■ ARTHROPLASTY

The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty

PART 1: THE DEVELOPMENT OF AN EVIDENCE-BASED MODEL LINKING PREOPERATIVE SCORE TO THE PROBABILITY OF GAINING BENEFIT FROM SURGERY

A. J. Price,
S. Kang,
J. A. Cook,
H. Dakin,
A. Blom,
N. Arden,
R. Fitzpatrick,
D. J. Beard

From Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, University of Oxford, Oxford, UK

Aims
To calculate how the likelihood of obtaining measurable benefit from hip or knee arthroplasty varies with preoperative patient-reported scores.

Methods
Existing UK data from 222,933 knee and 209,760 hip arthroplasty patients were used to model an individual's probability of gaining meaningful improvement after surgery based on their preoperative Oxford Knee or Hip Score (OKS/OHS). A clinically meaningful improvement after arthroplasty was defined as ≥ 8 point improvement in OHS, and ≥ 7 in OKS.

Results
The upper preoperative score threshold, above which patients are unlikely to achieve any meaningful improvement from surgery, is 41 for knees and 40 for hips. At lower scores, the probability of improvement increased towards a maximum of 88% (knee) and 95% for (hips).

Conclusion
By our definition of meaningful improvement, patients with preoperative scores above 41 (OKS) and 40 (OHS) should not be routinely referred to secondary care for possible arthroplasty. Using lower thresholds would incrementally increase the probability of meaningful benefit for those referred but will exclude some patients with potential to benefit. The findings are useful to support the complex shared decision-making process in primary care for referral to secondary care; and in secondary care for experienced clinicians counselling patients considering knee or hip arthroplasty, but should not be used in isolation.

Correspondence should be sent to A. J. Price; email: andrew.price@ndorms.ox.ac.uk

Introduction

Within the UK's National Health Service (NHS), more than 200,000 hip and knee arthroplasties are now performed each year, mostly in patients with painful osteoarthritis (OA).[1] In most cases, the procedures are highly successful in improving quality of life, although 10% to 20% of patients are dissatisfied with outcomes, mainly due to ongoing pain.[2] Following National Institute for Health and Care Excellence (NICE) guidance for managing OA, patients should be referred to secondary care to be assessed for a possible arthroplasty when pain persists after nonoperative care management.[3] In secondary care, shared decisions about surgery are based on balancing an individual's capacity to benefit from surgery and the potential risks associated with surgery (e.g. medical complications, revision, and dissatisfaction).[4]

Since 2009, patients in England have been asked to complete the Oxford Hip and Knee Score[5] (OHS/OKS) questionnaires preoperatively and six months after surgery. In addition to patient outcome evaluation, these data have been used to compare providers.[6-9] Over the last eight years, some Clinical Commissioning Groups have introduced guidelines for referral that have included a maximum preoperative Oxford score, set as a threshold that acts as hard boundary for referral or not. The thresholds have varied between 18 and 30 points.[10-13] The assumption underpinning this approach is that patients with higher preoperative scores have less capacity to benefit from

arthroplasty[4] and could be excluded from referral. However, there is currently no published evidence to support the use of thresholds or their current level.[14] In response to the need for greater evidence in this area, the National Institute of Health Research (NIHR) commissioned research to examine whether preoperative score thresholds for referral are appropriate, and to provide some evidence-based guidance for the selection of threshold levels.

The aim of the research was to: estimate and internally validate a model using existing data to link an individual's preoperative Oxford score to their probability of gaining a meaningful improvement after arthroplasty; quantify how the magnitude of change in Oxford score varies with preoperative score; and determine the effect of setting different and varying preoperative scores as thresholds for referral into secondary care.

Methods

Choice of the Oxford scores. As part of the wider project, we initially performed a comprehensive systematic review of the literature exploring all outcome measures used for total hip arthroplasty (THA) and total knee arthroplasty (TKA) patients.[15] This assessed the methodological quality and psychometric properties of 32 instruments. It was found that the best-performing condition-specific scores were the Oxford Hip Score, the Oxford Knee Score, and the Western Ontario and McMaster Universities Osteoarthritis Index[16] (WOMAC). As a result of their evidence base,[9] established and documented measurement properties,[17] consensus opinion, and existing widespread use in the NHS, OHS and OKS were selected for estimation of thresholds. These scores range between zero and 48 points (where zero indicates the most severe problems and higher values indicate better function) and have been shown to be validated measures of preoperative disease state and outcomes of arthroplasty.[5,18-20]

Dataset. The NHS patient-reported outcome measures (PROMs) dataset linked to Hospital Episode Statistics[21] (HES; 2009 to 2015) was used. Inclusion criteria were patients who had undergone primary hip or knee arthroplasty (including total knee, unicondylar knee and patellofemoral arthroplasty, and total and resurfacing hip arthroplasty) primarily for OA;[22] baseline and six-month postoperative Oxford scores are collected.

Definition of benefit after hip and knee arthroplasty. Benefit was defined as a meaningful improvement in OHS/OKS after surgery. An anchor-based best cut-point method for determining meaningful improvement using receiver operating characteristic (ROC) curves had previously been developed using an earlier extract of the same data source (NHS PROMs data 2009 to 2011).[17] Based on the minimal important change (MIC) identified in that study, we defined a clinically meaningful improvement ('much better' or 'a little better') after arthroplasty as comprising an eight points or greater improvement in OHS, and seven points or more for OKS. Although meaningful improvement could be defined using smaller values of OHS and OKS, we believe that the use of the calculated MIC, using established methodology, is the best estimate.

