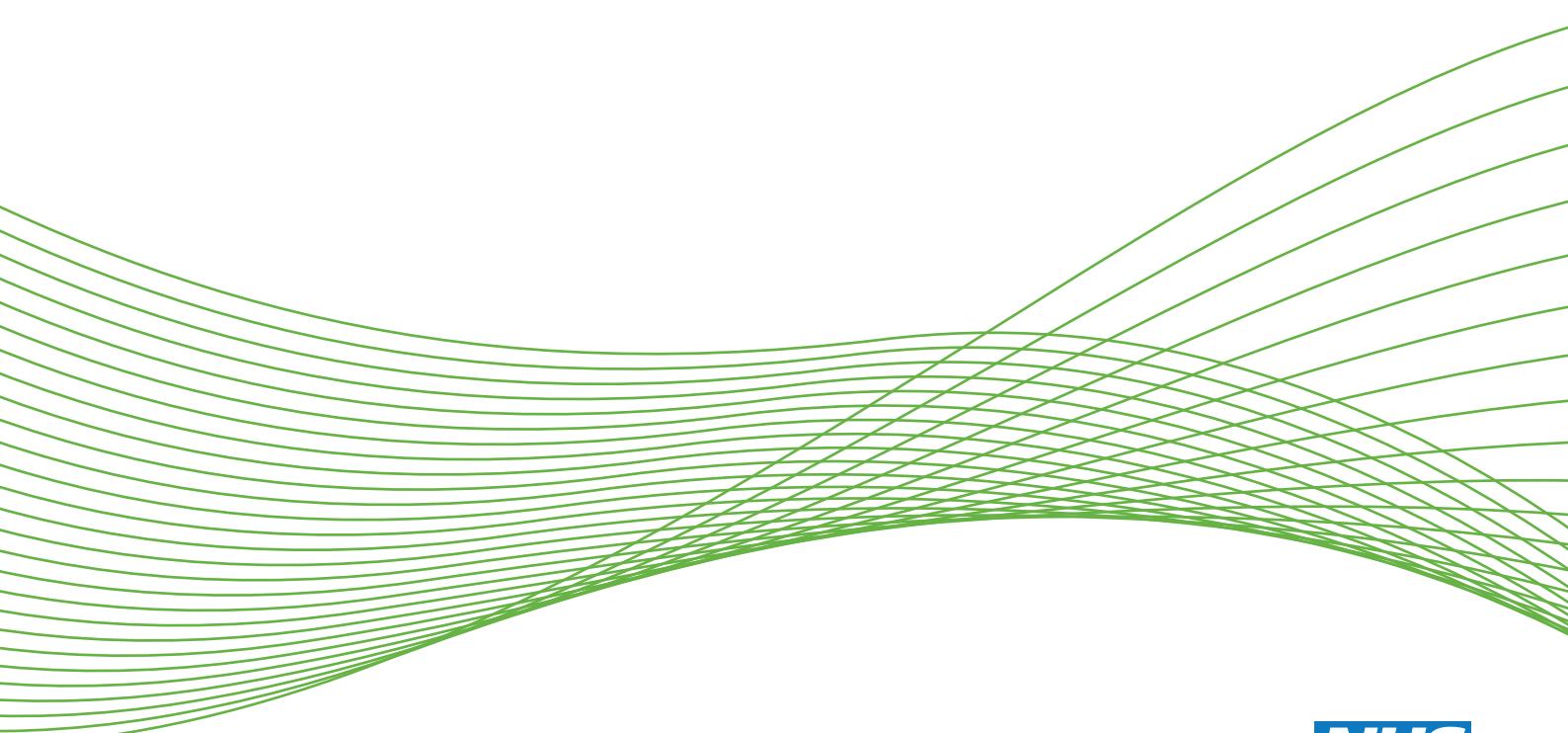


Lower limb arthroplasty: can we produce a tool to predict outcome and failure, and is it cost-effective? An epidemiological study

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



**National Institute for
Health Research**

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-Alhambra¹ and James Raftery⁷

¹Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

²Developmental Origins of Health & Disease Division, University of Southampton, Southampton, UK

³Medical Research Council, Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK

⁴Faculty of Health Sciences, University of Southampton, Southampton, UK

⁵Orthopaedic and Trauma Surgery, University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁶Research & Development Support Unit, University of Southampton, Southampton, UK

⁷Wessex Institute for Health Research and Development, University of Southampton, Southampton, UK

*Corresponding author

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

This report should be referenced as follows:

Arden N, Altman D, Beard D, Carr A, Clarke N, Collins G, *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 Appl Res* 2017;**5**(12).

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.

© Queen's Printer and Controller of HMSO 2017. This work was produced by Arden *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health. This issue may be freely reproduced for the purposes of private research and study and extracts (or indeed, the full report) may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NIHR Journals Library, National Institute for Health Research, Evaluation, Trials and Studies Coordinating Centre, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

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

Programme Grants for Applied Research Editor-in-Chief

Professor Paul Little Professor of Primary Care Research, University of Southampton, UK

NIHR Journals Library Editor-in-Chief

Professor Tom Walley Director, NIHR Evaluation, Trials and Studies and Director of the EME Programme, UK

NIHR Journals Library Editors

Professor Ken Stein Chair of HTA Editorial Board and Professor of Public Health, University of Exeter Medical School, UK

Professor Andree Le May Chair of NIHR Journals Library Editorial Group (EME, HS&DR, PGfAR, PHR journals)

Dr Martin Ashton-Key Consultant in Public Health Medicine/Consultant Advisor, NETSCC, UK

Professor Matthias Beck Chair in Public Sector Management and Subject Leader (Management Group), Queen's University Management School, Queen's University Belfast, UK

Dr Tessa Crilly Director, Crystal Blue Consulting Ltd, UK

Dr Eugenia Cronin Senior Scientific Advisor, Wessex Institute, UK

Ms Tara Lamont Scientific Advisor, NETSCC, UK

Dr Catriona McDaid Senior Research Fellow, York Trials Unit, Department of Health Sciences, University of York, UK

Professor William McGuire Professor of Child Health, Hull York Medical School, University of York, UK

Professor Geoffrey Meads Professor of Health Sciences Research, Health and Wellbeing Research Group, University of Winchester, UK

Professor John Norrie Chair in Medical Statistics, University of Edinburgh, UK

Professor John Powell Consultant Clinical Adviser, National Institute for Health and Care Excellence (NICE), UK

Professor James Raftery Professor of Health Technology Assessment, Wessex Institute, Faculty of Medicine, University of Southampton, UK

Dr Rob Riemsma Reviews Manager, Kleijnen Systematic Reviews Ltd, UK

Professor Helen Roberts Professor of Child Health Research, UCL Institute of Child Health, UK

Professor Jonathan Ross Professor of Sexual Health and HIV, University Hospital Birmingham, UK

Professor Helen Snooks Professor of Health Services Research, Institute of Life Science, College of Medicine, Swansea University, UK

Professor Jim Thornton Professor of Obstetrics and Gynaecology, Faculty of Medicine and Health Sciences, University of Nottingham, UK

Professor Martin Underwood Director, Warwick Clinical Trials Unit, Warwick Medical School, University of Warwick, UK

Please visit the website for a list of members of the NIHR Journals Library Board:
www.journalslibrary.nihr.ac.uk/about/editors

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

Abstract

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-Alhambra¹ and James Raftery⁷

¹Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

²Developmental Origins of Health & Disease Division, University of Southampton, Southampton, UK

³Medical Research Council, Lifecourse Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton, UK

⁴Faculty of Health Sciences, University of Southampton, Southampton, UK

⁵Orthopaedic and Trauma Surgery, University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁶Research & Development Support Unit, University of Southampton, Southampton, UK

⁷Wessex Institute for Health Research and Development, University of Southampton, Southampton, UK

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

Background: Although hip and knee arthroplasties are considered to be common elective cost-effective operations, up to one-quarter of patients are not satisfied with the operation. A number of risk factors for implant failure are known, but little is known about the predictors of patient-reported outcomes.

Objectives: (1) Describe current and future needs for lower limb arthroplasties in the UK; (2) describe important risk factors for poor surgery outcomes and combine them to produce predictive tools (for hip and knee separately) for poor outcomes; (3) produce a Markov model to enable a detailed health economic analysis of hip/knee arthroplasty, and for implementing the predictive tool; and (4) test the practicality of the prediction tools in a pragmatic prospective cohort of lower limb arthroplasty.

Design: The programme was arranged into four work packages. The first three work packages used the data from large existing data sets such as Clinical Practice Research Datalink, Hospital Episode Statistics and the National Joint Registry. Work package 4 established a pragmatic cohort of lower limb arthroplasty to test the practicality of the predictive tools developed within the programme.

Results: The estimated number of total knee replacements (TKRs) and total hip replacements (THRs) performed in the UK in 2015 was 85,019 and 72,418, respectively. Between 1991 and 2006, the estimated age-standardised rates (per 100,000 person-years) for a THR increased from 60.3 to 144.6 for women and from 35.8 to 88.6 for men. The rates for TKR increased from 42.5 to 138.7 for women and from 28.7

to 99.4 for men. The strongest predictors for poor outcomes were preoperative pain/function scores, deprivation, age, mental health score and radiographic variable pattern of joint space narrowing. We found a weak association between body mass index (BMI) and outcomes; however, increased BMI did increase the risk of revision surgery (a 5-kg/m² rise in BMI increased THR revision risk by 10.4% and TKR revision risk by 7.7%). We also confirmed that osteoarthritis (OA) severity and migration pattern of the hip predicted patient-reported outcome measures. The hip predictive tool that we developed performed well, with a corrected R^2 of 23.1% and had good calibration, with only slight overestimation of Oxford Hip Score in the lowest decile of outcome. The knee tool developed performed less well, with a corrected R^2 of 20.2%; however, it had good calibration. The analysis was restricted by the relatively limited number of variables available in the extant data sets, something that could be addressed in future studies. We found that the use of bisphosphonates reduced the risk of revision knee and hip surgery by 46%. Hormone replacement therapy reduced the risk by 38%, if used for at least 6 months postoperatively. We found that an increased risk of postoperative fracture was prevented by bisphosphonate use. This result, being observational in nature, will require confirmation in a randomised controlled trial. The Markov model distinguished between outcome categories following primary and revision procedures. The resulting outcome prediction tool for THR and TKR reduced the number and proportion of unsatisfactory outcomes after the operation, saving NHS resources in the process. The highest savings per quality-adjusted life-year (QALY) forgone were reported from the oldest patient subgroups (men and women aged ≥ 80 years), with a reported incremental cost-effectiveness ratio of around £1200 saved per QALY forgone for THRs. In the prospective cohort of arthroplasty, the performance of the knee model was modest ($R^2 = 0.14$) and that of the hip model poor ($R^2 = 0.04$). However, the addition of the radiographic OA variable 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). These data will ideally need replication in an external cohort of a similar design. The data are not necessarily applicable to other health systems or countries.

Conclusion: The number of total hip and knee replacements will increase in the next decade. High BMI, although clinically insignificant, is associated with an increased risk of revision surgery and postoperative complications. Preoperative pain/function, the pattern of joint space narrowing, deprivation index and level of education were found to be the strongest predictors for THR. Bisphosphonates and hormone therapy proved to be beneficial for patients undergoing lower limb replacement. The addition of new predictors collected from the prospective cohort of arthroplasty slightly improved the performance of the predictive tools, suggesting that the potential improvements in both tools can be achieved using the plethora of extra variables from the validation cohort. Although currently it would not be cost-effective to implement the predictive tools in a health-care setting, we feel that the addition of extensive risk factors will improve the performances of the predictive tools as well as the Markov model, and will prove to be beneficial in terms of cost-effectiveness. Future analyses are under way and awaiting more promising provisional results.

Future work: Further research should focus on defining and predicting the most important outcome to the patient.

Funding: The National Institute for Health Research Programme Grants for Applied Research programme.

Contents

List of tables	xiii
List of figures	xxi
List of abbreviations	xxv
Plain English summary	xxvii
Scientific summary	xxix
Chapter 1 Introduction	1
Background	1
Surgical trends	1
Predictors of surgery outcomes and prediction tool	2
Cost-effectiveness of implementation of the tools	4
External validation of the tool	5
Aims of the programme grant	5
Chapter 2 Work package 1: current and future rates of lower limb arthroplasties	7
Design and setting	7
Main exposure	8
Outcome measures	8
Rates of total hip replacement, total knee replacement and unicompartamental knee replacement in the UK	8
<i>Temporal trends</i>	8
Regional and national variation for hip and knee replacement surgery in the UK	11
Describing the revision rates for hip and knee arthroplasty, and quantifying the rates in the UK	12
Association of body mass index with time of revision	18
<i>A population-based survival analysis describing the association of body mass index on time to revision for total hip and knee replacements: results from the UK General Practice Research Database</i>	18
Projected future trends for total hip replacement/total knee replacement accounting for projected changes in age and obesity	21
<i>Estimating the lifetime risk of total knee and hip arthroplasties</i>	21
<i>Future projections of total hip and knee arthroplasties</i>	23
Discussion	26
Conclusion	28
Chapter 3 Work package 2: designing the statistical tool to predict surgery outcome	29
Predictors of outcome	29
Aims and objectives	29
Design and setting	30
<i>Cohorts and databases</i>	30
Outcome measures	31
<i>Oxford Hip Score/Oxford Knee Score</i>	31
<i>EuroQol-5 Dimensions</i>	31
<i>Satisfaction</i>	32

<i>Western Ontario and McMaster Universities Osteoarthritis Index</i>	32
<i>Short Form questionnaire-12 items</i>	32
Main exposure	32
Sample size	32
Variables	32
Statistical methods	33
Defining good and bad patient-reported outcome measures	33
<i>What is a good patient-reported outcome after total hip replacement?</i>	33
<i>Value-added publication</i>	39
Identifying the role of univariable as well as multivariable risk factors in patient-reported outcomes	41
<i>Preoperative Oxford Hip Score and Oxford Knee Score</i>	41
<i>Age</i>	46
<i>Sex</i>	48
<i>Body mass index</i>	50
<i>Deprivation</i>	56
<i>Indication for surgery</i>	57
<i>American Society of Anesthesiologists/comorbidities</i>	59
<i>Anxiety/depression</i>	60
<i>Clinical, surgical and drug predictors</i>	63
Conclusion	64
Developing a predictive model	65
<i>Statistical tool for predicting poor outcomes following total hip or knee replacement surgeries</i>	65
<i>Hip predictive tool</i>	66
<i>Knee predictive tool</i>	74
Chapter 4 Work package 3: economic evaluation	81
Introduction	81
Methods for total hip replacement	82
<i>Economic model</i>	82
<i>Patient subgroups</i>	83
Model inputs for total hip replacements	84
<i>Transition probabilities for total hip replacement</i>	84
<i>Costs for total hip replacements</i>	88
<i>Quality-adjusted life-years for total hip replacements</i>	94
<i>Parameter values when using the prediction tool</i>	99
<i>Assumptions</i>	101
Methods for total knee replacements	103
Economic model for total knee replacements	103
Model inputs	103
<i>Transition probabilities for total knee replacements</i>	103
<i>Costs for total knee replacements</i>	106
<i>Quality-adjusted life-years for total knee replacements</i>	108
<i>Parameter values when using the tool</i>	110
Results	112
<i>Total hip replacements</i>	112
<i>Total knee replacements</i>	116
Discussion	119
<i>Main findings</i>	119

