Lower limb arthroplasty: can we produce a tool to predict outcome and failure, and is it cost-effective? An epidemiological study

Nigel Arden,[1]* Doug Altman,[1] David Beard,[1] Andrew Carr,[1] Nicholas Clarke,[2] Gary Collins,[1] Cyrus Cooper,[3] David Culliford,[4] Antonella Delmestri,[1] Stefanie Garden,[1] Tinatin Griffin,[1] Kassim Javaid,[1] Andrew Judge,[1] Jeremy Latham,[5] Mark Mullee,[6] David Murray,[1] Emmanuel Ogundimu,[1] Rafael Pinedo-Villanueva,[1] Andrew Price,[1] Daniel Prieto-Alhambra[1] and James Raftery[7]

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Developmental Origins of Health & Disease Division, University of Southampton, Southampton, UK
[3]Medical Research Council, Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK
[4]Faculty of Health Sciences, University of Southampton, Southampton, UK
[5]Orthopaedic and Trauma Surgery, University Hospital Southampton NHS Foundation Trust, Southampton, UK
[6]Research & Development Support Unit, University of Southampton, Southampton, UK
[7]Wessex Institute for Health Research and Development, University of Southampton, Southampton, UK

*Corresponding author nigel.arden@ndorms.ox.ac.uk

Declared competing interests of authors: Nigel Arden reports personal fees from Bioventus, Flexion, Merck Sharp & Dohme, Regeneron, Smith & Nephew, Freshfields Bruckhaus Deringer, and grants from Bioibérica and Novartis outside the submitted work. Daniel Prieto-Alhambra reports grants from Servier Laboratories, AMGEN and Bioibérica S.A., and advisory board fees from AMGEN outside the submitted work. Andrew Judge reports personal fees from Anthera Pharmaceuticals, Inc., Servier, the UK Renal Registry, *Blood Journal* and Freshfields Bruckhaus Deringer, and grants and personal fees from the Oxford Craniofacial Unit, and grants from Roche-Chugai outside the submitted work. Jeremy Latham reports personal fees from Zimmer Biomet, Lima Corporate, MatOrtho and DePuy Synthes outside the submitted work. Rafael Pinedo-Villanueva reports personal fees from Freshfields Bruckhaus Deringer outside the submitted work. David Murray reports grants and personal fees from Zimmer Biomet and grants from Stryker and Health Technology Assessment (HTA) outside the submitted work. James Raftery reports that he is an editor on the National Institute for Health Research HTA and EME editorial board. In addition, James Raftery was previously a Director of the Wessex Institute and Head of the National Institute for Health Research (NIHR) Evaluation, Trials and Studies Coordinating Centre (NETSCC).

Published June 2017
DOI: 10.3310/pgfar05120

Scientific summary

Lower limb arthroplasty

Programme Grants for Applied Research 2017; Vol. 5: No. 12
DOI: 10.3310/pgfar05120

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

Scientific summary

Background

Hip and knee replacements are considered clinically successful and cost-effective. Revision surgery has historically been used to measure the outcome of primary total joint replacements; however, this indicated that up to one-quarter of patients are not satisfied after their operation. Moving towards patient-reported outcome measures (PROMs), therefore, seems warranted.

Several clinical determinants of poor patient outcome have been published, but they have not been used as a combined risk score to guide the clinicians, patients and health commissioners. The results of this programme will help plan future services, educate patients and doctors about benefits and risks of surgery at patient level, and inform the NHS with its planning.

Objectives

1. To describe current and future rates of primary and revision lower limb joint replacement surgery.
2. To describe regional variation in hip and knee replacement surgery in the UK.
3. To confirm the operational, clinical, biological and other important risk factors for poor outcomes.
4. To combine the risk factors to produce a clinically relevant instrument to predict poor outcomes.
5. To perform detailed health economic analyses of the statistical tools developed.
6. To test, and refine, the prediction tool on a pragmatic, prospective lower limb cohort to predict patient-specific outcomes at 12 months postoperatively.