Modelling. We used two statistical models to address the first two aims of our study: to use existing data to link an individual's preoperative OHS/OKS to their probability of gaining a meaningful improvement after arthroplasty, and to quantify how the magnitude of change in OHS/OKS varies with preoperative score.

First, we modelled the probability of achieving a meaningful improvement directly using fractional polynomial logistic regression. This assessed the prognostic value of the preoperative OHS/OKS only, and with additional covariates.[23] The performance of the models with the preoperative score only and the final covariate model was assessed in terms of discrimination and calibration by calculating the area under the (receiver operating) curve (AUC) and producing calibration plots. The baseline covariates investigated in the logistic regression models were age (as a continuous variable), sex, the presence or absence of 12 patient-reported comorbidities (heart disease, high blood pressure, problems caused by stroke, leg pain when walking due to poor circulation, lung disease, diabetes, kidney disease, disease of the nervous system, liver disease, cancer, depression, and arthritis) and length of time with hip or knee symptoms (< one year, one to five years, six to ten years, > ten years).

Second, we used quantile regression to quantify how the magnitude of change in Oxford score varies with preoperative score: a calculation that was not possible with the first modelling approach. These models had a third-degree polynomial form and robust variance estimation and directly modelled the absolute magnitude of change across the range of preoperative scores for the respective centile.[24] We assessed the prediction accuracy of this modelling approach by comparing the estimated values against the observed percentile value across the preoperative score range. Additionally, we calculated the probability of a meaningful improvement from the respective quantile (1 to 99 percentile) model and compared this against the observed data and the estimate from the logistic regression model. Given the very large size of the available dataset, modelling performance was internally validated by assessing sensitivity to key factors (time period, sex, and age) through undertaking subset analyses and not by selecting a random sample or a bootstrapping approach. The sensitivity and specificity (with confidence intervals (CIs)) of using the respective preoperative threshold to identify those who had a meaningful improvement was calculated.

Complete case analyses were performed throughout, given the large samples available and the lack of information about reasons for values being missing. Stata v. 14 (StataCorp, College Station, Texas, USA) was used for all statistical analyses. For binary measures, 95% CIs were calculated using the binomial exact method.

Calculation of example thresholds. We defined the 'absolute threshold' as the level of preoperative OHS or OKS above which a patient had a 0% probability of achieving a meaningful improvement in score. Below the absolute threshold value, an individual's probability of achieving a good outcome changes as the preoperative score reduces. It is therefore possible to calculate preoperative scores (or thresholds) that equate to a predetermined probability of achieving a meaningful improvement. Specifically, we report example threshold OHS/OKS values at which the probability of meaningful improvement is 90%, 80%, 70%, and 50%.

Table I. Demographics of the populations studied.

Characteristic	Hip	Knee
Total patients	209,761	222,933
Mean age (SD; range)	68.4 (10.5; 13 to 100)	69.5 (8.9; 16 to 102)
Age category, n (%)*		
< 60 yrs	37,904 (18.1)	29,349 (13.2)
60 to 80 yrs	144,064 (68.7)	164,132 (73.6)
> 80 yrs	27,793 (13.2)	29,452 (13.2)
Sex, n (%)†		
Male	84,673 (40.4)	96,006 (43.1)
Female	125,058 (59.6)	126,885 (56.9)
Comorbidity, n (%)		
Heart disease	19,679 (9.4)	23,340 (10.5)
High blood pressure	82,428 (39.3)	102,542 (46.0)
Stroke	2,912 (1.4)	3,733 (1.7)
Peripheral vascular disease	11,968 (5.7)	16,464 (7.4)
Lung disease	15,592 (7.4)	18,571 (8.3)
Diabetes	18,449 (8.8)	27,789 (12.5)
Kidney disease	3,550 (1.7)	4,022 (1.8)
Nervous system disorder	1,566 (0.7)	2,155 (1.0)
Liver disease	1,081 (0.5)	1,199 (0.5)
Cancer	10,085 (4.8)	10,416 (4.7)
Depression	15,264 (7.3)	18,375 (8.2)
Arthritis	151,331 (72.1)	174,391 (78.2)
Comorbidities, n (%)		
0	29,933 (14.3)	22,121 (9.9)
1	79,168 (37.7)	73,599 (33.0)
2	63,076 (30.1)	74,649 (33.5)
3+	37,584 (17.9)	52,564 (23.6)
Year of NHS PROMs, n (%)		
2009 to 2011	96,041 (45.8)	102,448 (46.0)
2012 to 2015	113,720 (54.2)	120,485 (54.0)
Living arrangement, n (%)‡		
I live with someone	151,669 (73.6)	164,451 (75.3)
I live alone	53,318 (25.9)	52,985 (24.3)
I live in a care home	259 (0.1)	219 (0.1)
Other	769 (0.4)	802 (0.4)
Symptom period, n (%)§		
< 1 year	29,053 (13.9)	11,041 (5.0)
1 to 5 years	142,960 (68.5)	116,195 (52.4)
6 to 10 years	23,108 (11.1)	48,340 (21.8)
More than 10 years	13,588 (6.5)	46,273 (20.9)

*Data missing for one hip patient.
†Data missing for 30 hip patients and 42 knee patients.
‡Data missing for 3,746 hip patients and 4,476 knee patients.
§Data missing for 1,053 hip patients and 1,84 knee patients.
PROMs, patient-reported outcome measures.

