The Arthroplasty Candidacy Help Engine tool to select candidates for hip and knee replacement surgery: development and economic modelling

Andrew Price,[1]* James Smith,[1] Helen Dakin,[2] Sujin Kang,[1] Peter Eibich,[2] Jonathan Cook,[1] Alastair Gray,[2] Kristina Harris,[1] Robert Middleton,[1] Elizabeth Gibbons,[3] Elena Benedetto,[1] Stephanie Smith,[1] Jill Dawson,[3] Raymond Fitzpatrick,[3] Adrian Sayers,[4] Laura Miller,[4] Elsa Marques,[4] Rachael Gooberman-Hill,[4] Ashley Blom,[4] Andrew Judge,[1] Nigel Arden,[1] David Murray,[1] Sion Glyn-Jones,[1] Karen Barker,[1] Andrew Carr[1] and David Beard[1]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Health Economics Research Centre, University of Oxford, Oxford, UK
[3]Nuffield Department of Population Health, University of Oxford, Oxford, UK
[4]Musculoskeletal Research Unit, University of Bristol, Bristol, UK

*Corresponding author andrew.price@ndorms.ox.ac.uk

Declared competing interests of authors: Andrew Price reports personal fees from Zimmer Biomet, DePuy Synthes and Smith & Nephew plc, and grants from the National Institute for Health Research (NIHR) and Arthritis Research UK, outside the submitted work. Alastair Gray reports grants from NIHR, during the conduct of the study. Rachael Gooberman-Hill reports grants from the NIHR Health Services and Delivery Research programme for this work during the conduct of the study. Helen Dakin reports grants from NIHR during the conduct of the study and consultancy for Halyard Health outside the submitted work. David Beard reports grants from NIHR outside the submitted work. Jonathan Cook was a member of the NIHR Health Technology Assessment (HTA) Efficient Trial Designs Board (2014–16). Jill Dawson reports grants from the NIHR HTA programme during the conduct of the study and royalty payments from Oxford University Innovation (a university technology transfer company) outside the submitted work, and is one of the original developers of the Oxford Hip and Knee Scores. Raymond Fitzpatrick is one of the developers of the Oxford Hip and Knee Scores. Ashley Blom is the principal investigator in a research project funded by Stryker Corporation. Andrew Judge reports personal fees for consultancy from Anthera Pharmaceuticals, Inc., and Freshfields Bruckhaus Deringer LLP, outside the submitted work. Nigel Arden reports grants from Bioibérica and Novartis Pharmaceuticals UK Ltd, and personal fees from Bioventus, Flexion Therapeutics, Freshfields Bruckhaus Deringer LLP, Janssen Pharmaceutica, Merck & Co. Inc. and Regeneron Pharmaceuticals, Inc., outside the submitted work. David Murray reports grants from the NIHR HTA programme and grants and personal fees from Zimmer Biomet outside the submitted work.

Published June 2019
DOI: 10.3310/hta23320

Scientific summary

The Arthroplasty Candidacy Help Engine tool

Health Technology Assessment 2019; Vol. 23: No. 32
DOI: 10.3310/hta23320

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Background

Numerous health-care measures, including patient-reported outcome measures (PROMs), are used to assess patients undergoing hip and knee replacement. It has been suggested that preoperative PROM scores could be used to guide referrals by general practitioners (GPs) or musculoskeletal hubs to secondary care. Local thresholds have been used in the NHS, although they have been arbitrary, not evidence based and may have resulted in the overtreatment of some patients while inappropriately limiting access to care for others. The aim was to develop a mechanism for identifying appropriate patients for hip and knee replacement referral from primary to secondary care using safe and equitable thresholds. We did this by creating an evidence-based tool, the Arthroplasty Candidacy Help Engine (ACHE), which used an existing assessment score to evaluate and describe patients' capacity to benefit from cost-effective surgery. To achieve this aim, the following questions that were set out in the National Institute for Health Research (NIHR) call for this research were addressed:

- Can clinical tools for assessment of a patient's suitability for knee or hip replacement be used to set thresholds for operation?
- How does the choice of threshold affect the cost-effectiveness of the procedure and subsequent improvements in patient quality of life?

Objectives

- Create a shortlist of scoring systems that are potentially useful for selecting candidates for arthroplasty surgery.
- Identify a single scoring system and threshold values that could to be used within the ACHE tool to select candidates for surgery.
- Establish the cost-effectiveness of hip and knee surgery as the referral threshold changes.
- Explore the potential impact of using the ACHE tool within the NHS.
- Determine the acceptability of the tool and thresholds to stakeholders and patients.