Chapter 5 Work package 4: external validation of the tool	125
Aim	125
Background	125
Description of the study	126
<i>Study design</i>	126
<i>Important changes to the design of the study</i>	126
<i>Participating centres</i>	126
<i>Clinical Outcomes in Arthroplasty Study team</i>	127
Study population	127
<i>Patient inclusion and exclusion criteria</i>	127
Patient pathway	128
<i>Consent and patient recruitment</i>	128
<i>Preoperative visits</i>	130
<i>Inpatient data and sample collection</i>	130
<i>Postoperative visits</i>	130
<i>Follow-up</i>	130
<i>Patient pathways at University Hospitals Southampton NHS Foundation Trust and Nuffield Orthopaedic Centre</i>	130
Ethics	133
<i>Recruitment figures</i>	133
Data, samples and radiographic variables	136
<i>Data</i>	136
<i>Samples</i>	136
<i>Radiographic variables</i>	137
Data storage and database	138
<i>Database</i>	138
<i>Data security and confidentiality</i>	138
<i>Error checking and validation</i>	138
<i>Audit trail</i>	138
<i>Backups</i>	138
<i>Safety reporting</i>	138
Patients and scientific community	139
<i>Data sharing</i>	139
<i>Patient and public involvement</i>	139
Statistical analysis (methods) and end points	140
<i>Total knee replacement: Clinical Outcomes in Arthroplasty Study knee model</i>	140
<i>Total hip replacement: Clinical Outcomes in Arthroplasty Study hip model</i>	140
<i>Assessment of Clinical Outcomes in Arthroplasty Study hip radiographs</i>	141
Results	142
<i>Part 1: Clinical Outcomes in Arthroplasty Study hip model</i>	142
<i>Part 2: Clinical Outcomes in Arthroplasty Study knee model</i>	153
Absolute differences between baseline and the 12-month follow-up Oxford scores	163
Patient expectation and outcome	163
Discussion	164
Conclusion	166
 Chapter 6 Summary of the programme	 167
Work package 1	167
Work package 2	168
Work package 3	170
Work package 4	171
Research implications	172
Conclusions	172

Acknowledgements	175
References	179
Appendix 1 Read codes	197
Appendix 2 Statistical methods	203
Appendix 3 Variable list	207
Appendix 4 Hip replacement model by Briggs <i>et al.</i>	213
Appendix 5 Full total hip replacement economic model schema	215
Appendix 6 Expert elicitation	217
Appendix 7 Estimating 2-year revision rates by Kalairajah outcome classification using data from the New Zealand Joint Registry	219
Appendix 8 Estimating primary care costs attributable to hip pain	221
Appendix 9 Logit model to predict surgery outcome	223
Appendix 10 Mapping the Oxford Hip Score onto the EuroQol-5 Dimensions	225
Appendix 11 Model assumptions and limitations	233
Appendix 12 Primary care costs for total knee replacements	239
Appendix 13 External validation of the model	241

List of tables

TABLE 1 Baseline clinical variables for THR and TKR	16
TABLE 2 Clinical and demographic characteristics, all participants undergoing hip and knee replacement	18
TABLE 3 Estimated subhazard of revision for total hip and knee replacements: competing risks analysis – hip	19
TABLE 4 Estimated subhazard of revision for total hip and knee replacements: competing risks analysis – knee	20
TABLE 5 Estimated hazard of revision for THR: univariable and adjusted Cox regression analysis with death as a censoring event	20
TABLE 6 Estimated hazard of revision for TKR: univariable and adjusted Cox regression analysis with death as a censoring event	21
TABLE 7 Estimated lifetime risk (95% CI) of undergoing primary TKR or THR based on age- and sex-specific incidence rates adjusted for mortality	22
TABLE 8 Demographic characteristics of CPRD subjects used to construct incidence rates	25
TABLE 9 Health Survey for England 1991 to 2010 showing number of female and male subjects by BMI	25
TABLE 10 Health Survey for England 1991 to 2010 showing number of female and male subjects by age	25
TABLE 11 Projected UK counts for total hip and knee replacements in adults to the year 2035	26
TABLE 12 Projected UK counts for total hip and knee replacements respectively by sex to the year 2035, with estimated TKR incidence rates fixed at 2010 level	27
TABLE 13 Cohorts and data sets used in work package 2	31
TABLE 14 Baseline characteristics of patients	34
TABLE 15 Descriptive statistics of satisfaction with surgery: OKS and OHS preoperatively and at 6 months	37
TABLE 16 Spearman's rank-order correlation of satisfaction with surgery (95% CI)	39
TABLE 17 Patient satisfaction with surgery 6 months after THR, stratified by baseline scores	40
TABLE 18 Patient satisfaction with surgery 6 months after TKR, stratified by baseline scores	41

TABLE 19 Associations found between pre- and postoperative score	45
TABLE 20 Associations found between age and postoperative score	49
TABLE 21 Associations found between sex and postoperative score	50
TABLE 22 Associations found between BMI and postoperative knee score	52
TABLE 23 Associations found between BMI and postoperative hip score	53
TABLE 24 Repeated measures regression model: estimates of pre- and postoperative OHS adjusted for age and sex	55
TABLE 25 Repeated measures regression model: estimates of pre- and postoperative OHS adjusted for adjusted for all confounders	55
TABLE 26 Associations found between deprivation and postoperative score	57
TABLE 27 Associations found between indication for OA or RA and postoperative score	58
TABLE 28 Associations found between ASA score and number of extra comorbidities and postoperative score	60
TABLE 29 Associations found between mental health and postoperative score	62
TABLE 30 Associations found between surgical and drug variables and postoperative score	64
TABLE 31 List of predictors available within the EUROHIP and EPOS cohorts, and distribution of the extent of missing data in each study	67
TABLE 32 Clinical baseline data between responders and non-responders	70
TABLE 33 Results of linear regression (ANCOVA) model to identify predictors of OHS at 12 months' follow-up	71
TABLE 34 Results of calibration and discrimination of the predictive model	72
TABLE 35 List of prognostic variables available for analysis in KAT trial	75
TABLE 36 Descriptive statistics and comparison of potential predictors between responders and non-responders	76
TABLE 37 Identification of the 12-month postoperative OKS for the total sample (general linear model)	78
TABLE 38 Preoperative probabilities: deterministic and probabilistic parameters	85
TABLE 39 Probability of poor outcome after primary THR: deterministic and probabilistic parameters	85

TABLE 40 Transition probabilities between outcome categories after primary THR from the first to second postoperative years: deterministic and probabilistic parameters	86
TABLE 41 Transition probabilities between outcome categories from second year after a primary THR onwards: deterministic and probabilistic parameters	87
TABLE 42 Transition probabilities between outcome categories and revision THR: deterministic and probabilistic parameters	87
TABLE 43 Probability of poor outcome after revision THR: deterministic and probabilistic parameters	88
TABLE 44 Transition probabilities between outcome categories after revision THR from first to second postoperative years: deterministic and probabilistic parameters	88
TABLE 45 Transition probabilities between outcome categories after second year following a revision THR onwards: deterministic and probabilistic parameters	89
TABLE 46 Primary care costs associated to all preoperative states: deterministic and probabilistic parameters	90
TABLE 47 Cost of a primary THR to the NHS by patient subgroup	90
TABLE 48 Costs associated to primary THR and first year thereafter by outcome category: deterministic and probabilistic parameters	91
TABLE 49 Costs associated to second and subsequent years after THR in either outcome category states: deterministic and probabilistic parameters	92
TABLE 50 Cost of a revision THR to the NHS by patient subgroup	93
TABLE 51 Costs associated to revision THR followed by first year in either outcome category states: deterministic and probabilistic parameters	93
TABLE 52 Costs associated to second and subsequent years after THR in either outcome category states: deterministic and probabilistic parameters	94
TABLE 53 Disutility associated to all preoperative states: deterministic and probabilistic parameters	95
TABLE 54 Disutility associated to primary THR and first postoperative year states: deterministic and probabilistic parameters	97
TABLE 55 Disutility associated to second and subsequent years after primary THR: deterministic and probabilistic parameters	98
TABLE 56 The HES PROMs data: pre- and postoperative health utility estimates by patient subgroup and outcome category	98
TABLE 57 Disutility associated to revision THR and first postoperative year states: deterministic and probabilistic parameters	99

TABLE 58 Disutility associated to second and subsequent years after revision THR: deterministic and probabilistic parameters	100
TABLE 59 Preoperative probabilities with the tool: deterministic and probabilistic parameters	100
TABLE 60 Probability of poor outcome after primary THR with the tool: deterministic and probabilistic parameters	101
TABLE 61 Disutility associated to preoperative states with the tool: deterministic and probabilistic parameters	102
TABLE 62 Probability of each outcome group following surgery, by patient subgroup	105
TABLE 63 Revision rates by outcome groups, first year after TKR	106
TABLE 64 Revision rates by outcome groups, ≥ 2 years after TKR	106
TABLE 65 Excess primary care costs of TKR patients compared with controls (estimates from CPRD data)	107
TABLE 66 Mean primary care costs during first year after primary TKR	109
TABLE 67 Effect of implementing the outcome prediction tool for TKR on COASt cohort	111
TABLE 68 Probability of 'poor' outcome after primary TKR, based on application of the outcome prediction tool	111
TABLE 69 Probability of 'poor' outcome after primary THR, based on application of the outcome prediction tool on the COASt cohort	111
TABLE 70 Deterministic cost-effectiveness results, by patient subgroup	112
TABLE 71 Sensitivity analysis on health utility estimate for long-term medical management, women aged 70 years	113
TABLE 72 Sensitivity analysis on the cut-off point applied by the outcome prediction tool, women aged 70 years	114
TABLE 73 Deterministic cost-effectiveness results for TKRs, by patient subgroup	116
TABLE 74 Sensitivity analysis on health utility estimate for long-term medical management, women aged 70 years	117
TABLE 75 Sensitivity analysis on the cut-off point applied by the outcome prediction tool, women aged 70 years	118
TABLE 76 Showing patient inclusion and exclusion criteria for COASt	128
TABLE 77 Detailing patient eligibility criteria by procedures for knee and hip category	129

TABLE 78 Detailing procedures performed in COAST cohort for knee and hip category	129
TABLE 79 Recruitment and figures and returned 1-year follow-up questionnaire for hip patients in COAST	134
TABLE 80 Recruitment figures and returned 1-year follow-up questionnaire for knee patients in COAST	135
TABLE 81 Sample types and derived aliquots collected by COAST	137
TABLE 82 Numerical description of the COAST variables and outcome used to validate the hip model	143
TABLE 83 Numerical description of the additional variables used that were not included in the COAST hip model (complete and incomplete records)	144
TABLE 84 Calibration and discrimination of the validated hip model using the complete data set	148
TABLE 85 Calibration and discrimination of the validated hip model using the imputed data: validated COAST data	150
TABLE 86 Corrected and uncorrected R^2 for the variables added to the COAST hip model	151
TABLE 87 Calibration and discrimination of the improved model using the hip COAST data	152
TABLE 88 Linear regression models for the OHS COAST data with one additional variable	153
TABLE 89 Numerical description of the COAST variables used to validate the knee model (incomplete data set, 608 records and the entire data set, 1025 records)	154
TABLE 90 Numerical description of the additional variables used that were not included in the COAST knee model (incomplete data sets, 608 and 1025 records)	155
TABLE 91 Calibration and discrimination of the validated knee model using the 608-record data set with imputed data: validated COAST data	159
TABLE 92 Calibration and discrimination of the validated knee model using the 1025-record data set with imputed data	160
TABLE 93 Corrected and uncorrected R^2 on adding the C-reactive protein variable to the COAST knee model	162
TABLE 94 Calibration and discrimination of the improved model using the knee COAST data	162
TABLE 95 Absolute difference between baseline and the 12-month scores for the OHS and OKS	163

TABLE 96 Patient expectation and outcome	164
TABLE 97 Medical diagnosis codes for hip osteoarthritis	197
TABLE 98 Medical diagnosis codes for knee osteoarthritis	197
TABLE 99 Medical diagnosis codes for hip replacement	197
TABLE 100 Medical diagnosis codes for knee replacement	199
TABLE 101 Medical diagnosis codes for hip pain	200
TABLE 102 Medical diagnosis codes for knee pain	200
TABLE 103 Medical diagnosis codes for hip arthritis	201
TABLE 104 Medical diagnosis codes for knee arthritis	201
TABLE 105 Medical diagnosis codes for degeneration of the knee	201
TABLE 106 Variable list of THR in different cohorts	207
TABLE 107 Expert opinion on patient referral from surgical assessment	217
TABLE 108 Expert opinion on reassessment after risk factor modification	217
TABLE 109 Expert opinion on reassessment after long-term medical management	218
TABLE 110 Number and rates of revision at 2 years based on New Zealand Joint Registry data	219
TABLE 111 Use of resources by outcome category, COASt cohort patients	223
TABLE 112 Surgery outcome predictive model at 1 year: logit regression for poor outcome, estimated on COASt cohort data	224
TABLE 113 Linear regression output from OLS continuous model	227
TABLE 114 Coefficients and <i>p</i> -values for OLS categorical and two-part models	228
TABLE 115 Performance of models and internal validation	229
TABLE 116 Performance of models and external validation	230
TABLE 117 Mean primary care costs during first year after revision TKR	239
TABLE 118 Mean primary care costs during year 2 and onwards after primary TKR	239
TABLE 119 Mean primary care costs during year 2 and onwards after revision TKR	239
TABLE 120 Linear regression models for the OKS COASt data with one additional variable	241

TABLE 121 Additional variables not included in the COAST hip model, based on the complete (550 records) and incomplete (1098 records) data sets	242
TABLE 122 Additional variables not included in the COAST hip model, based on the complete (550 records) and incomplete (1098 records) data sets	244
TABLE 123 Additional variables not included in the COAST knee model, based on the 608- and 1025-record incomplete data sets	245
TABLE 124 Additional variables not included in the COAST Knee model, using the 608- and 1025-record data sets	246

List of figures

FIGURE a Summary of work programme	xxx
FIGURE 1 Trends in primary THR and TKR rates in (a) women; and (b) men, based on data from Culliford <i>et al.</i>	9
FIGURE 2 The sex ratio for the number of hip and knee replacements vs. year of procedure based on data from Culliford <i>et al.</i>	10
FIGURE 3 Age-standardised regional rates between south-west England and London	12
FIGURE 4 Revision rates for (a) hip; and (b) knee	14
FIGURE 5 Trends in revision rates with 95% CIs 1991 to 2006: (a) hip; and (b) knee	15
FIGURE 6 Mean age at revision, with 95% CIs, in 1991–2006: (a) hip; and (b) knee	16
FIGURE 7 Ratio of incidence rates for knee to hip revisions for men and women between 1991 and 2006	17
FIGURE 8 Estimated lifetime risk at age 50 years of undergoing (a) THR; or (b) TKR based on age- and sex-specific incidence adjusted for mortality	23
FIGURE 9 Patient satisfaction at 2 years after surgery: (a) ROC plot; and (b) 75th percentile	35
FIGURE 10 The ROC curve cut-off points for satisfaction at 24 months using 24-month OHSs, stratified by baseline clinical variables	36
FIGURE 11 Bar charts illustrating the cut-off points of (a) OHS; and (b) OKS in relation to satisfaction 6 months post surgery	37
FIGURE 12 Percentage of potential changes	38
FIGURE 13 Scatterplots showing relationship between (a) baseline OKS; and (b) baseline OHS compared with satisfaction at 6 months after surgery	40
FIGURE 14 Distributions of OKS at (a) baseline; and (b) 6 months after surgery	42
FIGURE 15 The distribution of OHSs at (a) baseline; and (b) 1 year postoperatively	43
FIGURE 16 Change in OHS over time, stratified by baseline score, from Judge <i>et al.</i>	44
FIGURE 17 Cumulative incidence estimate for revision of TKR by age	46
FIGURE 18 The change in OHS over time, layered by age, from Judge <i>et al.</i>	47
FIGURE 19 Cumulative incidence estimate for revision of THR by age	47
FIGURE 20 Estimated cumulative incidence for TKRs by BMI	51