For each of the threshold levels, we used PROMs data to calculate the proportion of patients who currently undergo hip or knee arthroplasty in England who would potentially have been restricted from referral if that threshold was applied.

Results
Descriptive and demographic statistics. The NHS PROMs cohort linked to HES for the period 2009 to 2015 included 277,786 hip and 298,194 knee procedures once duplicates were removed (Table I). Of these, 209,761 hips and 222,933 knee patients who completed pre- and postoperative OHS/OKS with sufficient procedure-specific data to derive scores were included in the
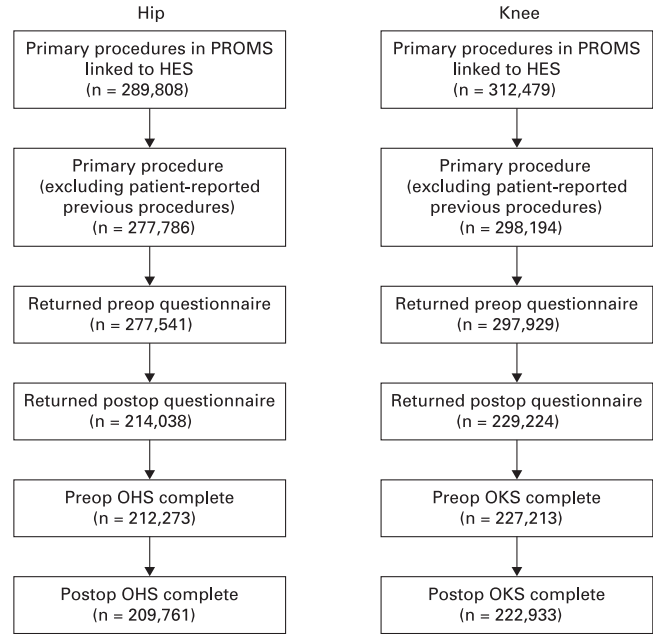
Hip

| Knee |

Fig. 1

Flowchart for hip and knee data sets. HES; Hospital Episode Statistics; OHS; Oxford Hip Score; OKS; Oxford Knee Score; PROMs, patient-reported outcome measures.

analysis (Figure 1). Individuals (< 1%) with partially completed OHS/OKS were excluded given the overall sample size. In the hip group, 59.6% were female, and there was a mean age of 68.4 years (12 to 100) (Table I). In the knee cohort, 56.9% were female, and there was a mean age of 69.5 years (16 to 102).

Absolute thresholds. We calculated the absolute threshold (above which there is 0% probability of a meaningful improvement) to be 40 points on the OHS, since patients with preoperative scores of 41 or higher cannot achieve an eight-point improvement in OHS. Similarly, the absolute OKS threshold (above which patients cannot achieve a seven-point improvement in OKS) was 41 points.

Linking preoperative score to probability of achieving a meaningful improvement. Figure 2 demonstrates how the probability of achieving a meaningful improvement in outcome varies according to preoperative OHS/OKS. All patients at or below the absolute threshold (40 points for hips and 41 points for knees) had, by definition, some chance of achieving a meaningful improvement from hip/knee arthroplasty. For hip arthroplasty, the maximum estimated probability of meaningful improvement was 95% (at a preoperative OHS score of eight points). The probability of meaningful improvement was over 90% at OHS scores between 0 and 23 points (Figure 2a). Between 24 and 38 points, the probability of benefit decreased slowly at first then more quickly as the preoperative OHS increased. Above a score of 38 points, where the probability of meaningful improvement was calculated to be about 50%, there was a faster drop to the absolute threshold of 40 points. Above 40 points there is, by definition, a zero probability of meaningful improvement, since these patients cannot achieve an eight-point improvement on the 48-point scale.
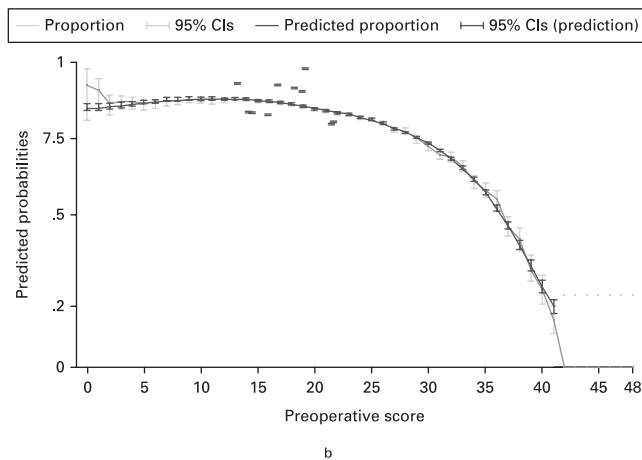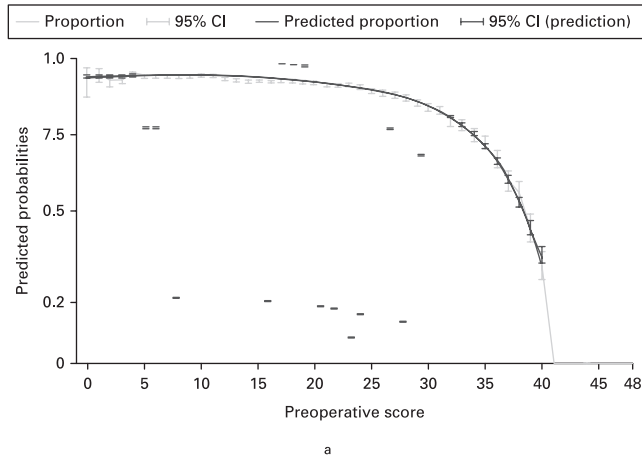
a

