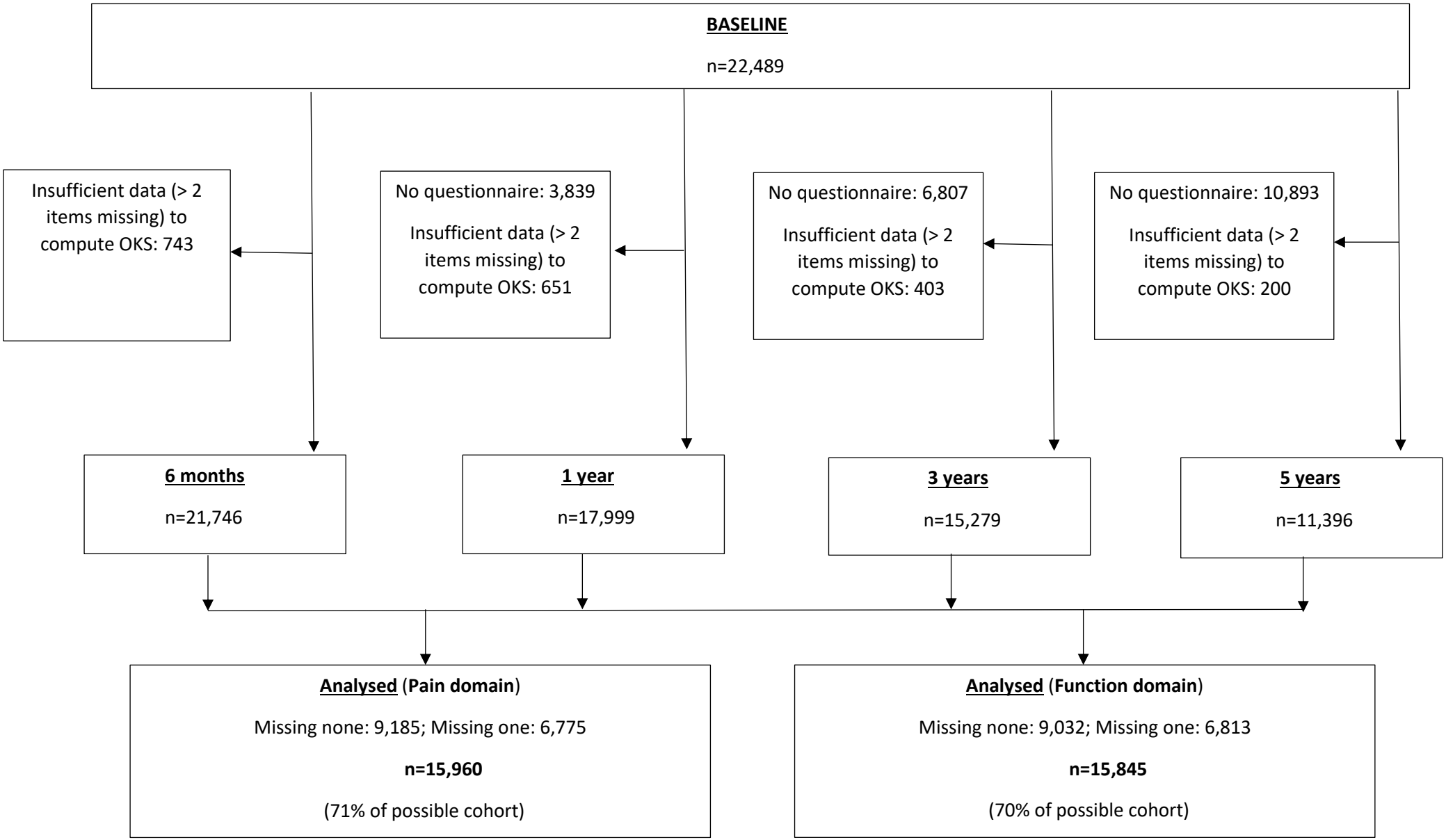
Supplementary Figure 1: Line graph to illustrate the optimal number of clusters was determined by a set of three non-parametric tests.