FIGURE 21 Figure showing estimated cumulative incidence for THRs by BMI	54
FIGURE 22 Fixed-effects meta-analysis from Batra <i>et al.</i>	54
FIGURE 23 Change in OHS over time, stratified by BMI from Judge <i>et al.</i>	56
FIGURE 24 Change in OHS over time, stratified by number of coexisting diseases, from Judge <i>et al.</i>	61
FIGURE 25 Change in OHS over time, stratified by SF-36 mental health categories from Judge <i>et al.</i>	62
FIGURE 26 Change in OHS over time, layered by femoral component offset (mm offset) categories, from Judge <i>et al.</i>	65
FIGURE 27 Kaplan–Meier estimates of probability of revision surgery according to hormone replacement therapy use	65
FIGURE 28 Example of a web-based calculator displaying the predictive tool based on the patient’s individual preoperative characteristics	72
FIGURE 29 The ROC curve showing discriminatory ability of the model to identify those patients with the worst outcome (change in OHS of 5 units or fewer)	73
FIGURE 30 Markov model for the cost-effectiveness of an outcome prediction tool for THR	83
FIGURE 31 Mean EQ-5D summary scores mapped from OHSs reported by EPOS patients	96
FIGURE 32 Expert elicitation exercise: opinions on distribution of probabilities of referral to risk modification state	104
FIGURE 33 Receiver operating characteristic curve: primaries	104
FIGURE 34 Percentage of good and poor outcome following primary TKR, by outcome groups	105
FIGURE 35 The ROC curve for the COAST-derived model predicting poor outcome based on resource use	108
FIGURE 36 Distribution of costs for the first year following primary TKR	108
FIGURE 37 Mean health utility accounted by sex and age	109
FIGURE 38 The mean EQ-5D score observed in KAT patients before and after TKR	110
FIGURE 39 Progression of health utility, based on the KAT	110
FIGURE 40 Results of Monte Carlo simulations on the cost-effectiveness plane	115
FIGURE 41 Cost-effectiveness acceptability curve	115

FIGURE 42 Results of the probabilistic sensitivity analysis in the cost-effectiveness plane for women aged 70 years	118
FIGURE 43 Pie chart showing the breakdown of COASt eligible patients cohorts	127
FIGURE 44 Showing personnel allocated to the study in Oxford and Southampton	128
FIGURE 45 Recruitment process in SGH	131
FIGURE 46 Recruitment flow chart at the NOC	132
FIGURE 47 Number of patients included in COASt	133
FIGURE 48 Correlation matrix plot for (a) the complete; and (b) imputed hip data sets	146
FIGURE 49 Density plot of the observed and predicted scores for the hip data using the complete data set	148
FIGURE 50 Histograms of the outcome using (a) baseline; and (b) absolute change in the OHS, using the complete data set (550 records)	149
FIGURE 51 Calibration plot of the validated hip model using the complete data set	149
FIGURE 52 Density plot of the (a) observed; and (b) predicted scores for the hip data using the complete data set	150
FIGURE 53 Calibration plot of the validated hip model using the imputed data	151
FIGURE 54 Density plot of the predicted scores for the improved model using the hip data	153
FIGURE 55 Correlation matrix plot for the incomplete knee data sets with (a) 608 records; and (b) 1025 records	157
FIGURE 56 Calibration plot of the validated knee model using the 608-record data set	160
FIGURE 57 Density plot of the (a) observed; and (b) predicted OKS using the 608-record data set	161
FIGURE 58 Histograms of the (a) baseline; and (b) absolute change in the OKS, using the 608-record data set with imputed data	161
FIGURE 59 Density plot of the predicted scores for the improved knee model	163
FIGURE 60 Summary of programme	167
FIGURE 61 Hip replacement model used by six economic evaluations and first reported by Briggs <i>et al.</i>	213
FIGURE 62 Full THR economic model schema	215

- FIGURE 63** Estimated GP day visits because of hip pain by THR patients during the year prior to surgery, patients aged > 45 years in the CPRD 221
- FIGURE 64** Receiver operating characteristic curve for the model predicting surgery outcome category 224

List of abbreviations

ACL	anterior cruciate ligament	HTA	Human Tissue Authority
ANCOVA	analysis of covariance	ICER	incremental cost-effectiveness ratio
ASA	American Society of Anesthesiologists	IMD	Index of Multiple Deprivation
AUC	area under the curve	IMSU	Information Management Services Unit
BMD	bone mineral density	IQR	interquartile range
BMI	body mass index	IT	information technology
BNF	<i>British National Formulary</i>	K/L	Kellgren and Lawrence grading system
CEAC	cost-effectiveness acceptability curve	KAT	Knee Arthroplasty Trial
CI	confidence interval	MICE	multivariate imputation by chained equations
COAST	Clinical Outcomes in Arthroplasty Study	NCOAST	North COAST study (Oxford)
CPRD	Clinical Practice Research Datalink	NDORMS	Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences
DEXA	dual-energy X-ray absorptiometry	NICE	National Institute for Health and Care Excellence
DICOM	Digital Imaging and Communications in Medicine	NIHR	National Institute for Health Research
DSAC	Data and Sample Access Committee	NJR	National Joint Registry
DVT	deep-vein thrombosis	NOC	Nuffield Orthopaedic Centre
EOC	Elective Orthopaedic Centre	NSAID	non-steroidal anti-inflammatory drug
EPOS	Exeter Primary Outcomes Study	OA	osteoarthritis
EQ-5D	EuroQol-5 Dimensions	OARSI	Osteoarthritis Research Society International
EUROHIP	European Collaborative Database of Cost and Practice Patterns of Total Hip Replacement	OHS	Oxford Hip Score
GP	general practitioner	OKS	Oxford Knee Score
GPRD	General Practice Research Database	OLS	ordinary least squares
HCS	high-compliance system	OMB	Oxford Musculoskeletal BioBank
HES	Hospital Episode Statistics	ONS	Office for National Statistics
HR	hazard ratio	OR	odds ratio
HRG	Healthcare Resource Group	OUH	Oxford University Hospitals NHS Trust
HRQoL	health-related quality of life	OXMIS	Oxford Medical Information System
hsCRP	high-sensitivity C-reactive protein		
HSE	Health Survey for England		

LIST OF ABBREVIATIONS

PACS	picture archiving and communication system	SGH	Southampton General Hospital
PASS	patient-accepted symptom state	SHR	subhazard ratio
PCT	primary care trust	SWLEOC	South West London Elective Orthopaedic Centre
PE	pulmonary embolism	THA	total hip arthroplasty
PIS	patient information sheet	THR	total hip replacement
PoPC	percentage of potential change	TJR	total joint replacement
PPI	patient and public involvement	TKA	total knee arthroplasty
PPR	patient and public representative	TKR	total knee replacement
PROM	patient-reported outcome measure	TTU	transfer to utility
PSA	probabilistic sensitivity analysis	UHS	University Hospitals Southampton NHS Foundation Trust
QALY	quality-adjusted life-year	UKR	unicompartmental knee replacement
RA	rheumatoid arthritis	VAS	visual analogue scale
RCT	randomised controlled trial	VIF	variance inflation factor
ROC	receiver operating characteristic	WOMAC	Western Ontario and McMaster Universities Osteoarthritis Index
SD	standard deviation		
SF-12	Short Form questionnaire-12 items		
SF-36	Short Form questionnaire-36 items		

Plain English summary

In this project we collected data on the number of hip and knee replacements across the UK. To help health-care planning, we have estimated the rapidly growing demand for hip and knee replacements. We also investigated patient outcomes and developed a new way of measuring the success of surgery for individual patients. We identified the factors that predict poor results of knee and hip replacements, postoperative complications and the need for further surgery. We have found that patients who have severe pain and poorer joint function before surgery experience more unsatisfactory outcomes after operation. Poor surgery outcomes were associated with older age, deprived socioeconomic background and poor mental health. Although the association between increased weight and poor outcomes was weak, we found that overweight patients are more likely to need revision surgery.

We combined risk factors and designed statistical tools separately for hip and knee to identify patients likely to have unsatisfactory outcomes. We then tested the tools for practicality in two NHS trusts, in Oxford and Southampton. Although the predictive tools slightly underperformed, inclusion of additional risk factors in the tools improved their performance. Better results would be likely if we used a wider spectrum of risk factors.

Economically, hip and knee replacements are, in general, undoubtedly cost-effective in improving the long-term quality of patients' lives. It is important, therefore, that any measure that directs patients away from surgery because of likely poor outcomes is accurate and offers good alternatives. Further work is in progress to help both patients and their doctors make reliable decisions about future treatment in individual cases.

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

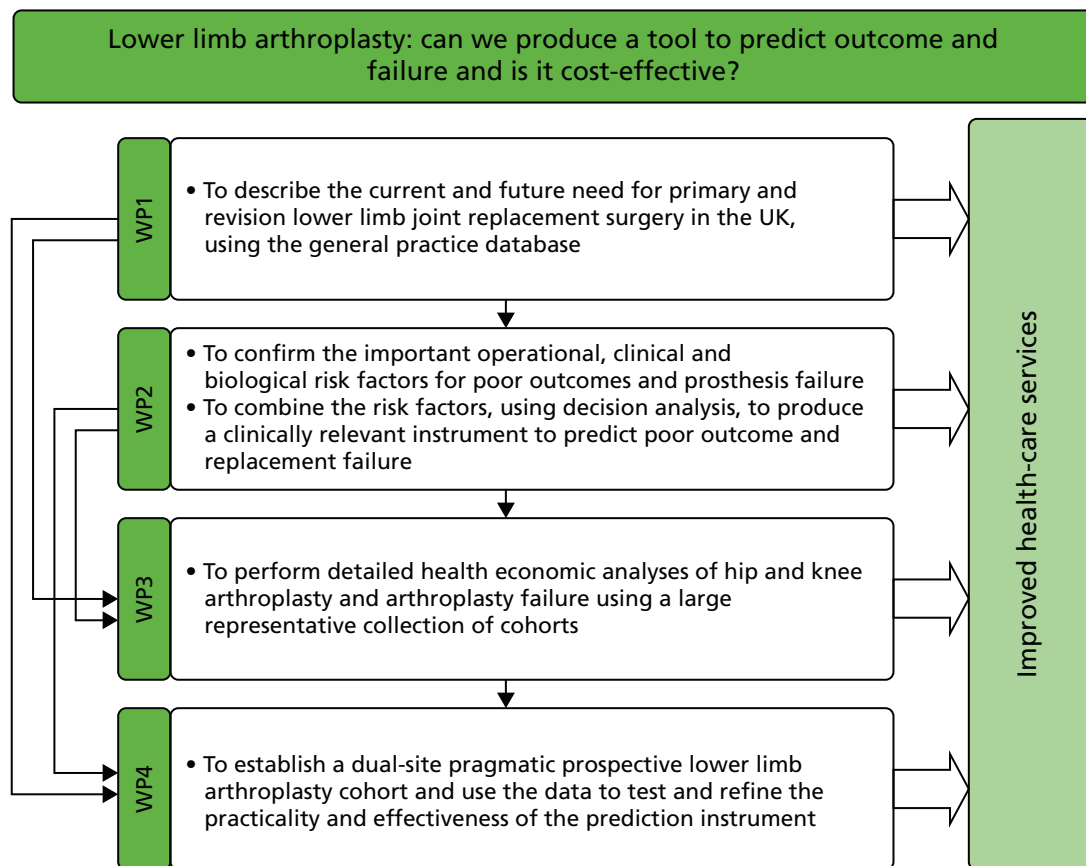


FIGURE a Summary of work programme. WP, work programme.

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.

Chapter 1 Introduction

Background

Osteoarthritis (OA) is one of the most common musculoskeletal conditions worldwide and accounts for > 90% of total knee replacement (TKR) and total hip replacement (THR) procedures in the UK.¹ There is currently no acceptable medical therapy for reducing the onset or progression of OA and the current treatments are aimed at symptom relief and increasing mobility.^{2,3} The only successful treatment for patients with OA of the lower limbs is arthroplasty.

A wide range of high-quality support measures and treatments are needed for individuals suffering from musculoskeletal conditions. Choice of treatment is based on type of condition, severity of symptoms and access to health-care services and professionals. An estimated 30% of general practitioner (GP) consultations⁴ and 40% of visits to NHS walk-in centres^{5,6} are for musculoskeletal-related conditions.

Currently, lower limb arthroplasty is the most common and most successful elective orthopaedic procedure undertaken in the UK in patients with OA affecting the hip and knee joints.^{7,8} It provides substantial relief from pain and improves physical function,^{2,8-10} and is considered the most successful and cost-effective operation for end-stage disease.^{2,9,11,12} It is therefore not unexpected that the workload of trauma and orthopaedic surgical services has intensified.

Musculoskeletal conditions are a major cause of ill health, pain and disability, placing a significant burden on the NHS.¹³ Evidence suggests that this burden will only increase as a result of a growing elderly population and an increase in obesity.¹⁴ A combination of these factors, in association with related comorbidities,¹⁵ highlights an urgent need for accurate and reliable data to ensure effective long-term planning and equitable resource allocation across all regions in the UK. It is therefore essential that the current needs for surgery are accurately described and future trends estimated for effective planning of health-care services.

Surgical trends

It has been estimated that a total of 52,048 THRs and 44,645 TKRs were performed in 2002, and the number of elective procedures increased markedly in the period 1989–2004.¹⁶ These surgical interventions remain the top elective procedures performed by orthopaedic surgeons in the NHS and, in the 12 months to April 2012, a total of 75,366 hip replacements and 76,497 knee replacements were performed¹⁷ in the NHS.

The reasons for the increase in demand are widely debated and suggested explanations include improved instrumentation and prostheses survival rates, but also increasing numbers of patients with OA (related or unrelated to comorbid chronic conditions such as obesity) and the significant increase in total joint replacements (TJR) among individuals aged > 55 years.¹⁸

Patient choice is important to understand and may be a factor affecting demand, as suggested by a Canadian study assessing OA patients' perception of total joint arthroplasty as an intervention, which found that the willingness to undergo surgery was inversely associated with misperceptions about its appropriateness.¹⁹ The same team also found low rates of willingness among those with disabling arthritis.²⁰ Women seemed to be less willing than men to undergo surgery²¹ and less willing to undergo TKR than THR²² as a suitable intervention. Jüni *et al.*²³ previously reported that 32% of patients considered for TKR, for a variety of reasons, did not consider surgery an option. The group conducted a population-based study of TKR in the

south-west of England, using an assessment of need based on the New Zealand Score,²⁴ and found that differences in the perception of disease severity may account for some of the underprovision reported by the study team.

Some estimates have been produced from historical rates of arthroplasty but up-to-date information is essential to estimate future rates. Historically, published findings were often based on small data sets unrepresentative of a general population and administrative codes lacked specificity. The potential of these codes in providing accurate and clinical descriptions was recognised, and revision codes were updated in October 2005.²⁵ Furthermore, published results have used simplistic models to predict changes in future rates and have produced unrealistic information.²²

Predictors of surgery outcomes and prediction tool

Although the majority of patients improve after hip and knee arthroplasty, an important group of patients continue to experience some pain and functional disability after THRs and TKRs, and some experience no improvement or get worse.^{26–35} This especially applies to TKR surgeries, as a number of studies have identified that a small minority of patients are not satisfied with their knee replacements.^{26,34,36}

Arthroplasties are successful interventions for end-stage disease and have been known to provide pain relief and improved physical function.⁸ However, for years, the approach has tended to focus on revision as an outcome, with few data on patient-reported outcomes. In more recent years, the government has accepted the importance of patient-reported outcomes for this operation and introduced the patient-reported outcome measures (PROMs) for monitoring the outcomes of such patients.