b

Fig. 2

Graphs showing observed and predicted probability of achieving a meaningful improvement plotted against preoperative Oxford score for a) hip (area under receiver operating characteristic (ROC) curve: 0.65 (95% confidence interval (CI) 0.64 to 0.65) and b) knee arthroplasty (area under ROC curve: 0.61 (95 % CI 0.61 to 0.62)), calculated using fractional polynomial logistic regression. In each plot, the light grey line with 95% CIs presents the observed proportion.
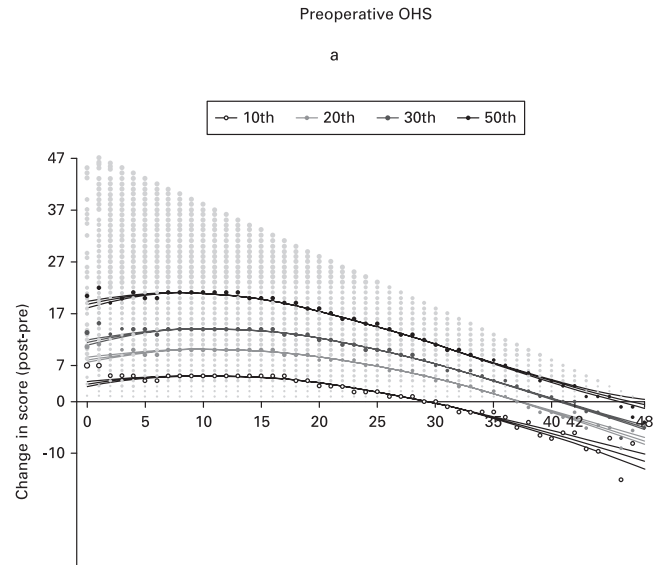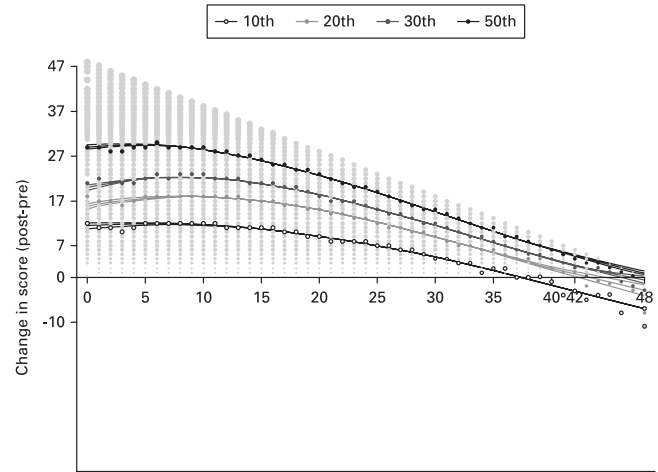

a

b

Fig. 3

Change from baseline Oxford Hip Score (OHS) and Oxford Knee Score (OKS) at the 10th, 20th, 30th, and 50th quantiles estimated with 95% confidence intervals (CIs) using the quantile regression model and based upon the observed data for a) hips and b) knees. Lightest grey dots indicate each observational percentile and all other dots indicate the observation density thresholds

The overall pattern for knee arthroplasty was similar to hip arthroplasty (Figure 2b). The peak probability of meaningful improvement of 88% was seen at 11 points. In contrast to the hip model, there was some suggestion of a modest decrease in probability of meaningful improvement decreased this, reaching a minimum value of around 85%. Above 20 points, the probability of improvement reduced more quickly with increasing preoperative scores; however, the curve drops less steeply than that for hips, over a much wider range of scores. For scores above 41 points (the absolute threshold), applying our definition there is no chance of a meaningful improvement.

The quantile regression models demonstrated that for almost all preoperative scores below the absolute threshold, most patients achieved improvements from hip and knee arthroplasty that were greater than the minimally important clinical difference, while a minority of patients did not achieve a meaningful improvement (Figure 3).

Calculation of example threshold levels. A selection of different example threshold levels were explored. For patients undergoing hip arthroplasty, setting a preoperative OHS threshold of 35 points would identify patients with a ≥ 70% chance of meaningful improvement. The other specific example relative threshold levels calculated for hip arthroplasty were 23, 32, and 38 points respectively. The proportion of patients who would have a meaningful improvement at each of these preoperative score categories was 90%, 80%, and 50% respectively (Table II). For knee arthroplasty, a threshold of 31 points would identify patients with a ≥ 70% chance of improvement. Relative thresholds of 25 and 36 points would identify patients with 80% and 50% chance of meaningful improvement, respectively (Table III).

Table II. Relative threshold values for hip arthroplasty at which 90%, 80%, 70%, and 50% of subjects achieved the meaningful improvement criteria of eight points on the Oxford Hip Score, for both observed data and predictions from quantile regression.

Baseline covariate	Total patients	Predicted preoperative score corresponding to preselected percentage probability of improvement.				Observed score			
		90%	80%	70%	50%	90%	80%	70%	50%
Total	209,761	223	32	35	38	24	31	35	38
Age category, n									
< 60 yrs	37,904	26	33	36	38	27	32	38	38
60 to 80 yrs	144,064	24	32	35	38	24	32	35	38
> 80 yrs	27,793	19	27	32	36	19	26	31	36
Sex, n*									
Male	84,673	25	32	35	38	26	33	35	38
Female	125,058	22	31	34	38	23	31	35	38
Year of NHS PROMs, n									
2009 to 2011	96,041	22	31	34	38	20	31	35	38
2012 to 2015	113,720	25	32	35	38	25	31	35	38

*Data missing for 30 patients.
PROMs, patient-reported outcome measures.