Methods

Work package 1: a systematic review of established scores/instruments used to assess hip and knee replacement

A sensitive filter for finding studies on measurement properties was used to search MEDLINE, EMBASE, PsycINFO, and the Allied and Complementary Medicine Database (AMED). The Patient-Reported Outcome and Quality Of Life Instruments Database (ProQolid), the Oxford PROMs Database, the Database of Abstracts of Reviews of Effects (DARE) and EconLit were also searched using medical subject headings and free-text terms. Titles and abstracts of all identified articles were assessed twice for inclusion/exclusion by two reviewers. Selected full-text articles were then screened for all outcome measures using agreed inclusion and exclusion criteria. From selected publications, data were extracted on the psychometric performance and operational characteristics of each PROM. The following characteristics were included: reliability (test–retest reliability and internal consistency), validity (content and construct validity), responsiveness, interpretability (precision of the measure when used at an individual patient level), evidence of minimal clinically important differences/changes, ceiling or floor effects and acceptability (respondents' willingness to complete). Measurement properties for each instrument were assessed for hip, knee, and mixed hip and knee populations (depending on the availability of published studies). Our initial search yielded 3448 publications, leaving 135 after screening,

from which 32 possible scoring systems were identified. Following data extraction, we identified the Western Ontario and McMaster Universities Arthritis Index (WOMAC®), Oxford Hip Score (OHS), Oxford Knee Score (OKS) and Short Form questionnaire-12 items (SF-12) to be the most promising scores, but all scores required more data to enhance characterisation of measurement properties.

Work package 1: calculation of additional measurement properties

The calculation of additional measurement properties was undertaken using five established pre-existing data sets of patients undergoing primary hip and knee replacement. The Knee Arthroplasty Trial (KAT) and the Exeter Primary Outcome Study (EPOS) data sets were used for the analyses for OHS and OKS, and the SF-12 physical component score (PCS) and mental component score (MCS). The European Collaborative Database of Cost and Practice Patterns of Total Hip Replacement (EUROHIP) data set was used for WOMAC scores. The Assessing Disability After Partial and Total Joint Replacement (ADAPT) study and the Arthroplasty Pain EXperience (APEX) study were used for SF-12 PCS, MCS (ADAPT) and WOMAC scores (both hip and knee) analyses.

The following measurement properties were evaluated:

- internal consistency (Cronbach's alpha and corrected item–total correlation)
- construct validity (the magnitude and direction of correlations with other measures)
- responsiveness (magnitude and direction of Pearson and Spearman correlations of change scores)
- floor and ceiling effects (proportion of the top and the bottom scores at pre and post surgery)
- interpretability [using various definitions of improvement including minimally detectable change (MDC) and group levels of minimally important change (MIC)/minimally important difference criteria].

High internal consistency of the instruments was observed with a Cronbach's alpha of around 0.9 at pre and post operation, and no improvement obtained by removal of any item [except for the preoperation EuroQol-5 Dimensions, three-level version (EQ-5D-3L) index and pre- and post-operation SF-12 (version 1 with US weighting)] score. Construct validity was supported, with strong correlations between the instruments pre and post operation (except the correlation between SF-12 MCS and other instruments: WOMAC total, pain, physical function, stiffness and SF-12 PCS). There was evidence of responsiveness (Spearman's rank-order correlation of > 0.5) in terms of the correlation of the change scores between the instruments, except the correlation between SF-12 MCS and other instruments. High ceiling effects were found in the EQ-5D-3L index (39–46% for hip and 25–30% for knee), OHS (19%) and WOMAC total (21%) for hip only post operation. MDCs [intracluster correlation coefficient (ICC) 0.9] were 0.23–0.24 for the EQ-5D-3L index and 12–16 for WOMAC total score across the data sets. After considering the evidence, four scoring systems were shortlisted and taken forward for further analysis: the OHS [range of scores from minimum = 0 (worst) to maximum = 48 (best)], the OKS [range of scores from minimum = 0 (worst) to maximum = 48 (best)], the SF-12 [range of scores from minimum = 0 (worst) to maximum = 100 (best)] and the WOMAC total [range of scores from maximum = 100 (worst) to minimum = 0 (best)].

Work package 2: calculation of threshold values for shortlisted scores

We estimated absolute and relative thresholds, using different definitions of improvement within the same data sets mentioned above and data from the NHS PROMs collection (2012–15). Preoperative scores were used to calculate absolute thresholds above which there is no potential for clinical benefit from surgery. This is defined as the largest observed presurgery value for which any improvement was achieved. Four improvement definitions included minimally clinically important difference (MCID) applying a 'medium' effect size (ES) (0.5) – criterion B. Linear and logistic regressions were used to estimate two relative thresholds for patient probability of improvement at 50% and 75%. Specificity of using the absolute threshold to rule out inability to benefit was also calculated in each data set.