There is consistent evidence in the joint arthroplasty literature that up to 30% of patients are dissatisfied with their outcomes.^{37,38} It still not entirely clear from the available evidence what factors contribute to dissatisfaction. For example, Gandhi *et al.*³⁸ and Hawker *et al.*²⁶ found that preoperative pain and function were not associated with patient satisfaction and, yet, Kim *et al.*³⁹ and Scott *et al.*⁴⁰ demonstrated that less preoperative pain is suggestive of increased satisfaction.

As a result of advances in arthroplasty devices and improvements in technical surgical skills and expertise, a successful long-term outcome is now achieved in the majority of patients undergoing hip or knee replacements. It has, however, also become evident that prosthesis survival may not be an accurate or true measure of success when patient satisfaction is taken into account and, by this criterion, a small but important group of patients do not improve or even get worse.²⁷ Following this understanding, the focus has moved away from implant survival to patient-reported outcomes that concentrate on the patient's experience and level of satisfaction with the operation.⁴¹ The difficulty lies with identifying the determinants of outcome, as well as using the most appropriate and accurate method for collecting and interpreting the patient-reported outcomes. A successful joint replacement should result in pain relief, function improvement and patient satisfaction.⁴²

Total hip replacement is successful in the majority of patients. However, there is growing evidence that a small, but important, minority of patients show no improvement or get worse.^{27,32,33,35} PROMs are now commonly used to determine the result of knee and hip surgical interventions. However, little work has been done to establish the predictors of good or bad patient-reported outcomes after THR.⁴³

Determinants of outcomes for THR have been widely researched and include baseline levels of pain and function,^{32,44–47} severity of clinical disease,⁴⁵ age,^{45,47,48} sex,^{45,46,49} radiographic grade,^{45,50} education,^{32,44,49,50} obesity,^{46,48} comorbidities,^{32,46} living alone,^{46,51} mental health⁴⁷ and patients' expectation of surgery.⁵² The results in literature are conflicting. For example, some studies have found that age, sex, body mass index (BMI) and comorbidities are not predictive of outcome,^{26,37–40} whereas others have found that a lower level

of education and higher BMI are associated with dissatisfaction,²⁶ and Noble *et al.*⁵³ found an association between age and satisfaction.⁵³

Validation of some key issues about age, sex, rates and indication for surgery is needed for a better understanding of these surgical interventions in order to effectively target treatment. Of the patient factors for poor outcome, there is ongoing controversy as to the role of obesity.⁵⁴ Although some have found no difference in clinical outcome after THR⁶ or TKR,⁵⁵ others have demonstrated that obesity is a recognised risk factor for poor outcome after TKR.^{6,56}

Surgical technique and implant type are important factors in the outcome of hip and knee arthroplasty. However, in a recent systematic review, Kynaston-Pearson *et al.*⁵⁷ reported that there is no clinical evidence of effectiveness in one-quarter of available hip prostheses in the UK. The relationships between implant users, manufacturers and suppliers have been in development over a number of years, as implant costs contribute appreciably to the overall cost of surgery. Lack of implant regulation became an area of focused concentration recently as a result of the adverse outcomes in metal-on-metal resurfacings and large-bearing-surface implants.⁵⁸

It is only recently that regulatory frameworks have started to focus on the safety regulations around implants, and much more needs to be done to ensure that tested prosthetics are both safe and clinically effective.³ Other technical factors include case volume, technique and choice of prosthesis. Technical factors in performing surgery, such as component alignment,⁵⁹ influence both the short- and long-term success rates. The technical ability of the surgeon also plays a vital role in the successful outcome of hip and knee replacements, and continues to drive the ongoing development and refinement of implants, surgical techniques, skills and training.

Historically, most research has focused on implant failure as the main outcome. There are three main causes of implant failure: aseptic osteolysis, infection and inflammation. Osteolysis results, in part, from resin wear, leading to local inflammation and accelerated bone resorption.⁶⁰ Differences between implant design, resin storage and type affect resin wear and osteolysis.^{61,62} Currently, radiographic assessment has poor sensitivity for detecting osteolysis, requiring at least 50% of demineralisation to occur before osteolysis can be detected. However, osteolysis may be better detected using structural measurements, such as fractal analysis, than changes in density.^{63,64} Although initially considered a purely degenerative disease, there is increased inflammation in both synovial fluid and cell membranes of osteoarthritic joints, and this may play a key role in arthroplasty failure.⁶⁵

Other common complications of arthroplasties include infection, vascular or thrombotic compromise, dislocation, instability and fracture. Infection is a common cause of early failure and can lead to significantly poorer clinical outcomes, such as amputation or revision surgery. Deep-seated infections may be the leading cause of implant failure over the next 20 years.⁶⁶ The diagnosis of surgical site infection following TJR requires a balance between quality and practicality. Revision after septic failures has a higher failure rate than revision after aseptic failures, highlighting the importance of accurate identification of sepsis.⁶⁰

Extensive research into the diseases commonly associated with degenerative changes in joints has been conducted, as an understanding of the underlying causes could assist with predicting the outcome of surgery. One of the most common musculoskeletal conditions, OA, accounted for > 90% of the total knee and hip arthroplasties (153,000 procedures) in the UK up to 2010.¹ OA is the major cause of health, pain and disability and increased mortality. Two main risk factors for OA are age and obesity, both of which are increasing in the population in the West.⁶⁷ For this reason, it is almost inevitable that the prevalence of OA will increase substantially in the next 20 years.⁶⁸

Osteoarthritis has been recognised as a global burden and is the most frequent primary indication for total hip and knee replacements in the UK,⁶⁹ accounting for 93% of hip replacements and 97% of knee replacements in England and Wales.⁷⁰ In the UK, 550,000 people have moderate to severe knee OA and

210,000 moderate to severe hip OA. Each year approximately 2 million people consult their GP for OA and 115,000 are admitted to hospital. The prevalence of hip and knee OA is particularly high in the population aged > 60 years.⁷¹⁻⁷⁴ The lifetime risk of hip OA has been calculated at 25%⁷⁵ and of knee OA has been calculated at 45%.^{14,71,76} Changes in the reported prevalence of OA, such as the overall increase and increased prevalence in younger patients found by Kim,¹⁴ have to be substantiated and validated to inform new treatment algorithms for local services.

Patient selection,^{3,77} implant design and surgical technique are all key factors that could affect the durability of a prosthetic implant.⁷⁸ Historically, outcome studies used continuous variables at population levels to identify statistically significant predictors; however, their clinical relevance is less clearly understood, especially by patients. Understanding and identifying patients at risk of poor patient-reported outcomes and presenting these in a clinically meaningful way to an individual patient will enable clinicians to evaluate the risks and benefits of surgery on an individual level.

The lack of information led to well-publicised decisions by primary care trusts (PCTs) in Suffolk to temporarily withhold hip and knee arthroplasty from obese subjects.⁷⁹ This decision was overturned because of a lack of supporting evidence. We urgently need data to identify patients at a high risk of poor outcome both before surgery, in order to minimise risk factors, and in the early postoperative period, to initiate urgent interventions to improve outcomes and prioritise resources.

A number of individual determinants of implant failure have been described in the literature; however, the majority of patients exhibit more than one cause of failure,⁵⁹ and the benefit of combining risk factors are not known. The current literature describes a wide range of risk factors in a prognostic model,^{80,81} including age, sex, education, obesity, mental health status, preoperative level of pain and function, indication for surgery, coexisting conditions, radiographic variables (radiographic grade) and surgery-related risk factors (i.e. femoral component offset). In this programme we aimed to develop similar prognostic models for the knee and hip, and to include a wider range of risk factors to predict pain and function outcomes.

As personalised medicine becomes increasingly common, it is essential that the correct patients are chosen to undergo hip and knee arthroplasties, which are important but complex procedures. This emphasises the importance of understanding the predictors of patient-reported outcomes of satisfaction and pain or function scores. This programme aims to address these issues. Previous work on outcomes focused very heavily on prosthesis and little attention was paid to surgical- and patient-related factors that predict outcomes. It is important to look at all three components and their interactions to predict surgery outcomes accurately. The information then can be used to identify patients with good or poor outcomes and form the clinical decision-making tool to allow stratification of patients for surgeries with patient-informed consent.

With increasing restrictions on funding in the NHS, it is critical to have accurate and reliable data from practice, alongside current and future population-based estimates, for a better understanding of these surgical interventions. This would aid our understanding of the clinical effectiveness and cost-effectiveness of lower limb arthroplasty and help to target resources more efficiently.

Cost-effectiveness of implementation of the tools

Lower limb arthroplasties are a considerable burden on NHS resources. Estimations by Jenkins *et al.*⁸² suggest the cost per procedure to be in excess of £7000. In the USA, TJR is a cost-saving or cost-effective procedure in those with significant functional limitation as it avoids high care costs resulting from the disability of OA.⁸³ Early improvements in the management of patients, such as decreasing length of stay, resulted in overall reduced costs.^{84,85} Although the overall costs of primary TJR have decreased, the procedural costs of revision surgery continue to increase.⁸⁶ In the UK implant survival data are impressive,

with a 5-year revision rate of 4.5% for THR and 5.1% for TKR.⁷ Yet, although it may be a technically successful replacement, up to 20% of knee replacement patients still have a poor outcome and a small, but important, proportion of patients who have had hip replacements do not achieve a clinically meaningful symptomatic improvement or their symptoms get worse.^{27,87}

Accurate cost-effectiveness data are essential for the appropriate evaluation within the NHS of the incremental cost-effectiveness ratio (ICER) of using more expensive prosthetic components that may improve implant survival.^{88,89} As well as validated predictors of poor outcome following TJR outcome, cost implications are important for informing patient expectations.⁹⁰

In addition to optimising the outcome of patients undergoing arthroplasty, it is important for NHS commissioners to have accurate data on the cost-effectiveness and cost-utility of these operations. Current health economic data are limited for several reasons. The main limitations of the data on utility gains post surgery are that they are from small cohorts, they do not differentiate between different patient profiles, they are limited to outcomes at 10 years, with limited data on short-term gains, and, importantly, they use revision surgery and not ongoing health-care utilisation as a result poor functional outcomes.

In this programme we aimed to design a clinical tool to predict patients who will experience poor outcomes following THR and TKR. Taking into account the fact that these procedures are costly and exert a significant burden on the NHS, we need to ascertain if the additional cost of the implementing tools would be worthwhile in terms of benefits to an already overstretched current health-care system; that is, if the tools would be a cost-effective use of resources in the UK health-care system. With this in mind, we aim to provide an economic evaluation of the implementation of the tool in the health-care setting. The availability of predictive tools, and detailed cost-effectiveness and cost-utility data will help to produce a coherent strategy for the provision of a clinically effective and cost-effective strategy for the provision of lower limb arthroplasty in the NHS. The information collected and analysed in the development of a predictive tool for hip and knee replacement will also support the development of patient-based, informed decision-making programmes.⁹¹

External validation of the tool

As part of the programme we aimed to test the productiveness, practicality and cost-effectiveness of the developed tool in the pragmatic cohort of NHS setting. This required us to recruit a cohort of patients undergoing hip and knee arthroplasties in which the productiveness, practicality and cost-effectiveness of the tools would be tested.

Aims of the programme grant

We aimed to inform the policy-makers of the current health-care system in the UK about predicting the outcomes and failure of lower limb arthroplasty, and give advice on the cost-effectiveness of implementation of predictive tools. We set out to achieve this through four work packages (1–4), as described in the subsequent chapters (see *Chapters 2–5*) of this report.

In *Chapter 2* (work package 1) we describe the current and future needs for primary and revision lower limb joint replacement surgeries in the UK using a national longitudinal prospective database.

In *Chapter 3* (work package 2) we look at the predictors of poor outcome following lower limb arthroplasties using extant databases. We report on combining these databases to produce predictive tools separately for knee and hip for patient-reported outcomes at 12 months.

In *Chapter 4* (work package 3) we describe the detailed body of work looking at the cost-effectiveness of implementing the tool to predict the outcomes following knee and hip arthroplasties using the extant databases, nationally available routine data and our prospective cohort of patients who were recruited in work package 4.

In *Chapter 5* (work package 4) we describe our prospective new cohort and the steps performed in this prospective pragmatic cohort to detail the external validation of the tools developed in work package 2.

Chapter 2 Work package 1: current and future rates of lower limb arthroplasties

This chapter describes the current and future needs for primary and revision lower limb joint replacement surgeries in the UK using a national prospective database.

The chapter contains information covered in work package 1. The objectives in this work package were to:

1. describe and estimate the rates of THR, TKR and unicompartmental knee replacement (UKR) in the UK
2. describe regional and national variation in hip and knee replacement surgery in the UK
3. describe the mechanics of revision for hip and knee arthroplasty and quantify the rates in the UK
4. predict future trends in hip and knee surgery in the UK, accounting for projected changes in age and obesity.

Design and setting

In order to quantify the rates of lower limb arthroplasty in the UK we used a prospective cohort from the Clinical Practice Research Datalink (CPRD), formerly known as the General Practice Research Database (GPRD). The CPRD is a recognised and frequently utilised database in epidemiological studies,⁹² and is validated.^{69,93} It is a computerised medical records system that is representative of the UK population and has been validated for a wide range of medical conditions.⁹⁴ The CPRD database has data from over 6 million patients across more than 600 practices and has been collecting data since 1987. The data set has been validated and audited, and only practices providing good-quality data are admitted. Every patient is registered with one general practice. GPs are responsible for providing primary care and referral services to their patients and keep comprehensive records that contain prescription data, clinical events, specialist referrals, hospital admissions and their major outcomes. These systems are commonly used in general practice for the classification of diseases. Personal details are encoded and all patients are provided with clinic identifiers to ensure confidentiality. The database is administered by the Medicines and Healthcare products Regulatory Agency. The CPRD database provides a unique resource to examine the outcome of joint arthroplasty.

The CPRD database is accepted as being broadly representative of the UK population with respect to age, sex, socioeconomic circumstances and region; the data and Read codes for diseases (see *Appendix 1*), which are cross-referenced to the *International Classification of Diseases*, Ninth Edition,^{94–96} are stored in the Oxford Medical Information System (OXMIS).

The CPRD data were used to answer several research questions within different outputs, which contributed to the results described in this chapter. The design details of the bespoke computer programs written to manipulate and post-process the raw CPRD data are not provided. The methods used in the published articles,^{67,97–99} produced as part of work package 1, are described in general terms.

The CPRD data are routinely gathered for contributing practices and are not explicitly censored, other than when patients die or when they leave a general practice. The data delivered to our research team were truncated at 31 December 2006 for practical purposes, as the data ‘cut’ were taken from the main CPRD database shortly thereafter. No minimum contribution time was imposed, but the CPRD does impose a practice-level requirement that the data delivered to the database by each practice should be ‘up to standard’ according to the CPRD’s definitions, which in effect means that each practice submits data for up to 6 months before the data were confirmed as being up to standard. Other than this exclusion criterion, which is applied to all CPRD studies, we applied no further inclusion/exclusion criteria other than those reported in the individual research outputs written. Consequently, studies using CPRD data are ‘real world’, in the sense that the data recorded in general practice are used as research data, and therefore constitute a sensible sample from which to make population-level inferences about the UK population of GP-registered patients.