Methods

This report describes the findings and outputs from the 5-year programme of health research into the outcomes of hip and knee arthroplasty. Work packages are displayed in *Figure a*. The programme identifies the current rates of primary and revision arthroplasty, and predicts the future requirements for health-care provision using routinely collected data, for example the Clinical Practice Research Datalink (CPRD), Hospital Episode Statistics (HES), PROMs and the National Joint Registry (NJR). It identifies and combines preoperative characteristics that predict poor patient-reported outcomes and produces a predictive model using a collection of extant UK cohorts. It evaluates the use of this model in a new prospective study of 3000 hip and knee replacement patients recruited in two hospitals in England (Southampton and Oxford). Finally, it evaluates the cost-effectiveness of implementing this model using all of the above sources of data.

Findings

Lower limb arthroplasty: the burden on the health service, present and future

In work package 1 we describe current rates of total hip replacements (THRs) and total knee replacements (TKRs) in the UK and regional variations in hip and knee replacements, and estimate projected future trends in these operations in the UK. We also describe, for the first time, the lifetime risk of THR and TKR.

We used the CPRD, in addition to the NJR, HES and Health Survey for England, to identify all male and female patients who underwent THR or TKR from 1991 to 2006. The rates of both procedures increased substantially over this period but with different trends. The rate of THR increased steadily, whereas the rate of TKR increased slowly initially but then more rapidly after 2000. We estimate the lifetime risk of TKR to

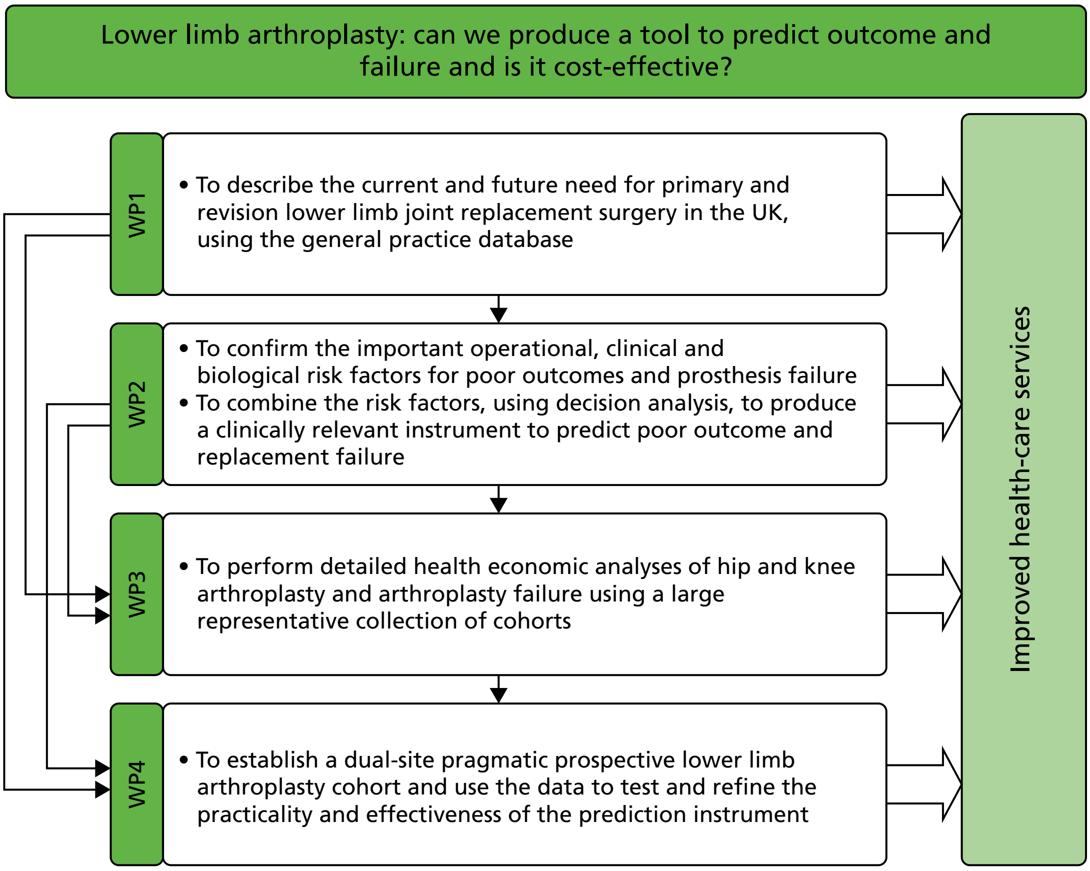

FIGURE a Summary of work programme. WP, work programme.