Supplementary Figure 2a: Flow-chart of THR participants to illustrate identification of analysed cohort



Supplementary Figure 2b: Flow-chart of TKR participants to illustrate identification of analysed cohort



Supplementary Table 1: Characteristics for patients following THR (n=20,089) tabulated by trajectory group according to 'pain' and 'function' domains.

	Pain				**Function**			
	Missing †	**Level 1**	**Level 2**	**OR (95% CI)***	**Missing †**	**Level 1**	**Level 2**	**OR (95% CI)***
n (% of total)	5,754 (29)	10,695 (53)	3,640 (18)		5,723 (28)	10,311 (60)	4,055 (20)	
Age [Mean (SD)]	67.2 (13.2)	68.4 (9.6)	68.5 (10.2)	1.01 (1.00, 1.01)	67.3 (13.2)	67.6 (9.4)	70.4 (10.4)	1.03 (1.03, 1.04)
BMI [Mean (SD)]	29.0 (5.5)	28.2 (4.8)	29.6 (5.4)	1.06 (1.05, 1.07)	29.0 (5.5)	28.0 (4.7)	29.9 (5.5)	1.08 (1.07, 1.09)
IMD [Mean (SD)]	20.8 (10.1)	18.4 (9.1)	20.9 (10.0)	1.03 (1.02, 1.03)	20.9 (10.1)	18.4 (9.1)	20.7 (9.9)	1.03 (1.02, 1.03)
Gender								
Males (%)	40	41	35	reference	40	42	33	reference
Females (%)	60	59	65	1.32 (1.22, 1.43)	60	58	67	1.43 (1.33, 1.55)
Ethnicity								
White (%)	97	99	98	reference	97	99	98	reference
Non-white (%)	3	1	2	3.33 (2.29, 4.88)	3	1	2	2.80 (1.93, 4.09)
Geographical Location								
Rural (%)	31	37	32	reference	31	37	32	reference
Urban (%)	69	63	68	1.32 (1.21, 1.43)	69	63	68	1.30 (1.20, 1.40)
ASA grade								
Fit and Healthy (%)	13	16	11	Reference	13	18	8	Reference
Mild Disease (%)	67	73	72	1.47 (1.31, 1.65)	67	73	71	2.05 (1.82, 2.33)
Incapacitating (%)	20	11	17	2.27 (1.97, 2.63)	20	9	21	4.72 (4.07, 5.49)
Operation satisfaction (at one year post-op)								
Very good	71	98	75	Reference	63	98	79	Reference
Poor	29	2	25	20.24 (17.15, 24.04)	37	2	21	10.79 (9.32, 12.53)

ASA - American Society of Anesthesiologists; BMI – body mass index; IMD – index of multiple deprivation score; SD – standard deviation; THR – total hip replacement

† Missing data for trajectory group membership occurs when there are 1 (or more) missing values in response to the OHS domains from baseline, six months, one year, three year and five year questionnaires.

* OR=odds ratio for logistic regression for association of trajectory group (level1=0; level2=1) with variable

Supplementary Table 2: Characteristics for patients following TKR (n=22,489) tabulated by trajectory group according to 'pain' and 'function' domains.

	Pain				**Function**			
	Missing †	**Level 1**	**Level 2**	**OR (95% CI)**	**Missing †**	**Level 1**	**Level 2**	**OR (95% CI)**
n (% of total)	6,529 (29)	11,082 (49)	4,878 (22)		6,644 (30)	10,000 (44)	5,845 (26)	
Age [Mean (SD)]	68.5 (10.7)	69.3 (8.5)	68.8 (9.2)	0.99 (0.99, 1.00)	68.6 (11.0)	68.8 (8.5)	69.7 (9.8)	1.01 (1.01, 1.02)
BMI [Mean (SD)]	31.3 (5.6)	30.2 (5.0)	31.8 (5.6)	1.06 (1.05, 1.07)	31.3 (5.7)	30.0 (5.0)	32.0 (5.6)	1.08 (1.07, 1.09)
IMD [Mean (SD)]	21.7 (10.6)	19.2 (9.3)	21.8 (10.4)	1.03 (1.02, 1.03)	21.6 (11.0)	19.0 (9.4)	21.7 (10.4)	1.03 (1.02, 1.03)
Gender								
Males (%)	41	46	40	reference	41	50	34	reference
Females (%)	59	54	60	1.28 (1.20, 1.37)	59	50	66	1.93 (1.81, 2.07)
Ethnicity								
White (%)	94	99	94	reference	94	99	95	reference
Non-white (%)	6	1	6	4.83 (3.83, 6.14)	6	1	5	4.01 (3.16, 5.13)
Geographical Location								
Rural (%)	28	34	28	reference	28	34	28	reference
Urban (%)	72	66	72	1.46 (1.35, 1.57)	72	66	72	1.46 (1.35, 1.57)
ASA grade								
Fit and Healthy (%)	10	12	9	Reference	10	13	10	Reference
Mild Disease (%)	71	76	73	1.22 (1.09, 1.37)	71	77	71	1.64 (1.47, 1.85)
Incapacitating (%)	19	12	18	1.96 (1.71, 2.25)	19	10	19	3.25 (2.84, 3.73)
Operation satisfaction (at one year post-op)								
Very good	69	97	62	Reference	65	96	69	Reference
Poor	31	3	38	19.86 (17.57, 22.52)	35	4	31	10.99 (9.79, 12.36)