Sets of Read codes (see *Appendix 1*) (including remapped OXNIS codes, which have been phased out) were used for all of the data selection from the CPRD. Two or more clinicians validated the code lists for replacement procedures and clinical consensus was reached. Regarding the potential miscoding of primary THR/TKR, we took the first code match as the date of primary replacement.

A small number of subjects were found to have more than two primary operations, which is not strictly possible according to the usual definition, and without linkage to register data it is impossible to know which are genuine and which are not. Similarly, sidedness (left or right) of the procedure is not identifiable from Read codes. We took a pragmatic but consistent approach by identifying the first primary encountered as the one to use. In addition, we acknowledge that, without detail on sidedness, we cannot be sure that, for example, a left-sided revision was matched with a left-sided primary, but register data suggest that the effect of these potential mismatches on estimated incidence rates, lifetime risks and hazard ratios (HRs) would be small.

The first phase of using the data set for this work package involved the construction and validation of the GPRD data set with internal checks for consistency. During the analyses we also compared the summary data set with other available and appropriate data sets such as the National Joint Registry (NJR) for England and Wales, Hospital Episode Statistics (HES) and Health Survey for England (HSE), and population forecast data from the Office for National Statistics (ONS) in order to establish external validity.

Main exposure

The main exposure is primary arthroplasty. The selected cohort contains all of the patients with a code for primary or revision hip or knee replacement surgery and meets the criteria for each particular analysis.

Outcome measures

This work package centres on describing the epidemiology of hip and knee replacements. The relevant outcome measures are incidence rates, estimates of revision risk and future projections of procedure counts.

Rates of total hip replacement, total knee replacement and unicompartmental knee replacement in the UK

Temporal trends

Temporal trends in hip and knee replacement in the UK⁶⁷

Total joint arthroplasty is a successful surgical intervention and is considered reliable and effective for pain relief and improved function and quality of life, with 90% prosthesis survival at 10 years.^{3,8} In total, 160,000 hip and knee replacements were carried out in England and Wales in the 12 months prior to April 2010.⁷⁰ This number is expected to rise, and studies from the USA predict an increase in hip replacements (174%) and knee replacements (673%) to nearly 3.5 million per annum by 2030.⁶⁶ More than 650,000 knee replacements alone were carried out in the USA in 2008,¹⁰⁰ and almost 80,000 in the UK in 2009.¹⁷ The USA saw the number of knee surgeries increase from 31.2 per 100,000 person-years [95% confidence interval (CI) 25.3 to 37.1 per 100,000 person-years] in 1971–1976, to 220.9 per 100,000 person-years (95% CI 206.7 to 235.0 per 100,000 person-years) in 2008.¹⁰¹

For the analysis we selected the patient data with a medical diagnosis code for THR ($n = 27,113$) or TKR ($n = 23,843$) between 1991 and the end of 2006. Patients were included if they were aged > 18 years at the time of operation. Evidence suggests that 20% of THR/TKRs in England are carried out in private institutions¹⁰² and the impact of this on arthroplasty provision needs further investigation. However, as the CPRD database had not been validated at this time, private practice codes were not included in the analysis. The NHS rates were validated using the HES (2005–6)¹⁰³ and NJR.¹⁰⁴

Directly age- and sex-standardised replacement rates for calendar years were calculated using 10-year age groups with the mid-year population estimates for 2003 as the reference standard, as published by the ONS,^{105,106} the General Register Office for Scotland and the Northern Ireland Statistics and Research Agency. The 95% CI was computed using the Poisson model appropriate for directly standardised rates. The mean age at total replacement was calculated for the hip and knee for each calendar year and 95% CIs computed. To investigate patterns over time we calculated age distribution at operation by sex for three consecutive 5-year periods for the hip and knee.

The estimated age-standardised rate of THR increased from 60.3 per 100,000 person-years (95% CI 53.7 to 67.0 per 100,000 person-years) to 144.6 per 100,000 person-years (95% CI 138.1 to 151.1 per 100,000 person-years) for women (*Figure 1a*), and from 35.8 per 100,000 person-years (95% CI 30.4 to 41.3 per 100,000 person-years) to 88.6 per 100,000 person-years (95% CI 83.4 to 93.7 per 100,000 person-years) for men (*Figure 1b*). The increase in rates over time for THR were steady between 1993 and 2005. The rate of

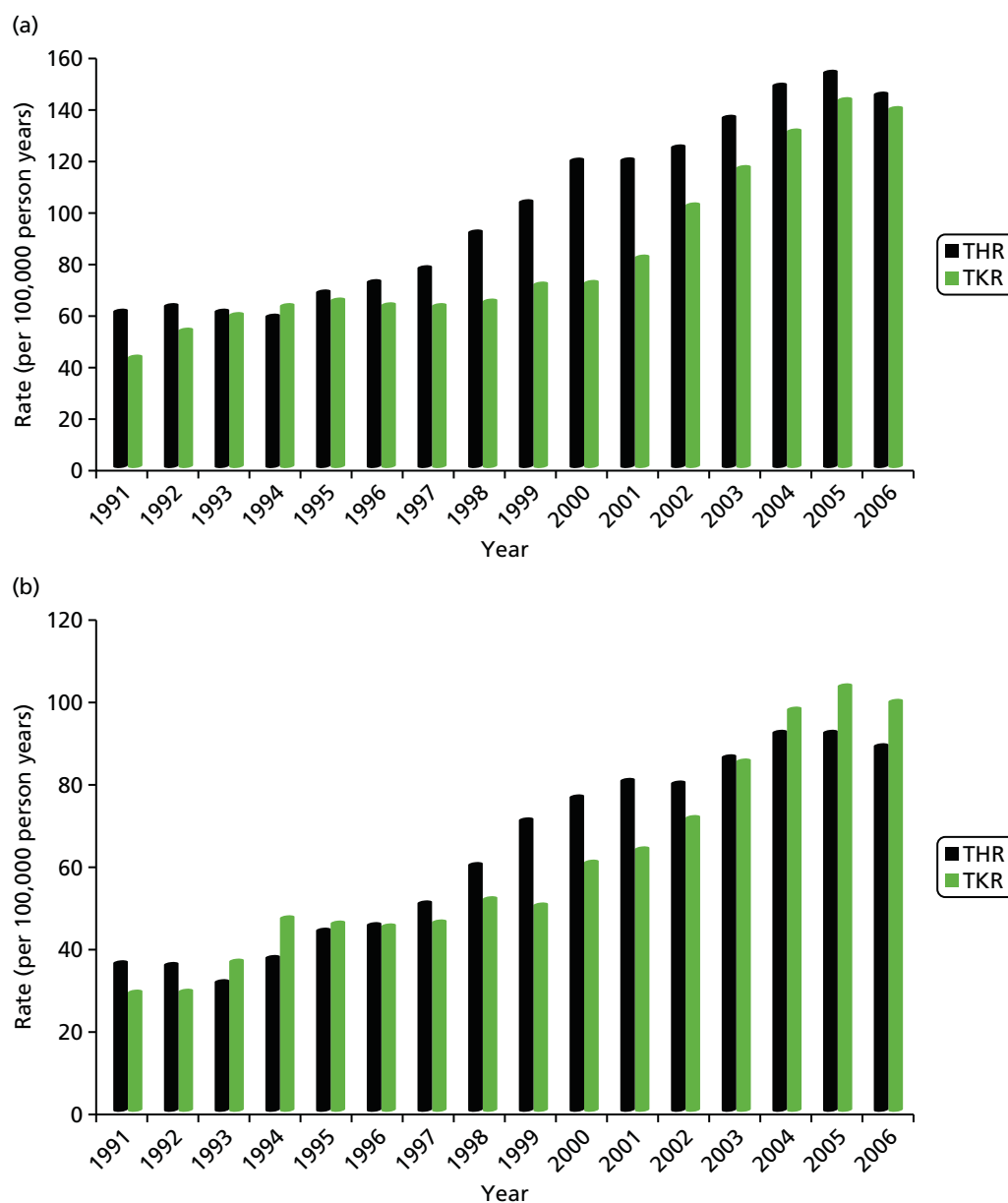


FIGURE 1 Trends in primary THR and TKR rates in (a) women; and (b) men, based on data from Culliford *et al.*⁶⁷

TKR increased from 42.5 per 100,000 person-years (95% CI 37.0 to 48.0 per 100,000 person-years) to 138.7 per 100,000 person-years (95% CI 132.3 to 145.0 per 100,000 person-years) for women (see *Figure 1a*), and from 28.7 per 100,000 person-years (95% CI 23.9 to 33.6 per 100,000 person-years) to 99.4 per 100,000 person-years (95% CI 93.9 to 104.8 per 100,000 person-years) for men (see *Figure 1b*). The temporal trend for knees showed a marked plateau from the mid-1990s, followed by a sharp rise from 2000.

The mean age at operation was significantly higher for women than for men for all years after 1991: the mean age at THR was 70.3 years (95% CI 69.8 to 70.8 years) for women and 67.6 years (95% CI 66.9 to 68.2 years) for men; and the mean age at TKR was 70.1 years (95% CI 69.6 to 70.5 years) for women and 69.2 years (95% CI 68.6 to 69.7 years) for men. The highest rates of THRs and TKRs were for women aged between 70 and 79 years, with a mean rate of THR of 541.8 per 100,000 person-years (95% CI 501.0 to 582.5 per 100,000 person-years) and of TKR of 555.3 per 100,000 person-years (95% CI 514.1 to 596 per 100,000 person-years).

The final results showed that the rates of hip and knee arthroplasty continued to increase, but that the rise was more marked for knees than for hips. Women were 67% more likely than men to undergo THR, and 45% were more likely to undergo TKR, but sex ratios have been consistent over time, as demonstrated in *Figure 2*.

Women were, on average, 3 years older than men at THR, but the age difference between men and women undergoing knee replacements was only half as great. BMI was significantly higher for patients undergoing TKR than for those undergoing THR ($p < 0.0001$) and was higher for women than for men. There was little sex difference in the number of replacements carried out in patients between the ages of 60 and 79 years, who made up almost two-thirds of the total number of patients undergoing arthroplasty during 1991–2006.

Unicompartmental knee replacements are also becoming more popular, and the rates have increased over the last decade. We, again, used CPRD data and Read/OXMIS codes to identify all patients who underwent

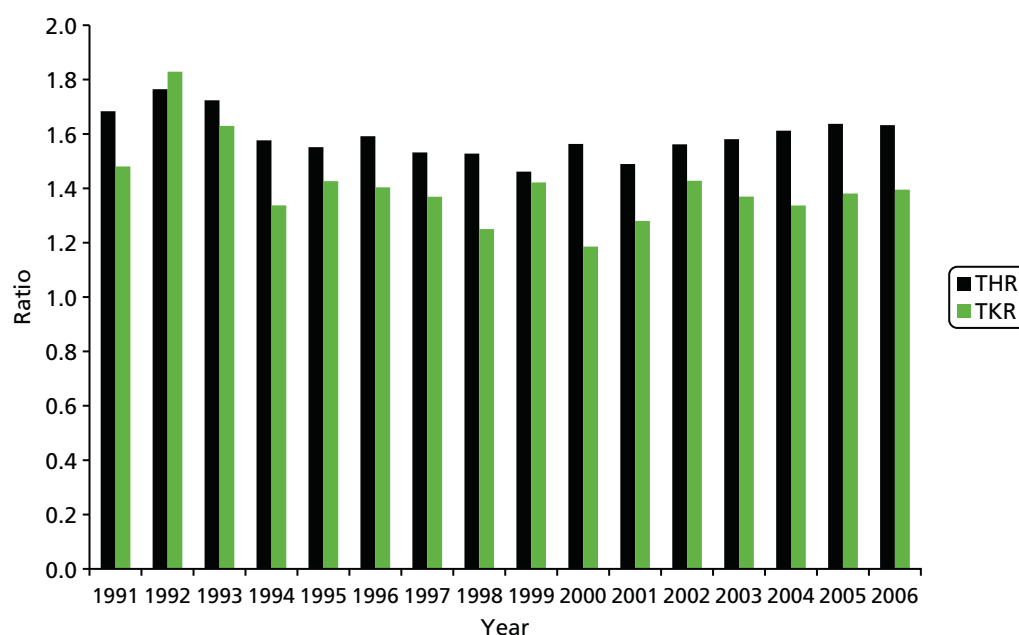


FIGURE 2 The sex ratio for the number of hip and knee replacements vs. year of procedure based on data from Culliford *et al.*⁶⁷

primary TKR or UKR between 1986 and 2006. However, the final analysis was restricted to the use of data between 1999 and 2006, as very few UKRs were carried out before this time.

The results of the statistical analysis give the number of TKRs and UKRs performed in each year; the mean age [and standard deviation (SD)] of patients of each sex and undergoing each type of operation was calculated to explore the profile for each intervention. The total numbers of TKRs and UKRs performed in the UK in 2006 were estimated by applying the CPRD rates to the population of the UK in that year.

There were substantially more TKRs ($n = 18,450$) than UKRs ($n = 266$) in 2006. The rate of TKRs increased from 55.4 per 100,000 person-years in 1999 to 123.5 per 100,000 person-years in 2006. The rate of UKRs increased from 0.25 per 100,000 person-years in 1999 to 3.0 per 100,000 person-years in 2006. Both men and women undergoing UKR were, on average, younger than those undergoing TKR ($p < 0.0001$). Men who underwent TKR were, on average, younger than women undergoing TKR ($p < 0.0001$). There was no statistically significant difference between the mean age of men and women undergoing UKRs ($p = 0.74$). TKR was performed more often in women ($n = 10,836$) than in men ($n = 7614$), but UKR was performed less often in women ($n = 126$) than in men ($n = 140$). The ratio of TKRs to UKRs fell from 250 : 1 in 1999 to 40 : 1 in 2006. The estimated numbers of operations performed in 2006 were 74,800 TKRs and 1800 UKRs.

The results showed a 12-fold increase in UKRs since 1999, and that this was still significantly less than TKRs, and UKRs are performed on a younger age group than TKRs.

Regional and national variation for hip and knee replacement surgery in the UK

Geographical and sociodemographic variations play an important role in the provision of, and access to, health care. Estimates of the mismatch of need and provision have been published by Judge *et al.*¹⁰⁷ and were found to be greater for TKR than for THR. There seems to be a wide variation in intervention rates for revision surgery across PCTs and the reasons need to be understood more clearly.

We looked at regional variations in the UK using CPRD data for 1986–2006 and found inter-regional differences in joint replacement rates. Using Read/OXMIS codes we identified 28,068 THRs and 24,364 TKRs.

Incidence was calculated by dividing the number of replacement operations by the number of person-years in the GPRD population. The rates were directly age and sex standardised, and computed by region, using a reference population (mid-2003 ONS population estimates). A 95% CI was calculated using a Poisson model.

Marked temporal changes were observed within and between certain regions. The reason for these differences is not clear, but factors such as medical indications and contraindications, personal and social perceptions of surgery as well as the availability of orthopaedic services should be considered. *Figure 3* shows the example of regional differences in hip replacement rates between south-west England and London, standardised by age.

The GPRD data showed significant inter-regional differences in joint replacement rates in the UK in the period 1991–2006. Marked temporal changes were observed within and between certain regions, and the differences are larger for hips than for knees. This is supported by other studies also using national databases and registries within the UK. Inequities and inequalities currently exist within the UK health-care system,¹⁰⁹ but the reasons need further investigation.