In reporting the WOMAC score, we inverted the range [inverted range of scores from minimum = 0 (worst) to maximum = 100 (best)] for consistency with the other measures (OKS/OHS/SF-12), giving in all measures a high score, indicating better health status than a low score. The ranges of scores for the following

measures are: OKS [minimum = 0 (worst) to maximum = 48 (best)], OHS [minimum = 0 (worst) to maximum = 48 (best)], SF-12 (PCS and MCS) [minimum = 0 (worst) to maximum = 100 (best)] and the inverted WOMAC [minimum = 0 (worst) to maximum = 100 (best)].

The absolute and relative thresholds for the OHS were 43 (specificity 2–9%) and 38–43 (specificity 2–6%), respectively, based on criterion B. The absolute and relative thresholds for WOMAC in hip arthroplasty were 89–91 (specificity 0–22%) and 78–86 (specificity 20–56%), respectively. SF-12 PCS and MCS findings were similar, with absolute threshold values of 65 for PCS and 66 for MCS (specificity 0% for both) and relative thresholds of 35–47 for PCS (specificity 20–48%) and 37–42 for MCS (specificity 91–100%). Considering knee replacement, the absolute threshold for OKS was 43 (specificity 1%) with relative thresholds of 29–40 (specificity 2–14%). The absolute and relative thresholds for WOMAC total in knee arthroplasty were 90–91 (specificity 0–7%) and 71–86 (specificity 5–19%), respectively. Relative thresholds using different improvement definitions were calculated: thresholds calculated using a medium ES (0.5) MCID showed similar outcomes with a MDC at 90% certainty using an ICC of 0.9. There was substantial variation in the magnitude of absolute change between and within each preoperative score subset. The SF-12 PCS and MCS findings were variable, with absolute threshold values of 66–71 (specificity 0%) and 65–74 (specificity 0–2%) and relative thresholds values of 22–43 (specificity 16–94%) and 26–49 (specificity 72–100%).

Work package 2: health economic evaluation of threshold scores

We conducted a cost–utility analysis comparing total hip arthroplasty (THA) and total knee arthroplasty (TKA) with no arthroplasty from a UK NHS perspective. Six Markov models, each with probabilistic sensitivity analysis (PSA), simulated progression of patient cohorts with different preoperative data to evaluate how the cost-effectiveness of THA/TKA varies with OHS, OKS, WOMAC and SF-12 and with age and sex. Model parameters were initially based on regressions of the parameter of interest on age, sex and preoperative clinical tool score using patient-level data from the APEX study, the Clinical Outcomes in Arthroplasty Study (COASt), EPOS, KAT and web-based PROMs data. Mortality and revision rates were taken from published studies. The reference year for costs was 2014. We took a 10-year time horizon and used a 3.5% discount rate. We considered arthroplasty to be cost-effective if it cost < £20,000 per quality-adjusted life-year (QALY) gained. The results demonstrated that THA/TKA is cost-effective in almost all patients currently undergoing surgery and that economic thresholds could be estimated for OKS and OHS. WOMAC failed to identify any 60- or 70-year-old patients for whom knee replacement was not cost-effective; thresholds for 50- and 80-year-old patients were higher than any scores observed in the available data sets. Hip replacement was cost-effective for all WOMAC scores except for 90-year-old patients scoring 100.

Work package 2: further threshold analysis using the Oxford Hip and Knee Scores

After considering the evidence provided in the initial part of work package 2, our recommendation, guided by the user group (see *Work package 3: user group opinion*), was that the OHS and OKS should be selected to use in the ACHE tool. The decision was based on their measurement properties and the fact that evidence-based thresholds could be calculated. In addition, the scores are already widely used in the NHS patient pathway and this was felt to support future adoption of the ACHE tool. We then undertook more extensive analysis using the NHS PROMs data set linked to Hospital Episode Statistics (HES) (2009–16). The raw improving proportion was calculated and plotted by presurgery score. Improvement was defined as receiver operating characteristic (ROC) curve best cut-off point-based MIC. Furthermore, two modelling approaches were used for analyses of the Oxford Hip and Knee Scores. First, polynomial-based quantile regression models were used to estimate the change score (postoperative minus preoperative) using the presurgery Oxford Hip or Knee Score. Accuracy was assessed against observed percentiles and internal comparison of subsets by key prognostic factors (e.g. gender). The second approach used was the fractional polynomial logistic regressions to predict probability of improving. Using this second modelling approach, the benefit of the baseline covariates on the capacity of benefit was investigated. Internal model validation of the logistic regression models was performed in terms of discrimination and calibration. Sensitivity and specificity values for the estimated relative threshold were calculated with corresponding 95% confidence intervals. The raw probability of improvement was calculated with a pattern similar for both hip and knee patients, although hip patients had a greater chance of improvement given preoperative score. The peak