ASA - American Society of Anesthesiologists; BMI – body mass index; IMD – index of multiple deprivation score; SD – standard deviation; TKR – total knee replacement

† Missing data for trajectory group membership occurs when there are 1 (or more) missing values in response to the OKS ‘Pain’ and ‘Function’ domains from baseline, six months, one year, three year and five year questionnaires.

* OR=odds ratio for logistic regression for association of trajectory group (level1=0; level2=1) with variable

Supplementary Table 3: Table to illustrate the baseline, 6-month, 1-year, 3-year and 5-year OHR and OKS scores (overall) and pain and function subsets for both missing and analysed THR and TKR cohorts.

	Hip		**Knee**	
	Missing	Analysed	Missing	Analysed
OHS/OKS	5,806 (29)	14,283 (71)	6,763 (30)	15,726 (70)
Baseline [Mean (SD)]	15.8 (8.3)	18.7 (8.2)	16.8 (7.8)	19.8 (7.7)
6-month [Mean (SD)]	22.8 (15.5)	38.1 (11.5)	21.3 (17.2)	34.5 (11.4)
1-year [Mean (SD)]	35.6 (9.1)	40.2 (8.8)	32.3 (10.7)	36.5 (9.7)
3-year [Mean (SD)]	37.1 (9.2)	40.2 (9.2)	34 (11.1)	36.6 (10.1)
5-year [Mean (SD)]	36.6 (9.6)	40 (9.3)	33.3 (11.7)	36.4 (10.4)
Pain Domain	5,754 (29)	14,335 (71)	6,529 (29)	15,960 (71)
Baseline [Mean (SD)]	5.9 (4)	7.1 (4.1)	8.6 (4.7)	10.1 (4.7)
6-month [Mean (SD)]	11.7 (9.7)	19 (6)	13 (10.5)	20.8 (6.9)
1-year [Mean (SD)]	18.3 (5.8)	20.2 (4.7)	19.8 (6.7)	22.1 (5.9)
3-year [Mean (SD)]	19.3 (5.3)	20.2 (5)	20.5 (6.6)	22.2 (6.1)
5-year [Mean (SD)]	18.5 (6.2)	20.2 (5.1)	20.4 (7.3)	22.2 (6.3)
Function Domain	5,723 (28)	14,366 (72)	6,644 (30)	15,845 (70)
Baseline [Mean (SD)]	9.9 (4.8)	11.6 (4.7)	8.1 (3.7)	9.6 (3.7)
6-month [Mean (SD)]	11.2 (9.4)	19 (5.9)	8.3 (6.9)	13.7 (4.9)
1-year [Mean (SD)]	17.3 (5.7)	20 (4.6)	12.4 (4.7)	14.4 (4.3)
3-year [Mean (SD)]	17.7 (5.7)	20.1 (4.8)	13.3 (5)	14.4 (4.5)
5-year [Mean (SD)]	18.2 (5.8)	19.9 (4.9)	12.9 (5)	14.2 (4.6)

OHS – Oxford Hip Score; OKS – Oxford Knee Score; SD: standard deviation