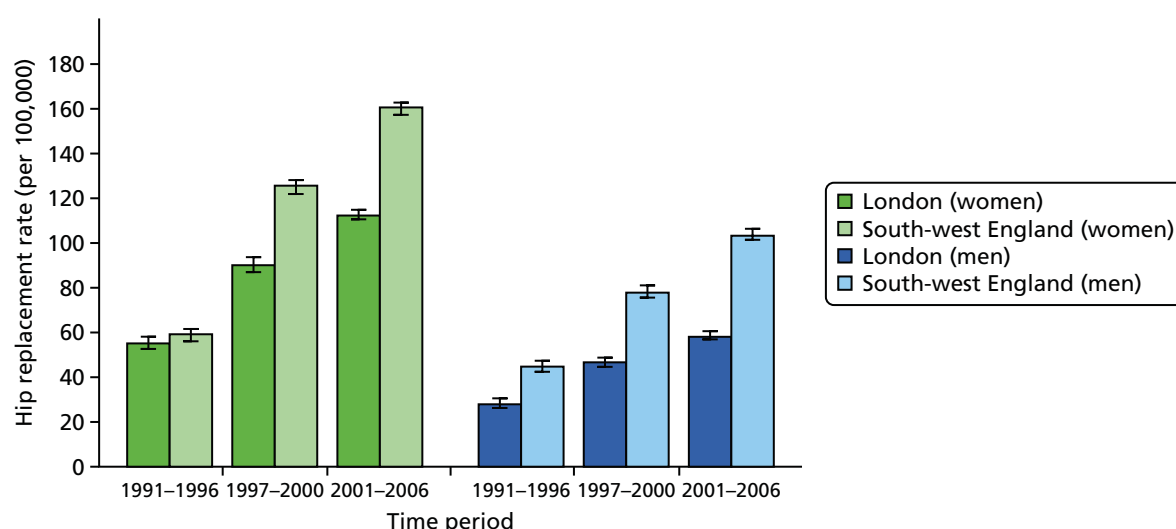


FIGURE 3 Age-standardised regional rates between south-west England and London.¹⁰⁸

Describing the revision rates for hip and knee arthroplasty, and quantifying the rates in the UK

Total joint replacements are very successful operations, but a number of patients continue to have problems or are not satisfied with the outcome. Hawker²⁷ estimated that up to 30% do not have symptomatic improvement after surgery. A further 20% of patients report unfavourable long-term pain.⁸⁷ There are many types of implants, and they are likely to be revised within the lifetime of a patient; for example, it is expected that a metal-on-polyethylene implant will need to be replaced after 20 years because of wear or prosthetic loosening. This is one of the reasons why THR and TKRs were indicated for mainly older patients. Even more modern prosthetics, making use of the latest technological advances, are not routinely recommended for younger patients. A UK population-based survey of patients after TKR found that 20% were not satisfied.³⁶ TJRs have, on average, a prosthesis survival rate of 90% at 10 years,¹¹⁰ and none has an indefinite lifespan.^{17,111} They are considered economical because of the low failure rate,^{78,112,113} but surgical intervention is recommended when they do fail. Revision surgery is a high-risk procedure with a significantly higher mortality and morbidity than primary joint replacement, and is more costly than primary replacements.^{114,115}

The revision rate is expected to increase as the population that requires hip and knee surgery increases because of an increase in lifespan in developed countries and changing demographics. Dixon *et al.*¹¹⁶ examined the trend in primary and revision TJRs in England and found a rapid increase in the proportion of hip surgeries requiring subsequent revision between 1991 and 2000, from 1 in 12 to 1 in 5; the number of knee revisions tripled over the same period, from 1 in 33 to 1 in 11.¹¹⁶ The increase in revision rates is expected to continue in parallel with the steady increase of primary joint replacements.¹¹⁷ Kurtz *et al.*⁶⁶ predicted an increase of 137% in hip revisions and of 601% in knee revisions in the USA by 2030 than 2005. Evidence from the Scandinavian National Joint Registries¹¹⁸ further demonstrates that the mean age for joint replacement is also decreasing.

An understanding of the reasons for failure, and success, of arthroplasty surgery is essential for guidance with implant design and clinical decision-making. Revision surgery is primarily indicated by implant loosening, instability through implant wear, or osteolysis and complications.

National and international registry data have been used extensively to estimate time to revision¹¹⁹ and to model prosthesis survival time in order to assess which specific demographic, clinical and prosthetic-specific factors are associated with time to failure.^{120,121} Appropriate commissioning of services will reduce waiting times by matching demand with capacity and improved health-care delivery. Compared with primary TJR, revision TJR is more costly and more technically difficult, and results in only a 65% improvement in symptoms, although it remains a cost-effective method for improving function, pain relief and quality of life.¹¹⁷

Revision rates continue to increase despite advances in surgical technique and implant design and the reasons for this remain unclear. Without this understanding it is difficult to address implant survival and long-term patient outcomes.¹²²

To determine the revision rates for the UK, we obtained data between 1991 and 2006 from the CPRD database. We used Read/OXMIS codes to identify all revisions to hip and knee replacements. Patients aged > 18 years at the time of operation were included in the analysis. Private practices were excluded because of lack of validation within the CPRD at this time.

For the analysis we calculated directly age- and sex-standardised rates for the incidence of revision for each calendar year. We used 10-year age groups, with 2003 mid-year population estimates as the reference standard. These rates have been constructed to represent the incidence of revision in the overall UK population and do not reflect the risk of revision for those already having undergone hip or knee replacement. The population estimates used for standardisation were as published by the ONS,¹⁰⁶ the General Register Office for Scotland and the Northern Ireland Statistics and Research Agency. We computed 95% CIs using a Poisson model appropriate for directly standardised rates.

Mean age at revision was calculated for hips and knees for each calendar year and 95% CIs computed. The distribution of age at revision was calculated for three consecutive 5-year periods, separately for hip and knee by sex, to investigate patterns over time.

A total of 1689 hip revisions and 634 knee revisions were identified in the CPRD database. During the period 1991–2006, women underwent 59% more hip revisions and 6% more knee revisions than men. Women were, on average, > 3 years older than men at hip revision and approximately 2.5 years older in the case of knee revisions. Since 1994, the female-to-male ratio among patients undergoing revision surgery has remained reasonably stable, with ratios for hips varying around 2 : 1 and for knees 1.4 : 1, with further variation by age group, showing higher ratios for 70- to 79-year-olds (*Figures 4 and 5*).

Patients undergoing TKR had a significantly higher BMI than those undergoing THR ($p < 0.0001$) and the difference in BMI was greater for women than for men (*Table 1*).

Between 1991 and 2006, the estimated age-standardised rates of hip revision arthroplasty increased from 2.3 per 100,000 person-years (95% CI 1.2 to 3.8 per 100,000 person-years) to 7.7 per 100,000 person-years (95% CI 6.2 to 9.2 per 100,000 person-years) for women and from 1.3 per 100,000 person-years (95% CI 0.5 to 2.5 per 100,000 person-years) to 6.3 per 100,000 person-years (95% CI 5.0 to 7.8 per 100,000 person-years) for men. The majority of the increase occurred between 1991 and 1994, with rates stabilising between 1994 and 2006. When the rates of revision hip replacement in 2006 were applied to the mid-2006 population estimates for the UK, we obtained an estimated total number of primary THRs (excluding private practice) of 1887 (95% CI 1538 to 2270) for women and 1447 (95% CI 1148 to 1780) for men.

Over the same period, the estimated age-standardised rates of knee revision arthroplasty increased from 0.9 per 100,000 person-years (95% CI 0.3 to 1.9 per 100,000 person-years) to 5.0 per 100,000 person-years (95% CI 3.9 to 6.3 per 100,000 person-years) for women and from 0.2 per 100,000 person-years (95% CI 0.0 to 3.1 per 100,000 person-years) to 4.1 per 100,000 person-years (95% CI 3.1 to 5.3 per 100,000 person-years) for men. The temporal trend in rates of knee revision shows a marked increase, with a steep rise after 1995. Estimated rates for women increased almost fivefold between 1996 and 2006. When we apply the 2006 rates for knee replacement to the mid-2006 UK population estimates, we obtain an estimated total number of revision TKRs (excluding private practice) of 1225 (95% CI 946 to 1540) for women and 942 (95% CI 706 to 1211) for men.

In 2006, the mean age at operation for hip revisions was 72.7 years (95% CI 70.3 to 75.0 years) for women and 69.5 years (95% CI 67.3 to 71.7 years) for men, and for knee revisions it was 71.0 years

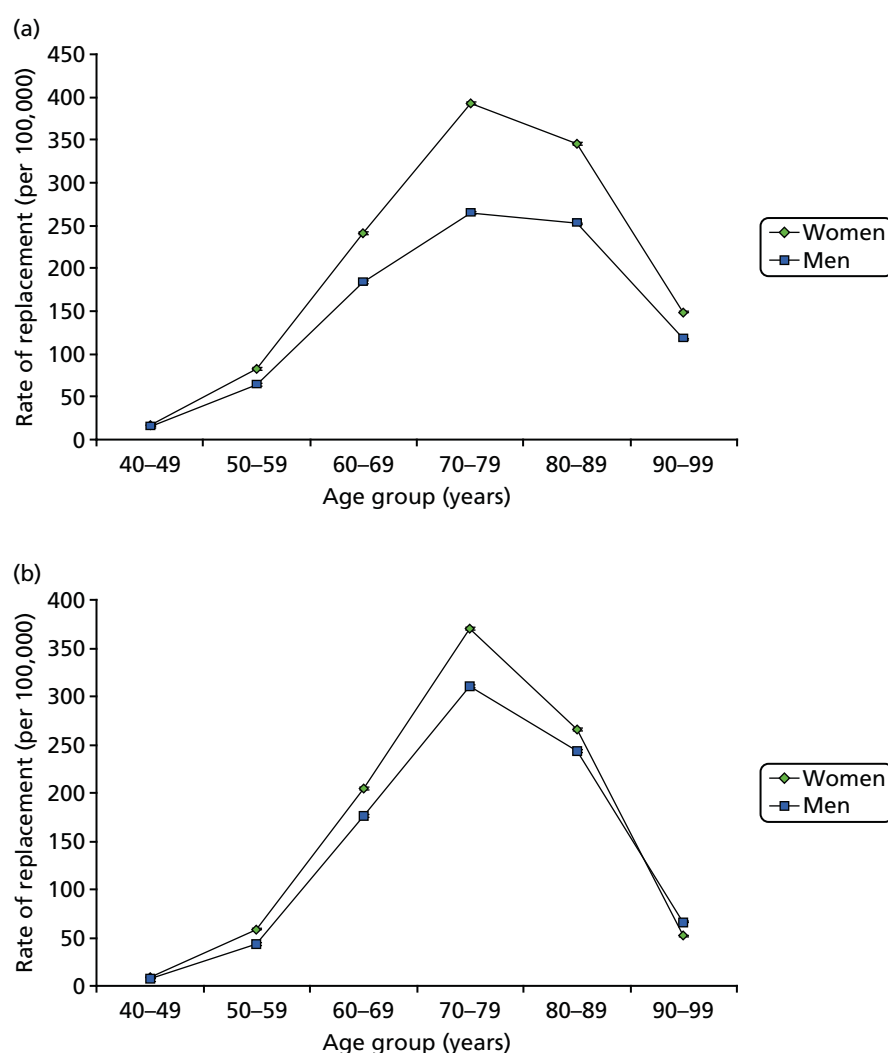


FIGURE 4 Revision rates for (a) hip; and (b) knee. Age profile of replacement rates by sex.

(95% CI 68.8 to 73.2 years) for women and 67.8 years (95% CI 65.3 to 70.3 years) for men. Among women, the highest incidence rate of hip revision is in the 80–89 years age group and of knee revision is in the 70–79 years age group. The rates in these groups are 35.1 (95% CI 22.1 to 48.1) for hips and 19.9 (95% CI 12.1 to 27.8) for knees. The number of replacements for those aged 60–79 years comprises almost two-thirds of the total for knees (64.7%) and a similar proportion for hips (61.0%), with men having a higher proportion than women in this age group for both hips and knees.

The mean age at hip revision was higher in women than in men for almost all years after 1991 (*Figure 6a*), but the difference was statistically significant in only 2 of the years. For knee revision (*Figure 6b*), the sex difference in mean age at operation is much narrower than for hip revision. Since 1999, the sex-specific mean ages at knee revision have been very similar, with women slightly older than men, but by 2006 there is virtually no discernible difference between the sexes.

To explore the possibility that there had been a change in the distribution in age of people undergoing revision surgery, we examined the distribution of age in 10-year age bands over three time periods: 1991–5, 1996–2000 and 2001–5. For the two earlier periods the counts of revision operations were generally too low to enable an effective comparison between age distributions over time. However, in the period 2001–5 it was observed that the distributions were similar between the sexes and also between hip

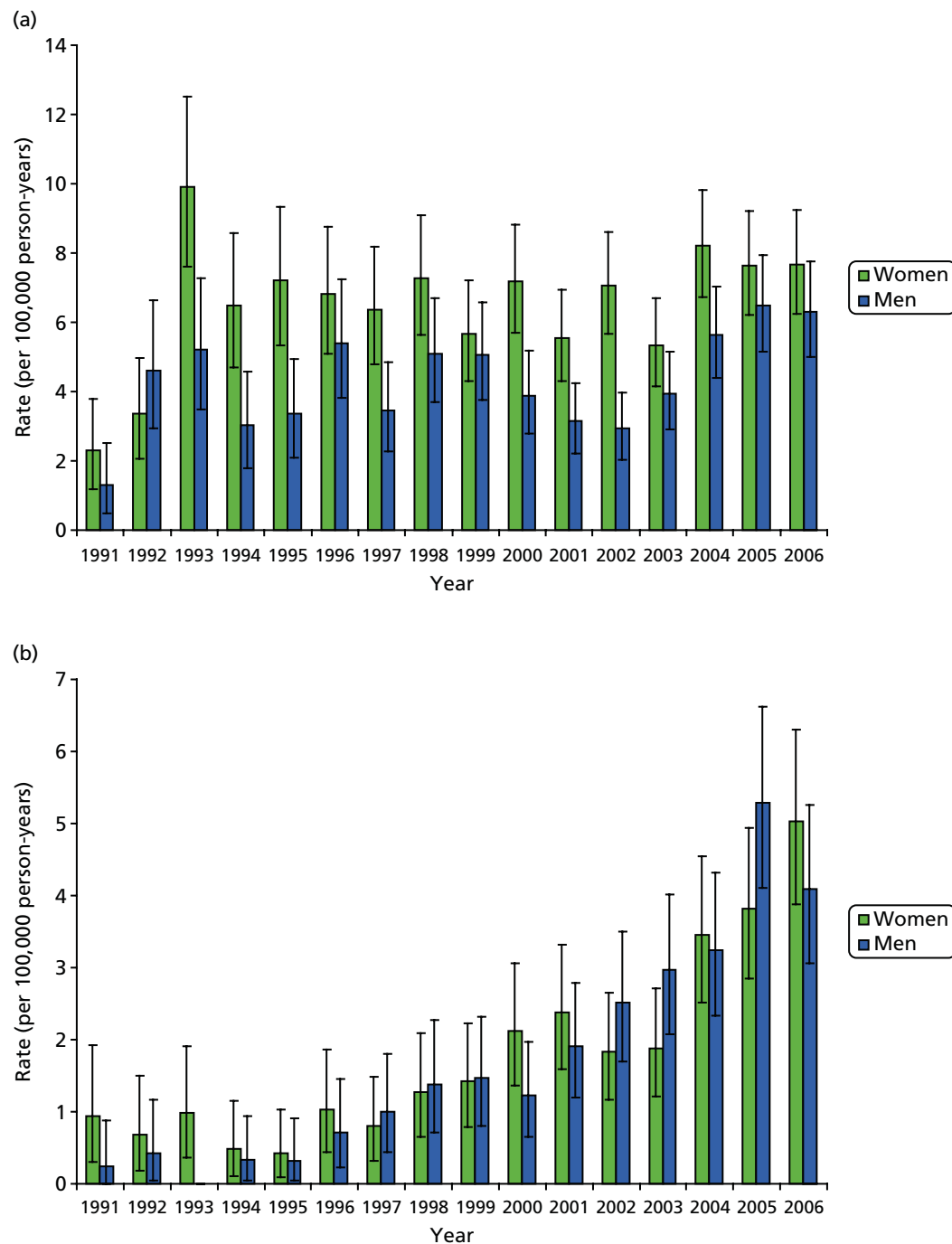


FIGURE 5 Trends in revision rates with 95% CIs 1991 to 2006: (a) hip; and (b) knee.

and knee revisions. During the same period, the 10-year age group at which most revisions were carried out was 70–79 years (between 37% and 39% of revisions, whether for hips or knees for either sex).