Table III. Relative threshold values for knee arthroplasty at which patients have a 90%, 80%, 70%, or 50% probability of achieving the meaningful improvement criteria of seven points on the Oxford Knee Score, for both observed data and predictions from quantile regression.

Baseline covariate	Total patients	Predicted preoperative score corresponding to preselected percentage probability of improvement				Observed score			
		90%	80%	70%	50%	90%	80%	70%	50%
Total, n	222,933	N/A	25	31	36	1	25	30	36
Age category, n									
< 60 yrs	29,349	N/A	18	29	35	1	21	30	35
60 to 80 yrs	164,132	N/A	26	31	36	1	26	31	36
> 80 yrs	29,452	13	24	30	35	15	26	30	36
Sex, n*									
Male	96,006	N/A	26	31	36	0	26	32	36
Female	126,885	N/A	25	30	35	1	25	30	36
Year of NHS PROMs, n									
2009 to 2011	102,448	N/A	24	30	35	1	23	30	36
2012 to 2015	120,485	N/A	27	31	36	1	26	32	36

*Data missing for 42 patients.
N/A, not applicable; PROMs, patient-reported outcome measures.

Model performance. The fractional polynomial logistic regression demonstrated that both OHS and OKS had prognostic value (based upon AUC around 0.6) for predicting whether patients have a meaningful improvement (Figures 3a and 4a). However, there was still substantial unexplained variation between patients. Following the model building process, the AUC was only marginally improved by the final models, which contained the selected covariates (Figures 1 and 2). Quantile regression results showed good correspondence between observed and estimated change (Figure 3). The quantile regression model did not show much difference across subpopulations (internal validation). There was also a good level of agreement of the proportion achieving a meaningful improvement based upon the quantile regression models and the corresponding values from the fractional polynomial logistic model and the observed data.

Effect of thresholds on current provision of hip and knee arthroplasty in the NHS. Based on the PROMs data, less than 1% of hip and knee arthroplasties performed between 2009 and 2016 were undertaken in patients with preoperative scores above our calculated absolute thresholds. This is reassuring for current practice and suggests that introducing such an absolute threshold for potential to benefit would not restrict access to care for most of the patients currently selected for arthroplasty. However, introducing lower thresholds could potentially restrict access to care for some patients (Figure 4). For example, setting a threshold OHS of 35 or OKS of 31 points (which would identify patients with a ≥ 70% probability of meaningful improvement) would have precluded referral in 2% of current hip arthroplasty patients (approximately 2,000 patients/year in England) and 6% (approximately 6,000/year) of knee arthroplasty patients, respectively. A threshold set at 25 points for knee patients would identify patients with an 80% or greater probability of meaningful improvement, but would exclude 20% of current patients. For patients with hip OA, a threshold of 23 would identify those with a 90% or greater probability of meaningful improvement, but would exclude 26% of current patients. In these circumstances a significant proportion of those excluded in each model would potentially have done well from surgery.
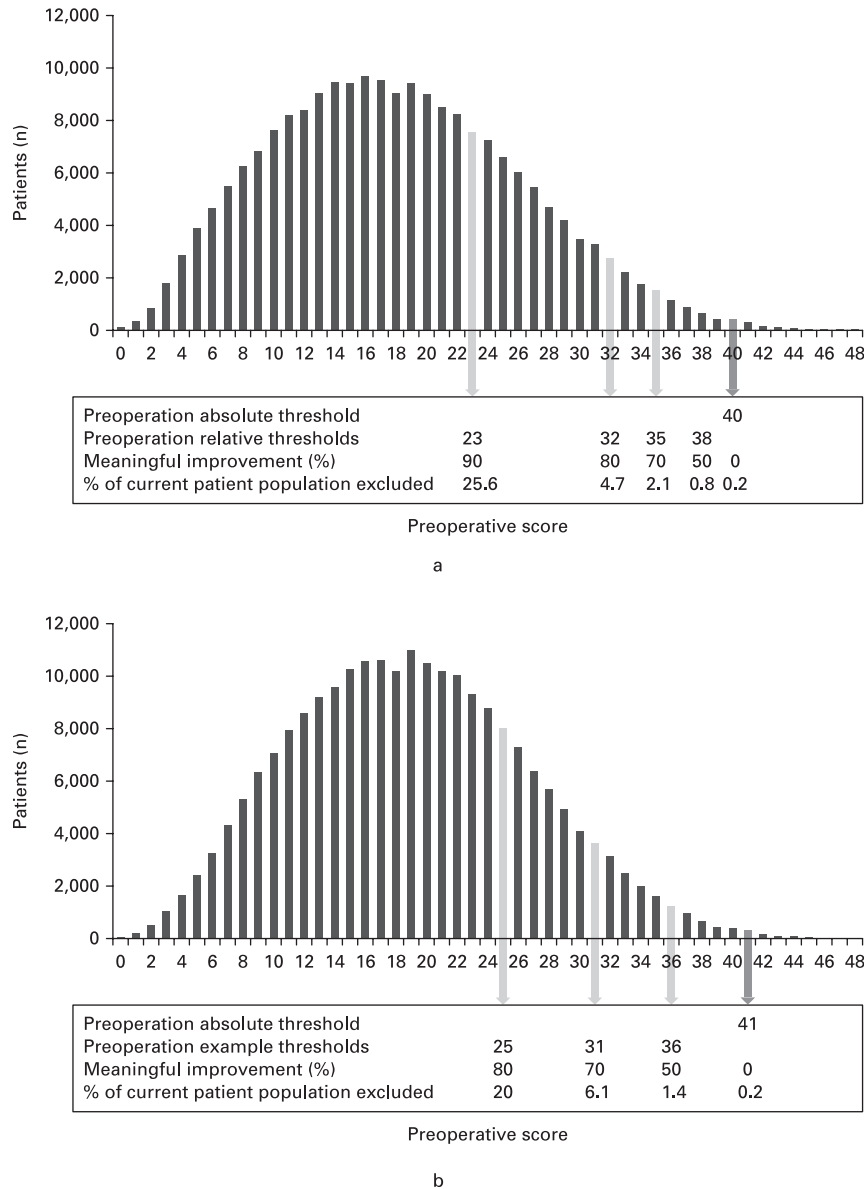
Preoperation absolute threshold				40	
Preoperation relative thresholds	23	32	35	38	
Meaningful improvement (%)	90	80	70	50	0
% of current patient population excluded	25.6	4.7	2.1	0.8	0.2