be 8.1% for 50-year-old men and 10.8% for 50-year-old women, and the lifetime risk of THR to be 7.1% for 50-year-old men and 11.6% for 50-year-old women. The popularity of unicompartmental knee replacement (UKR) has also increased, with the ratio of TKRs to UKRs falling from 250 : 1 in 1999 to 40 : 1 in 2006. We found considerable variation in the rates of THR and TKR between regions included in the General Practice Research Database (GPRD).

We found that the rates of revision surgery are low. The rates of hip revision remained essentially stable for the last 10 years, whereas the rates of knee revision increased substantially. This may partially reflect the recent increase in the rates of primary TKR, but also the established techniques and prostheses for THR. We have demonstrated that increasing body mass index (BMI) is a risk factor for revision surgery of both the hip and the knee. This is an important piece of information, but it has to be considered along with the effect of BMI on PROMs and of complications for which the adverse effects of obesity on outcome are more obvious.

We predict that the number of THRs and TKRs performed will increase dramatically over the next 20 years. The different methodologies used give very different estimates. For THR, we feel that the model using rates fixed at 2010 levels and varying BMI is the most sensible one, and this suggests an annual figure of 95,877 THRs in 2035, a 32% increase in the number of procedures than in 2015. For the knee, we feel that the rates in 2010 do not represent a balance between need and provision, and that the rates will continue to rise. The real number required will therefore be greater than the fixed rates and varying BMI model (118,666), but less than the estimates produced using the log-linear model (1,219,362).

Risk factors and predictive models for poor patient-reported outcome

In work package 2 we describe the predictors of poor patient-reported outcome of THR and TKR at 12 months and combine them into a statistical tool that could be used to identify, prior to lower limb replacement surgeries, patients who are likely to experience a poor outcome.

First, using existing cohorts from the South West London Elective Orthopaedic Centre, with PROMs at 6, 12 and 24 months, we defined values of the Oxford Hip Score (OHS) and Oxford Knee Score (OKS) that were associated with patient satisfaction with the operation at 12 months. We used two different statistical methods to identify the cut-off points: the receiver operating characteristic curve and the 75th percentile approach. The values were 30 units for TKR and 33 units for THR; however, a single score is not recommended, as stratified analyses demonstrated varying values depending on age, sex and baseline score. We propose a new score, percentage of potential change (PoPC), in order to assess the outcome of surgery. PoPC is computed as the actual change divided by the potential improvement multiplied by 100. PoPC is a measure to express relativity of an actual change in PROMs in relation to a potential change, that is, what could have been attained. It demonstrated less heterogeneity when stratifying for important baseline characteristics of the patients.

We then used these scores to identify predictors of patient-reported outcome. We identified a number of important variables that could be used to predict outcomes following THR and TKR. These include preoperative Oxford scores, age, sex, BMI, deprivation index, indication for surgery, anxiety and depression, and radiographic variables. Age had a variable association, with PROMs being worse in both the youngest and oldest patients, whereas the risk of revision surgery was higher in younger patients. Increasing BMI was associated with a higher rate of revision and, although it was associated with a poorer PROM, the effect size was very small, suggesting that it should not be a barrier to surgery. A radiographic pattern of joint space narrowing was found to be a strong predictor of outcome following THR, with outcomes being better in patients in whom joint space loss showed a superolateral pattern than in those with medial, superomedial or concentric patterns of joint space loss.

We have demonstrated for the first time that the use of bisphosphonates reduces the risk of revision knee and hip surgery by 46%. Furthermore, we found that hormone replacement therapy reduced the risk by 38% if used for at least 6 months postoperatively. We have since validated these findings in a Danish registry. In addition, we found an increased risk of postoperative fracture, which is prevented by bisphosphonate use.