The ratio of knee-to-hip revision incidence rates (*Figure 7*) was low for both men and women during the mid-1990s, at around 0.15 : 1. This ratio then began to rise steeply in both sexes in 1996 such that, by 2006, the incidence of knee revision was two-thirds of that of hip revision.

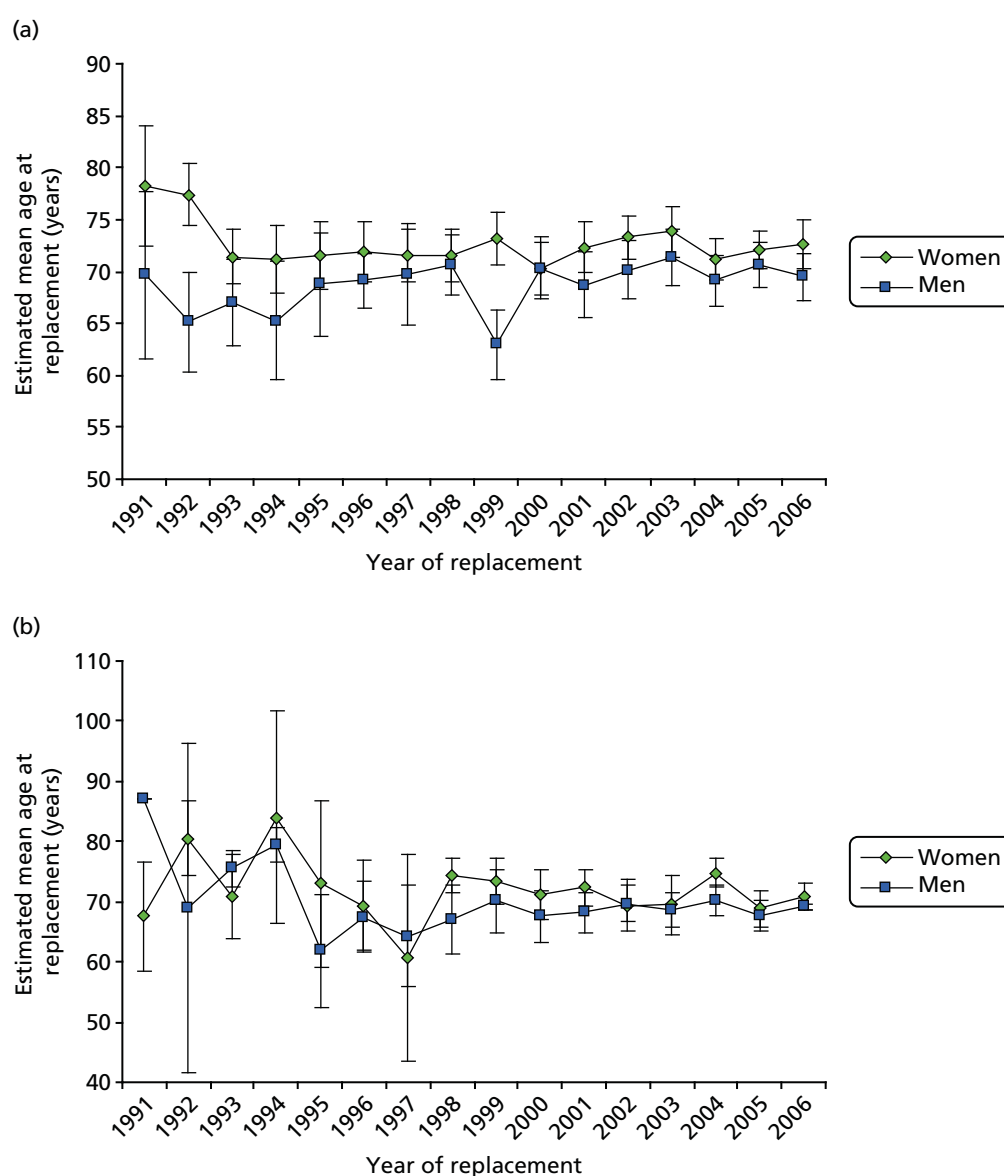
When we compared our estimated revision incidence rates for 2006 with the corresponding rates for primary operations using the same GPRD data set, we found that the ratio of primary operations to revisions was approximately 17 : 1 for hips and 25 : 1 for knees.

TABLE 1 Baseline clinical variables for THR and TKR

Variable	Surgery		TKR	
	THR			
	Female (n = 17,560)	Male (n = 10,508)	Female (n = 14,462)	Male (n = 9902)
Age (years), mean (range)	70.4 (18–103)	67.5 (19–100)	70.9 (18–99)	69.4 (19–98)
BMI (kg/m ²), mean (IQR)	27.0 (23.3–30.1)	27.5 (24.6–29.9)	29.1 (25.2–32.5)	28.4 (25.4–30.9)
Smoker (%)	11.3	14.4	8.1	12.3
Deprivation (% from practices in the most deprived quintile)	17.3	17.1	18.8	19.8

IQR, interquartile range.

Reproduced from oral presentation by David Culliford, Faculty of Health Sciences, University of Southampton, at the British Society for Rheumatology in 2008. Copyright David Culliford.


FIGURE 6 Mean age at revision, with 95% CIs, in 1991–2006: (a) hip; and (b) knee.

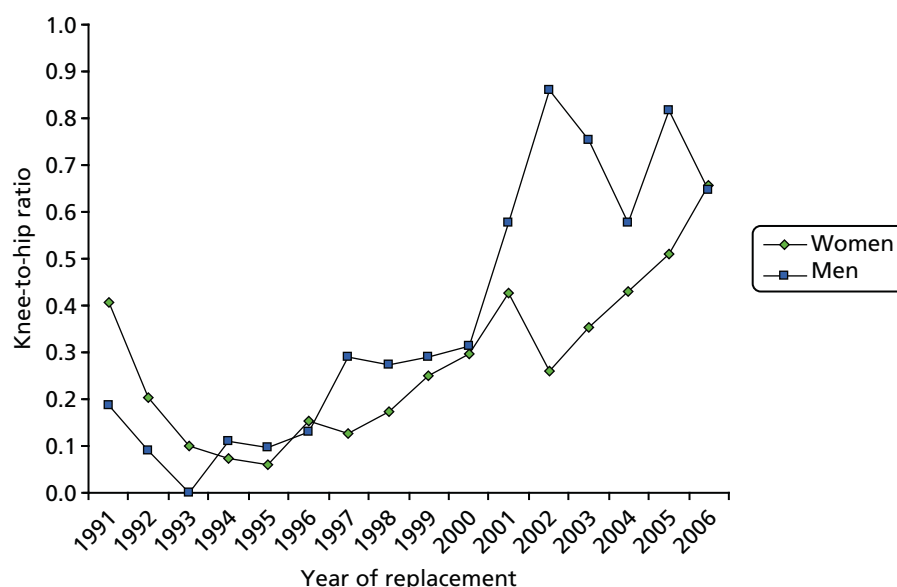


FIGURE 7 Ratio of incidence rates for knee to hip revisions for men and women between 1991 and 2006.

Analysis of the UK national database over a 10-year period revealed that knee revision rates increased more than fivefold over this period in both men and women. The rate of THR revision has remained relatively static, with no significant increase identified between 1994 and 2006.

The fivefold increase in knee revision rates may be multifactorial and a reflection of the increased number of primary replacements over this 10-year period as well as rapid advances in implant technology alongside improved surgical experience.⁶⁶ Clinicians may be much more likely to intervene at an earlier stage, especially in complex cases that in the past would have been considered beyond salvage. Another contributing factor may be the increased usage of unicompartmental knee arthroplasty, which has been associated with higher revision rates.

The picture looks different for hip replacements, as the results showed a marginal increase in revision rates. Technological advances, such as improved bearing surfaces and fixation methods, should have decreased the need for revision but may not have had the predicted impact on primary total hip arthroplasty (THA) longevity. Registry data from Scandinavia^{120,121} demonstrate that the longevity of more conventional cemented implants is superior to that of modern cementless or resurfacing designs. In 1996 there was good evidence to support THA in < 30% of primary cases in the UK; in 2010 there was good evidence for their use in < 40% of cases. This is an important observation, as newer implants tend to be more expensive and may in fact be adding to the revision burden if they are not introduced in a co-ordinated manner. Significant demographical differences were found, with women 59% more likely to require hip revision than men.

In comparison with our findings, Kurtz *et al.*⁶⁶ have previously reported a 79% increase in revision THA and a 200% increase in revision total knee arthroplasty (TKA) in the USA between 1990 and 2002.⁶⁶ This group also looked at future projections between 2006 and 2030 and estimated that revision THA would increase by 137% and revision TKA by 601% by 2030. These findings confirm the predicted trends in revision arthroplasty in the UK, with dramatic increases in knee revisions and a smaller, but still significant, increase in hip revisions. The cost implications for this increase would be significant, and accurate modelling of revision THA and TKA demand is therefore required for adequate and appropriate long-term health planning.

We have further investigated the role of a risk factor, particularly the role of BMI, on the time of revision for hip and knee arthroplasties. We used methods from survival analysis to present population-based estimates for the risk of revision following TJR of the hip and knee. We described these associations and published the results.⁹⁷

Association of body mass index with time of revision

A population-based survival analysis describing the association of body mass index on time to revision for total hip and knee replacements: results from the UK General Practice Research Database⁹⁷

For this task we selected 63,162 THR and 54,276 TKRs from the CPRD database. The average age at replacement was similar in both groups and the proportion for women in both procedures was greater (Table 2).

Table 2 also describes the baseline characteristics of the cohort, including summary statistics and missing data percentages for all explanatory variables for which complete data were not observed.

Eighty per cent of preoperative BMI values used were recorded within 5 years of the primary operation; among those with a recorded BMI, the proportion of obese patients ($\text{BMI} \geq 30 \text{ kg/m}^2$) was 26.2% for THR and 39.8% for TKR, and of morbidly obese patients ($\text{BMI} \geq 40 \text{ kg/m}^2$) was 1.6% for THR and 3.6% for TKR.

In a single-predictor (univariable) survival model allowing for the competing risk of death over the entire period of follow-up, we estimated that THR participants had a 3% increase in the subhazard of revision [subhazard ratio (SHR) 1.030, 95% CI 1.020 to 1.041; $p < 0.001$] for each extra unit (kg/m^2) of BMI, with TKR participants showing a 2.6% increase per unit (SHR 1.026, 95% CI 1.013 to 1.038; $p < 0.001$). The SHR was significantly greater for men than for women for both THR (SHR 1.35, 95% CI 1.23 to 1.48; $p < 0.001$) and TKR 2 (SHR 1.54, 95% CI 1.37 to 1.72; $p < 0.001$).

Age at TJR was also a significant univariable predictor of both hip and knee revision surgery, with THR participants estimated to have a 3% reduction (SHR 0.970, 95% CI 0.967 to 0.973; $p < 0.001$) for each extra year of age, and TKR participants showing a 4.3% reduction (SHR 0.957, 95% CI 0.952 to 0.961; $p < 0.001$). The effects for all three variables (sex, age and BMI) were then estimated in multivariable

TABLE 2 Clinical and demographic characteristics, all participants undergoing hip and knee replacement

Characteristic	Surgery			
	THR (N = 63,162)		TKR (N = 54,276)	
	Female (n = 39,292)	Male (n = 23,870)	Female (n = 31,682)	Male (n = 22,594)
Age (years), mean (SD)	70.5 (11.1)	67.7 (11.0)	70.7 (9.6)	69.4 (9.4)
Sex (%)	62.2	37.8	58.3	41.6
BMI (kg/m^2), mean (SD)	27.2 (5.1)	27.7 (4.3)	29.6 (5.6)	28.8 (4.4)
Missing BMI (%)	19.1	19.3	13.8	14.0
Revisions, n (%)	1000 (2.55)	811 (3.40)	572 (1.8)	614 (2.7)
Deaths pre revision, n (%)	6615 (16.8)	4201 (17.6)	4110 (13.0)	3349 (14.8)
Number of comorbid conditions (%)				
0	42.8	48.1	37.5	43.7
1	34.2	31.0	37.4	35.8
≥ 2	23.0	20.9	25.2	20.6

Note

Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) licence, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

competing risks regression models after adjusting for smoking status, drinking status and the number of comorbid conditions over the entire period of follow-up. For age, the estimates were almost exactly the same as those from the univariable model for both hip and knee, but for sex (SHR 1.23 for hip and 1.51 for knee) and BMI (SHR 1.020 for hip and 1.015 for knee) the estimates were smaller. Nevertheless, all three variables remained statistically significant for both hip and knee in the presence of adjustment.

For a 5-kg/m² and 10-kg/m² increase in BMI, this represents an increase in THR revision risk of 10.4% and 21.9%, respectively (7.7% and 16.1% for TKR). Testing for two-way interactions between age, sex and BMI did not produce any significant effects. All subhazard estimates (with 95% CI and *p*-values) from the univariable and multivariable models are given in *Tables 3* and *4*.

To further explore the effect of estimates for BMI, we ran the same adjusted age–sex–BMI model described but used categorical rather than continuous BMI. For morbidly obese TKR participants (BMI > 40 kg/m²) there was a 43.9% increase (95% CI 2.6% to 103.9%; *p* = 0.040) in the rate than those with a normal BMI (18.5–25 kg/m²), but the effect for THR was larger (an increase of 65.5%) and stronger (95% CI 15.4% to 137.3%; *p* = 0.006).

The effect sizes were similar to those obtained when using the adjusted SHR estimate of continuous BMI for a participant with a BMI of 45 kg/m² relative to one with a BMI of 22 kg/m² (an increase of 57.7% for THR and 40.8% for TKR). For obese patients in the range 30–40 kg/m² compared with those with a normal BMI, the estimated SHR for revision was weakly significant for THR (15.7% increase, 95% CI 0.2% to 33.7%; *p* = 0.048) but not for TKR (17.9% increase, 95% CI –1.9% to 41.6%; *p* = 0.079). As a sensitivity analysis, we also performed standard Cox regressions with revision surgery as the event of interest, and when no distinction was made between death and other censoring events. Univariable models for age, sex and BMI gave very similar results to the competing risks analysis, as did the multivariable models that adjusted for the same factors as in the competing risks regression. Results from the Cox regression models are given in *Tables 5* and *6*.

In addition, we also calculated that it would take 175 patients with TKR to reduce their baseline BMI from obese to normal in order to prevent one revision operation after 5 years. For patients with THR this number reduces to 152.