Preoperative score

a

Preoperation absolute threshold				41
Preoperation example thresholds	25	31	36	
Meaningful improvement (%)	80	70	50	0
% of current patient population excluded	20	6.1	1.4	0.2

Preoperative score

b

Fig. 4

Frequency plots of the preoperative scores recorded for a) hip and b) knee arthroplasty between 2009 and 2015, showing the effect of preoperative thresholds on current practice.

Discussion

We have estimated a model that links a patient's preoperative disease status, measured by a valid PROM, to their probability of gaining meaningful improvement from hip or knee arthroplasty. The results provide absolute preoperative Oxford score thresholds, above which patients have no capacity (0% probability, as defined by this study) to achieve clinically meaningful benefit from arthroplasty surgery: 40 points for hip patients and 41 points for knee patients. Above these thresholds, patients may increase their score postoperatively but cannot reach our definition of meaningful improvement. It is our opinion that they should probably not be routinely referred for consultation with an arthroplasty surgeon for consideration for hip and knee arthroplasty. Exceptions to this may (and will) occur such as individuals with an Oxford score above the absolute threshold for whom arthroplasty is required to treat other reasons beyond existing pain and dysfunction (e.g. progressive deformity).

No previously published studies have used preoperative PROMs to determine a threshold for referral. This lack of evidence has been a major criticism of previous use of Oxford scores to set referral levels. Previous work has shown that preoperative functional status is the strongest predictor of an individual's final functional outcome.[25-27] These findings support our model of using the preoperative Oxford score to predict an individual's capacity to improve. Importantly, our study supports the findings of many other studies observing that up to 85% of patients undergoing knee arthroplasty and 90% undergoing hip arthroplasty register meaningful improvement, with many demonstrating very large improvements.[5,6]

The model predicts an individual patient's probability of gaining meaningful improvement for any preoperative OHS/OKS below the absolute threshold; in other words it indicates whether arthroplasty will provide benefit for an individual patient. It shows that most patients have the potential to benefit from hip and knee arthroplasty.

We also assessed the effect of applying different thresholds for referral on current practice. Our analysis shows that some of the lower referral threshold levels that have been used within the NHS (e.g. an OKS of 20 points) would exclude significant numbers of patients who potentially have much to gain from surgery. Introducing the absolute threshold (e.g. an OKS of 41 points) would only limit access to surgery for less than 1% of patients who currently undergo hip and knee arthroplasty. This suggests that the current referral and assessment process in the UK does not result in many patients undergoing surgery where there is no potential for them to benefit. Lower absolute thresholds would exclude an increasing numbers of patients who currently undergo arthroplasty, despite them having some potential to benefit.

It is important to note that the model provides evidence about who may benefit from surgery and therefore could be referred for surgical assessment. It does not identify who should or will undergo surgery. This requires further detailed assessment and a shared decision-making process in secondary care, including assessment of medical risks and the likelihood of further surgery. However, the predicted probability of achieving a good outcome could be usefully integrated into a discussion regarding referral or the final decision to undergo surgery and this may be a useful application of our work. A web-based Arthroplasty Candidacy Help Engine (ACHE tool) using these data is under development and testing.

The implementation of the absolute threshold in patient-orientated decision-making is relatively straightforward to adopt. Above this level, a patient's score may improve after surgery, though they are unlikely to have a clinically meaningful increase in quality of life. A consultation between a suitably qualified clinician and patient with preoperative OKS/OHS above this threshold should lead to referral only in exceptional circumstances. The evidence should protect patients against undergoing surgery that is unlikely to be of value. Furthermore, the accompanying paper[28] demonstrates that hip and knee arthroplasty is highly cost-effective at the absolute thresholds, and that there is no health economic basis for restricting access to care for patients with OHS below 45 points or OKS below 43 points. Even these higher boundaries, if strictly applied, could exclude some who would benefit due to measurement error and the benefits not detectable with patient-reported pain and function measures.