probability for improvement for both hip and knee replacement occured when the preoperative score was < 20 (approximately 90% of hip patients and 85% of knee patients significantly improving). These values reduced as the preoperative score increased, with 75% of patients obtaining meaningful benefit at scores of 35 for hips and 30 for knees. For a 50% chance of gaining meaningful benefit, the figures are 36 for knee patients and 38 for hip patients. The absolute ROC–MIC-based threshold was 40 for hip replacement and 41 for knee replacement. Quantile regression showed good fit against observed values except at very high and very low preoperative values. Additional covariates did not substantively improve prognostic accuracy with a substantial amount of unexplained variation in patient outcome. A smoothed curve of the raw proportion of meaningful improvement was used in the ACHE tool.

Work package 2: further health economic analysis using the Oxford Hip and Knee Scores

Nine parameters for the OKS and OHS Markov models were re-estimated using PROMs/HES-linked data. The final models using PROMs/HES data found that hip and knee arthroplasty is cost-effective (i.e. costs < £20,000 per QALY) for > 99.9% of patients who currently undergo surgery. Averaging across men and women of all ages, it is cost-effective to conduct THA on patients with an OHS of ≤ 45 [95% credible interval (CrI) 44 to 45] and to conduct TKA on patients with an OKS of ≤ 43 (95% CrI 43 to 44). The economic threshold varied slightly with age but not with gender. PSAs suggested that there was relatively little parameter uncertainty around the conclusions, and sensitivity analyses suggested that the results were robust to large changes in the assumptions.

Work package 3: determining the outcome of using the Arthroplasty Candidacy Help Engine tool in the NHS

We conducted an audit of anonymised data extracted from the medical records of patients who were referred by Oxfordshire GPs with hip ($n = 607$) or knee ($n = 315$) osteoarthritis symptoms to the musculoskeletal hub at the Nuffield Orthopaedic Centre in Oxford between July 2015 and July 2016. These data were combined with PROMs/HES data and the results of the economic evaluation to model the potential impact that the ACHE tool may have on cost and health benefits using different thresholds. This preliminary analysis suggested that using the ACHE tool in a musculoskeletal hub would not reduce but may increase the number of referrals to secondary care. In turn, this may increase costs to the NHS while still supporting cost-effective care.

Work package 3: patient, public and general practitioner survey

We used the probability of good outcome models to develop a prototype ACHE tool, which was web based. We then undertook two web-based surveys in which we demonstrated the use of the ACHE tool to patients/the public and GPs to gain their opinion regarding its use. We had a very low response to the surveys from patients/the public ($n = 22/271$) and GPs ($n = 10/348$). The study data should be considered a pilot analysis, but, encouragingly, those who did respond were broadly supportive of the ACHE tool being used to assist in the decision to refer patients for possible joint replacement surgery.

Work package 3: user group opinion

The user group brought together stakeholders from across the hip and knee pathway in the NHS: patients, members of the public, GPs, surgeons, extended-scope physiotherapists, commissioners, musculoskeletal hub representatives and representatives of the British Orthopaedic Association, the British Hip Society and the British Association for Surgery of the Knee. The user group was consulted four times in the process of producing the ACHE tool, each time for opinions and guidance from users as the work progressed. The process culminated in the final user group meeting in which opinion was gathered as to the ACHE tool's potential real-world use in the NHS. The group's opinion was that the ACHE tool was potentially a very useful tool for assisting and standardising the process of referral from primary to secondary care. There was agreement that the ACHE tool should now be piloted and tested in the NHS to determine its uptake and effect on referral patterns.

Conclusions

The study has shown that the OHS and OKS can be used for assessment of a patient's suitability for knee or hip replacement using thresholds for candidacy based on the individual's capacity for or probability of improving. Our work has shown that hip and knee replacement, when undertaken in any patients with preoperative scores below the absolute OKS and OHS thresholds, is extremely cost-effective. The ACHE tool has been created and should now be carefully tested in the NHS.

Recommendations for future research

Future research could include (1) a real-world study of the ACHE tool to determine its acceptability with patients and GPs and (2) a study of the role of the ACHE tool in supporting referral decisions.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the NIHR.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 11/63/01. The contractual start date was in November 2016. The draft report began editorial review in February 2017 and was accepted for publication in April 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).