We have developed separate predictive models for the hip and knee. For the hip model we used data from two prospective cohorts of patients undergoing primary THR for osteoarthritis (OA), the European Collaborative Database of Cost and Practice Patterns of Total Hip Replacement study and the Exeter Primary Outcomes Study, and for the knee model we used the data from Knee Arthroplasty Trial. We identified risk factors to predict poor outcomes at 12 months after the hip and knee replacements. We used multivariate imputation to combine data from two studies and allow for missing variables. To validate, and to allow for overoptimism of, the model we applied automatic backward selection per 200 bootstrap samples of imputed data sets. The variables retained were those consistently selected for at least 70% of the analyses.

The hip predictive tool included age, sex, baseline OHS, BMI, education, Short Form questionnaire-36 items (SF-36), SF-36 mental component summary score, number of joints with OA, number of joints with surgery, radiographic pattern of OA and two surgical variables (femoral offset size and surgical approach). The model performed well, with a corrected R^2 of 23.1%, and had good calibration, with only slight overestimation of OHS in the lowest decile of outcome.

The knee predictive tool included age, sex, baseline OKS, BMI, deprivation score, Short Form questionnaire-12 items (SF-12), SF-12 mental component summary score, American Society of Anaesthesiologists grade, other conditions affecting mobility, previous knee surgery, fixed flexion deformity, valgus/varus deformity at baseline and preoperative anterior cruciate ligament state (intact yes/no). The model performed less well than the hip model with a corrected R^2 of 20.2%; however, it had good calibration.

Validation of predictive models

Work package 4 evaluated the performance of the models produced above to predict PROMs at 12 months. We recruited a large unique cohort of approximately 3000 patients from two hospitals (in Southampton and Oxford) with very comprehensive phenotyping and biological samples collected at baseline. All patients are followed up annually by means of a postal questionnaire designed to elicit data on PROMs, complications of surgery and health service utilisation.

The cohort confirmed the excellent PROMs of each operation: the THR preoperative OHS was 18.63 units [standard deviation (SD) 8.05 units] and the postoperative OHS was 41.06 units (SD 8.96 units), 92% satisfied; and the TKR preoperative OKS was 20.31 units (SD 7.69 units) and the postoperative OHS was 37.46 units (SD 9.74 units), 87% satisfied. It provided essential data on health resource use, both pre- and postoperatively. The performance of the knee model was modest ($R^2 = 0.14$) and that of the hip model poor ($R^2 = 0.04$). However, when the same variables used to develop the original model from the Knee Arthroplasty Trial (KAT) were used to produce a new model using Clinical Outcomes in Arthroplasty Study (COASt) data, the performance was improved for the knee model ($R^2 = 0.216$) as opposed to the levels of performance as the development model ($R^2 = 0.202$). Both models performed better in predicting good rather than poor outcomes. The addition of radiographic OA severity improved the performance of the hip model ($R^2 = 0.125$ vs. 0.110) and high-sensitivity C-reactive protein improved the performance of the knee model ($R^2 = 0.230$ vs. 0.216), demonstrating the importance of expanding the number of patient-based predictors, a large number of which have been collected in this cohort.

Several factors that affect the performance are discussed in detail, but the degree of imputation required in the development cohorts, the different case mix and non-comparability of the variables collected in the development and validation cohort are the most important.

Cost-effectiveness of predictive models

Work package 3 described the costs and utility outcomes of THR and TKR and assessed the cost utility of implementing the predictive model in clinical practice. We produced a novel Markov model that started at the orthopaedic surgeon's assessment and distinguished between outcome categories following primary and revision procedures. It was populated with the best-quality patient-level data available from numerous sources (such as GPRD, PROMs and HES), in addition to the cohorts involved in this project. We mapped the OHS onto the EuroQol-5 Dimensions (EQ-5D) index to enable production of health utilities estimates from the Oxford scores.

At 12 months post surgery, mean EQ-5D scores were substantially higher among both patients who underwent hip replacement and those who underwent knee replacement (0.44 and 0.32 units, respectively). Even patients who were defined as having poor outcomes in this programme experienced a substantial improvement in score (0.28 units and 0.19 units, respectively). As the cost of surgery is £4000–6000, the operations are cost-effective interventions, setting a high hurdle for the predictive tool to be cost-effective. The developed outcome prediction tool for THRs and TKRs did reduce the number of unsatisfactory and poor outcomes; however, the tool would deny surgery to patients who would have improved significantly, thereby producing fewer quality-adjusted life-years (QALYs) than current practice.