Implementation of lower thresholds for referral (i.e. creating a hard boundary for referral or not) is less straightforward and requires further consideration. Our work demonstrates that any lower threshold would inevitably exclude some patients who would potentially benefit from surgery. The difficulty around setting any minimum OKS/OHS threshold has already been recognized by some UK commissioning groups. They have moved away from setting formal thresholds for referral, avoiding a postcode lottery of access to hip and knee arthroplasty and the gaming that a threshold may introduce.[13]

It is important to reflect on our chosen definition for ascribing benefit. We chose the specific gain of at least seven and eight points in OKS and OHS, respectively, to define meaningful benefit. This was based on our previously published work where the Minimally Important Change (MIC) for individual level analysis was calculated using ROC analysis[14] . Whatever the level picked it may exclude some individuals who gain less than the MIC but nonetheless have improved quality of life and are very satisfied with the outcome of their surgery. Our analysis could not include all possible confounding variables (for example, socioeconomic status) that could have affected the model, as we were limited to the variables included within the existing data. However, our study used a very large and rich dataset collected over seven years, which enables robust and reliable estimates of the model's performance. Internal validation showed the model results to not be affected much by the explored factors. There was a substantial amount of patient post-surgical outcome variability which was unexplained (reflected in the low AUC value) making the models of limited use to predict the outcome of an individual based upon the available preoperative information.

A limitation to note is that the current work does not predict the probability that patients would be made worse by surgery. Showing the potential to gain benefit is different from predicting the likelihood of doing poorly. We accept that some patients may be worsened but have not included this in the present model or the decision-making process. Such differences are subtle, but this understanding is essential to how a new ACHE tool could work in practice and be usefully integrated into clinical pathways. The accompanying paper[28] also takes account of surgical mortality, revisions, and the magnitude of quality of life changes (both above and below the MIC) when assessing how cost-effectiveness varies with preoperative OHS/OKS.

Notional thresholds are not designed to, nor should, reduce or remove any autonomy from clinical decision-making. It would be wrong to use any threshold as a standalone barrier to referral as each patient's symptoms and clinical problem must be assessed on an individual basis. They can be used to help both the healthcare professional and the patient with shared decision-making. Any consultation using thresholds for arthroplasty should be conducted by professionals with appropriate levels of competency and knowledge in the management of musculoskeletal disease. In complex situations, such as progressive deformity with limited pain, referral for expert assessment by an orthopaedic surgeon is required regardless of the Oxford score. Following published NICE treatment guidance for managing OA, patients who are being considered for referral should have a persistent and intrusive level of preoperative symptoms that could be improved by surgery.[3] A patient must weigh up the potential benefits of the procedure against the real risks of surgery (e.g. infection, medical complications, revision surgery, and poor functional outcome). The most effective way of using the model we have produced may be to support the shared decision-making process for patients, beginning at the referral stage and extending into secondary care. Our model could be used to inform patients of their likelihood

of improvement following surgery and, as such, the tool helps deliver effective care rather than restrict access.

In conclusion, we estimated models linking an individual's preoperative OHS/OKS to the probability that they will achieve a meaningful improvement in symptoms from hip or knee arthroplasty surgery. By our definition of 'meaningful improvement', there is little chance of benefit from surgery when patients report OKS and OHS of 41 and 40 points, respectively, and referral may not be appropriate or necessary. Our work suggests that in current practice very few patients undergo hip or knee arthroplasty above these levels. Introducing minimum thresholds (i.e. creating a hard boundary for referral or not) is not straightforward because any threshold below the absolute value would increase the proportion of patients achieving a meaningful benefit from surgery, but would reduce access for some patients who have the potential to benefit. The models created could be usefully employed to support shared decision-making for individual patients at the time of referral and in secondary care.

Take home message

- The preoperative Oxford Hip or Knee Score can be usefully linked to the likelihood of improving after arthroplasty.
- Patients with an OKS above 41 and OHS above 40 should not be routinely referred for surgery and in UK current practice the vast majority of patients are referred below these levels.
- The models developed should be used to support patients' shared decision-making but thresholds should not be deployed to unfairly restrict patients who would benefit from arthroplasty.

Twitter
Follow H. Dakin @HERC_Oxford

References
1. **No authors listed**. 14th Annual Report, 2017. National Joint Registry for England, Wales, Northern Ireland and the Isle of Man (NJR). https://www.hqip.org.uk/wp-content/uploads/pelerous_media_manager/public/253/NJR/NJR%2014th%20Annual%20Report%202017.pdf (date last accessed 06 February 2020).
2. **Beswick AD**, **Wylde V**, **Gooberman-Hill R**, **Blom A**, **Dieppe P**. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open*. 2012;2(1):e000435.
3. **No authors listed**. Osteoarthritis: care and management: Clinical guideline [CG177]. National Institute for Health and Care Excellence (NICE). 2014. https://www.nice.org.uk/guidance/cg177 (date last accessed 06 February 2020).
4. **Dieppe P**, **Lim K**, **Lohmander S**. Who should have knee joint replacement surgery for osteoarthritis? *Int J Rheum Dis*. 2011;14(2):175–180.
5. **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.
6. **Judge A**, **Arden NK**, **Kiran A**, **et al**. Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery. *J Bone Joint Surg Br*. 2012;94-B(3):412–418.
7. **Kiran A**, **Bottomley N**, **Biant LC**, **et al**. Variations In Good Patient Reported Outcomes After Total Knee Arthroplasty. *J Arthroplasty*. 2015;30(8):1364–1371.
8. **Judge A**, **Arden NK**, **Price A**, **et al**. 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-B(12):1660–1664.
9. **No authors listed**. Provisional Quarterly Patient Reported Outcome Measures (PROMs) in England, April 2016 to March 2017. 2018. https://digital.nhs.uk/data-and-information/publications/statistical/patient-reported-outcome-measures-proms/provisional-quarterly-patient-reported-outcome-measures-proms-in-england-april-2016-to-march-2017-february-2018-release (date last accessed 06 February 2020).
10. **No authors listed**. Musculosketetal. Harrogate and Rural District Clinical Commissioning Group. 2016. https://www.harrogateandruraldistrictccg.nhs.uk/rss2/musculoskeletal/ (date last accessed 06 February 2020).
11. **No authors listed**. Commissioning Policy: Musculoskeletal Surgery Interventions. Redditch Bromsgrove, South Worcestershire, and Wyre Forest Clinical Commissioning Groups. 2011. http://www.worcestershire.nhs.uk/EasySiteWeb/GatewayLink.aspx?alId=30974 (date last accessed 06 February 2020).
12. **No authors listed**. Knee replacement – Referral Criteria/Commissioning position. Scarborough and Ryedale Clinical Commissioning Group. 2016. https://www.scarboroughryedaleccg.nhs.uk/home/referral-information/orthopaedics/knee-replacement/ (date last accessed 06 February 2020).
13. **No authors listed**. Is access to surgery a postcode lottery? The Royal College of Surgeons of England. 2014. https://www.rcseng.ac.uk/-/media/files/rcs/news-and-events/is-access-to-surgery-a-postcode-lottery.pdf (date last accessed 13 May 2020).
14. **Dakin H**, **Gray A**, **Fitzpatrick R**, **et al**. Rationing of total knee replacement: a cost-effectiveness analysis on a large trial data set. *BMJ Open*. 2012;2(1):e000332.
15. **Harris K**, **Dawson J**, **Gibbons E**, **et al**. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101–108.
16. **Bellamy N**, **Buchanan WW**, **Goldsmith CH**, **Campbell J**, **Stitt LW**. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833–1840.
17. **Beard DJ**, **Harris K**, **Dawson J**, **et al**. Meaningful changes for the Oxford hip and knee scores after joint replacement surgery. *J Clin Epidemiol*. 2015;68(1):73–79.
18. **Dawson J**, **Fitzpatrick R**, **Carr A**, **Murray D**, **et al**. Questionnaire on the Perceptions of Patients About Total Hip Replacement.. *J Bone Jt. Surg Br*. 2015;78-B(2):185–190.
19. **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-B(1):63–69.
20. **Harris KK**, **Dawson J**, **Jones LD**, **Beard DJ**, **Price AJ**. Extending the use of PROMs in the NHS-using the Oxford Knee Score in patients undergoing non-operative management for knee osteoarthritis: A validation study. *BMJ Open*. 2013;3(8):e003365.
21. **Herbert A**, **Wijlaars L**, **Zylbersztejn A**, **Cromwell D**, **Hardelid P**. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol*. 2017;46(4):1093–1093i.
22. **Eibich P**, **Dakin HA**, **Price AJ**, **et al**. Associations between preoperative Oxford hip and knee scores and costs and quality of life of patients undergoing primary total joint replacement in the NHS England: an observational study. *BMJ Open*. 2018;8(4):e019477.
23. **Royston P**, **Altman DG**. Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling. *Appl Stat*. 1994;43(3):429.
24. **Koenker R**, **Gilbert B Jr**. Regression Quantiles. *Econometrica*. 1978;46(1):33–50.
25. **Arden N**, **Altman D**, **Beard D**, **et al**. Lower limb arthroplasty: can we produce a tool to predict outcome and failure, and is it cost-effective? An epidemiological study. *Programme Grants for Applied Research*. 2017;5(12):1–246.
26. **Hawker GA**, **Badley EM**, **Borkhoff CM**, **et al**. Which patients are most likely to benefit from total joint arthroplasty? *Arthritis & Rheumatism*. 2013;65(5):1243–1252.
27. **Scott CEH**, **Howie CR**, **MacDonald D**, **Biant LC**. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. *J Bone Joint Surg Br*. 2010;92-B(9):1253–1258.
28. **Dakin H**, **Eibich P**, **Beard D**, **Gray A**, **Price A**. The use of patient-reported outcome measures to guide referral for hip and knee arthroplasty Part 2: A cost-effectiveness analysis. *Bone Joint J*. 2020;102-B(7):950–958.

Author information:
A. J. Price, MA FRCS(Orth), PhD, Professor of Orthopaedic Surgery, Musculoskeletal Science
S. Kang, PhD, Senior Researcher
J. A. Cook, BSc, PhD, CStat, Associate Professor
N. Arden, MBBS, FRCP, MSc, MD, Professor in Rheumatic Diseases
D. J. Beard, GDPhys, MCSP, MSc, MA, DPhil, FRCS(Hon), Professor of Musculoskeletal Science
Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Science, University of Oxford, Oxford, UK.

H. Dakin, DPhil, Senior Researcher, Health Economics Research Centre, University of Oxford, Oxford, UK.

A. Blom, MBChB, FRCS, MD, PhD, FRCS (Tr&Orth), FMedSci, Professor of Orthopaedic Surgery, Musculoskeletal Research Unit, University of Bristol, Bristol, UK.

R. Fitzpatrick, BA, MA, MSc, PhD, F Med Sci, Professor of Public Health and Primary Care, Nuffield Department of Population Health, University of Oxford, Oxford, UK.