The highest savings per QALY forgone for THRs were in the oldest patients (aged ≥ 80 years), with an incremental cost-effectiveness ratio (ICER) of around £1200 saved per QALY forgone. In the case of TKRs, the highest ICER was reported by younger women (entering the model at 45 years of age), with £637 saved per each QALY forgone. Keeping patients from surgery, therefore, appears unlikely to be cost-effective for any tool applied to such a highly successful operation, unless the tool is extremely sensitive and specific.

The Markov model produced will now be extremely useful for assessing the impact of current strategies aimed at restricting access to lower limb arthroplasty, including those based on BMI thresholds. It will also be essential for modelling new interventions produced.

Conclusions

This programme has calculated the number of hip and knee replacement operations performed in the past and projected to be performed in the future, which will help in planning services. It has defined a poor outcome using PROMs and identified a number of important predictors of PROMs. Increased BMI is statistically significantly associated with poorer PROMs, but the effect size is small and almost certainly not clinically significant. It is, however, associated with an increased risk of revision surgery and postoperative complications, something that needs to be considered when making a decision to operate. We found that bisphosphonates reduce postoperative fractures and the need for revision surgery, and subsequently validated this finding.

Although we have produced a predictive tool for outcome, it cannot be cost-effectively implemented in its current form. However, we have demonstrated that the addition of extra variables to the previously described list of predictors does improve the performance of the predictive tools. Further work is being performed to refine and improve the predictive tools using more extensive and novel risk factors. We believe that the work will prove to be very useful as part of patient decision aids in the future.

The Markov model and the collection of cohorts produced in this programme will prove beneficial for assessment of any future therapeutic or health-care delivery interventions.

Several areas of future research would use this programme as a foundation. This could involve:

- developing and testing a potential postoperative prediction model
- exploring novel potential predictors such as bone mineral density, vitamin D, bone markers and better phenotyping of mental status
- producing a new bespoke model in the validation cohort with external validation in a contemporary Geneva Arthroplasty Registry
- obtaining annual follow-up scores from the validation cohort for a longer period with a view to having long-term PROMs and also revision rates
- exploring the economic effects of new therapeutic and care delivery interventions (e.g. BMI restriction criteria) using the Markov model.

Funding

Funding for this study was provided by the Programme Grants for Applied Research programme of the National Institute for Health Research.

Programme Grants for Applied Research

ISSN 2050-4322 (Print)

ISSN 2050-4330 (Online)

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 PGfAR archive is freely available to view online at www.journalslibrary.nihr.ac.uk/pgfar. 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 *Programme Grants for Applied Research* journal

Reports are published in *Programme Grants for Applied Research* (PGfAR) if (1) they have resulted from work for the PGfAR programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Programme Grants for Applied Research programme

The Programme Grants for Applied Research (PGfAR) programme, part of the National Institute for Health Research (NIHR), was set up in 2006 to produce independent research findings that will have practical application for the benefit of patients and the NHS in the relatively near future. The Programme is managed by the NIHR Central Commissioning Facility (CCF) with strategic input from the Programme Director.

The programme is a national response mode funding scheme that aims to provide evidence to improve health outcomes in England through promotion of health, prevention of ill health, and optimal disease management (including safety and quality), with particular emphasis on conditions causing significant disease burden.

For more information about the PGfAR programme please visit the website: http://www.nihr.ac.uk/funding/programme-grants-for-applied-research.htm

This report

The research reported in this issue of the journal was funded by PGfAR as project number RP-PG-0407-10064. The contractual start date was in September 2008. The final report began editorial review in August 2015 and was accepted for publication in March 2016. As the funder, the PGfAR programme agreed the research questions and study designs in advance with the investigators. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The PGfAR editors and production house have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the final report 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, CCF, NETSCC, PGfAR or the Department of Health. 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 PGfAR programme or the Department of Health.

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