TABLE 3 Estimated subhazard of revision for total hip and knee replacements: competing risks analysis – hip

Variable	Analysis					
	Univariable			Adjusted ^a		
	HR	95% CI	<i>p</i> -value	HR	95% CI	<i>p</i> -value
BMI (kg/m ²) (per additional unit) ^b	1.030	1.020 to 1.041	< 0.001	1.020	1.009 to 1.032	< 0.001
Sex						
Female (reference)	1.00			1.00		
Male	1.35	1.23 to 1.48	< 0.001	1.23	1.10 to 1.38	< 0.001
Age (years at THR) (per additional year)	0.970	0.967 to 0.973	< 0.001	0.971	0.966 to 0.975	< 0.001

a Adjusted for smoking (yes/no/ex), drinking (yes/no/ex) and number of comorbid conditions.

b BMI available in 86.1% of patients.

Note

Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) licence, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

TABLE 4 Estimated subhazard of revision for total hip and knee replacements: competing risks analysis – knee

Variable	Analysis					
	Univariable			Adjusted ^a		
	HR	95% CI	p-value	HR	95% CI	p-value
BMI (kg/m ²) (per additional unit) ^b	1.026	1.013 to 1.038	< 0.001	1.015	1.002 to 1.028	0.023
Sex						
Female (reference)	1.00			1.00		
Male	1.54	1.37 to 1.72	< 0.001	1.51	1.32 to 1.73	< 0.001
Age (years at THR) (per additional year)	0.957	0.952 to 0.961	< 0.001	0.957	0.951 to 0.962	< 0.001

a Adjusted for smoking (yes/no/ex), drinking (yes/no/ex) and number of comorbid conditions.

b BMI available in 80.9% of patients.

Note

Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) licence, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

TABLE 5 Estimated hazard of revision for THR: univariable and adjusted Cox regression analysis with death as a censoring event

Variable	Analysis					
	Univariable			Adjusted ^a		
	HR	95% CI	p-value	HR	95% CI	p-value
BMI (kg/m ²) (per additional unit) ^b	1.029	1.017 to 1.040	< 0.001	1.019	1.008 to 1.031	0.001
Sex						
Female (reference)	1.00			1.00		
Male	1.36	1.24 to 1.49	< 0.001	1.26	1.13 to 1.41	< 0.001
Age (years at THR) (per additional year)	0.978	0.974 to 0.983	< 0.001	0.977	0.972 to 0.982	< 0.001

a Adjusted for smoking (yes/no/ex), drinking (yes/no/ex) and number of comorbid conditions.

b BMI available in 86.1% of patients.

Note

Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) licence, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

Finally, we assessed whether or not the higher incidence of hip revision surgery during the first year following THR might compromise the proportionality assumption and, therefore, suggest the inclusion of time-dependent effects. Separate univariable piecewise competing risk models for hip revision were fitted for sex, age (≤ 65 years vs. > 65 years) and BMI (> 40 kg/m² vs. ≤ 40 kg/m²). A single change point at 1 year was used to simultaneously estimate two SHRs for revision (before and after 1 year following THR).

The only model that provided some evidence for a different SHR during the first year was with BMI (> 40 kg/m² vs. ≤ 40 kg/m²) as the predictor (SHR 2.619, 95% CI 1.502 to 4.560; $p = 0.001$), but this was not matched with a statistically significant estimate for revision after the first year (SHR 0.575, 95% CI 0.238 to 1.170; $p = 0.130$).

TABLE 6 Estimated hazard of revision for TKR: univariable and adjusted Cox regression analysis with death as a censoring event

Variable	Analysis					
	Univariable			Adjusted ^a		
	HR	95% CI	p-value	HR	95% CI	p-value
BMI (kg/m ²) (per additional unit) ^b	1.024	1.012 to 1.037	< 0.001	1.015	1.003 to 1.028	0.019
Sex						
Female (reference)	1.00			1.00		< 0.001
Male	1.58	1.41 to 1.77	< 0.001	1.55	1.36 to 1.77	
Age (years at THR) (per additional year)	0.962	0.956 to 0.967	< 0.001	0.961	0.955 to 0.968	< 0.001

a Adjusted for smoking (yes/no/ex), drinking (yes/no/ex) and number of comorbid conditions.

b BMI available in 80.9% of patients.

Note

Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) licence, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

Cumulative incidence rates of revision were higher for men than for women and higher for hips than for knees. Age, sex and BMI were estimated to be significant predictors of time to revision in an adjusted model allowing for the competing risk of death. Severely obese patients undergoing THR had a higher risk of revision surgery during the first year following replacement, but the same effect was not observed for knee replacement.

Projected future trends for total hip replacement/total knee replacement accounting for projected changes in age and obesity

Estimating the lifetime risk of total knee and hip arthroplasties

The lifetime risk of total hip and knee arthroplasty: results from the UK General Practice Research Database⁹⁹

Lifetime risk is a patient-centred measure of risk for the onset of disease or the occurrence of specific events. The concept is easily understood by clinicians and policy-makers, and can be made even more informative by also calculating interval risks (e.g. 10 years) at different ages to establish the periods of greatest lifetime risk. Population-based estimates are needed for effective and efficient health-care planning and resource allocation. No lifetime risk estimates were available in the literature for patients who were undergoing these surgical procedures, but published incidence rates existed for hip and knee replacement in the UK^{67,116} and internationally.^{119,123,124}

The primary aim of this analysis was to use the CPRD database combined with the ONS mortality data to provide estimates for the lifetime risk of undergoing a primary THR or TKR in the UK. OXMIS/Read codes were used to identify THRs and TKRs for the period 1991–2006 in the CPRD database. Patients were included if aged ≥ 50 years at the time of replacement. Sex-specific all-cause mortality data from the ONS were obtained for the same period.¹²⁵

The analysis was done with CPRD data that were aggregated into single-year age intervals, with the age label defined as age at last birthday at the end of a calendar year, starting at the age of 50 years. We used data for the time period 1991–2006 and identified 49,105 patients who had undergone a THR (*n* = 25,845) or TKR (*n* = 23,260). Consistent definitions were applied to death data and exposure to risk. Incidence rates for joint replacement were computed by dividing the count of primary THRs and TKRs in the CPRD data by

the corresponding amount of person-time spent by those in the entire CPRD population who matched the age band, sex and time interval of interest. This was achieved by a life table method similar to that described by Kim *et al.*³⁹ CIs at the 95% level were estimated under a Poisson model.¹²⁶ Risks were estimated separately for sex and hip/knee. This was repeated with 60, 70 and 80 years of age as the starting point for the risk of replacement. We further computed 10-year risk percentages from age 50 years up to the age of 80 years. All estimates for single calendar years used mortality data matched to the same calendar years, but for the estimates based on the entire study period we used 2006 mortality rates with a sensitivity analysis. Lifetime risks of THR and TKR, stratified by sex for individual calendar years, were estimated in order to compare temporal trends.

The results, using rates from 2005, showed that the estimated mortality-adjusted lifetime risk of THR at age 50 years was 11.6% for women and 7.1% for men. For the aggregated data over the period 1991–2006, the mortality-adjusted lifetime risk of THR at age 50 years was estimated at 8.3% for women and 5.2% for men. The lifetime risk of THR at age 50 years rose from 4.0% (95% CI 3.0% to 5.0%) to 11.1% (95% CI 9.9% to 12.2%) for women and from 2.2% (95% CI 1.4% to 3.0%) to 6.6% (95% CI 5.7% to 7.5%) for men. Therefore, our findings estimated that, between 1991 and 2006, the lifetime risk of THR at age 50 years rose from 4.0% (95% CI 3.0% to 5.0%) to 11.1% (95% CI 9.9% to 12.2%) for women and from 2.2% (95% CI 1.4% to 3.0%) to 6.6% (95% CI 5.7% to 7.5%) for men.

Again, using the rates from 2005, we estimated that the mortality-adjusted lifetime risk of TKR at age 50 years was 10.8% for women and 8.1% for men. The aggregated data for the period 1991–2006 estimated the mortality-adjusted lifetime risk for TKR at age 50 years at 7.0% for women and at 5.2% for men. The same time period for TKR saw an increased risk for women from 2.9% (95% CI 2.1% to 3.8%) to 10.6% (95% CI 9.5% to 11.7%) and for men from 1.8% (95% CI 1.1% to 2.6%) to 7.7% (95% CI 6.8% to 8.7%). As with hips, TKR estimates of risk also increased, for women from 2.9% (95% CI 2.1% to 3.8%) to 10.6% (95% CI 9.5% to 11.7%) and for men from 1.8% (95% CI 1.1% to 2.6%) to 7.7% (95% CI 6.8% to 8.7%).

As a sensitivity analysis these estimates were recalculated using 1991 mortality data, but this resulted in only small reductions in the lifetime risk estimates of between 0.6 and 0.8 percentage points at age 50 years and of 0.2 and 0.3 percentage points at age 80 years. These reductions were seen for both THR and TKR, and for men and women.

The lifetime risk decreases with increasing age for both THR and TKR in men and women. At age 80 years, the sex gap in risk of THR reduced to 40% higher for women than for men (22% higher for TKR). Estimated risk percentages at ages 50, 60, 70 and 80 years are presented in *Table 7*.

The sex gaps in the estimates obtained for the whole study period were similar to those for the 2005 estimates.

Our results showed that between 1991 and 2006 the lifetime risk of THR at age 50 years increased from 4.0% (95% CI 3.0% to 5.0%) to 11.1% (95% CI 9.9% to 12.2%) for women and from 2.2% (95% CI

TABLE 7 Estimated lifetime risk (95% CI) of undergoing primary TKR or THR based on age- and sex-specific incidence rates adjusted for mortality. Data from GPRD in 2005⁹⁹

Current age (years)	Risk of primary TKR, % (95% CI)		Risk of primary THR, % (95% CI)	
	Female	Male	Female	Male
50	10.8 (9.7 to 11.9)	8.1 (7.1 to 9.1)	11.6 (10.4 to 12.7)	7.1 (6.2 to 8.0)
60	10.1 (9.0 to 11.2)	7.9 (6.9 to 8.9)	10.8 (9.7 to 12.0)	6.7 (5.8 to 7.7)
70	7.8 (6.7 to 8.8)	6.2 (5.2 to 7.2)	8.1 (7.1 to 9.2)	5.3 (4.3 to 6.2)
80	3.3 (2.6 to 4.1)	2.7 (1.8 to 3.6)	3.8 (3.0 to 4.7)	2.7 (1.8 to 3.6)

1.4% to 3.0%) to 6.6% (95% CI 5.7% to 7.5%) for men. For TKR, the risk increased for women from 2.9% (95% CI 2.1% to 3.8%) to 10.6% (95% CI 9.5% to 11.7%) and for men from 1.8% (95% CI 1.1% to 2.6%) to 7.7% (95% CI 6.8% to 8.7%) (Figure 8).

The lifetime risks of hip and knee replacements are estimated to be between 5% and 10%, which is substantially below the estimated lifetime risk of hip and knee OA. Our estimates based on UK GPRD data from 2005 suggest a lifetime risk of THR and TKR for women or men aged 50 years living in the UK of 10–11% and 6–7%, respectively.

Future projections of total hip and knee arthroplasties

Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink⁹⁸

Future predictions for lower limb arthroplasty in the UK are limited,^{116,123} and international predictions are mainly concentrated around the USA and Europe.^{34,68,127,128} With the steady increase in rates of hip and knee surgery, up-to-date future predicted rates are necessary as part of our understanding of this treatment intervention. The most recent published future projections of the UK covered England only,¹¹⁶ were based on a 10-year period of HES data and did not account for BMI changes or other important risk factors for arthroplasty.

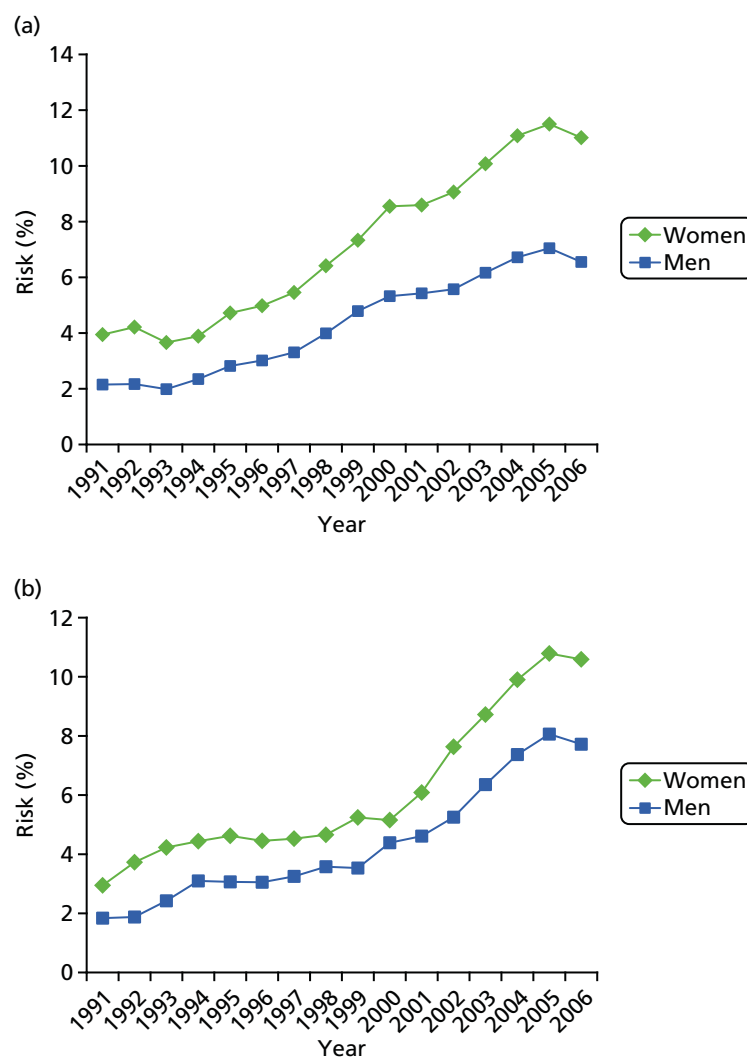


FIGURE 8 Estimated lifetime risk at age 50 years of undergoing (a) THR; or (b) TKR based on age- and sex-specific incidence adjusted for mortality. Data from the GPRD.⁹⁹

Deciding on the correct method for forecasting is important and dependent on high-quality research data. More sophisticated modelling approaches require at least one population-based cohort or survey data set with long-term follow-up.¹²⁹

We produced and published the national future projections for THR and TKR.⁹⁸ We used three national data sets that were representative of age, sex and BMI in the UK population. Using the CPRD database, in combination with national population forecasts from the ONS, we aimed to calculate age- and sex-specific forecasts for the number of THR and TKR operations per year in the UK between 2010 and 2035. Secondary analysis aimed to produce forecasts that reflect the changing distribution of BMI during the same period. To project estimated THR and TKR rates, HSE data were used. We constructed a denominator to estimate BMI-specific rates and obtained sex-specific population projections from the ONS for the period 2011–35.¹⁰⁶ The methods of Kurtz *et al.*⁶⁶ were further extended to incorporate the inclusion of BMI, in addition to age and sex.

Analysis: estimation

The CPRD data (1991–2010) were used to estimate annual incidence rates for THR/TKRs, and standard log-linear regression models were used to produce calendar year-, age- and sex-specific rates, but were extended to include BMI-specific rates. Unweighted aggregated data from the HSE, for the same period, were used as a proxy for the change in the distribution of BMI in the UK population. The CPRD data were remodelled by calendar year, age, sex and BMI. Age and BMI were grouped in categories, and rates for hips and knees were estimated separately. The calendar year-/age-/sex-specific values of BMI in the HSE were used to partition the calendar year-/age-/sex-specific denominator values in the CPRD to further break them down by BMI. Regarding the numerator for the rate (i.e. the counts of TJRs), the counts were weighted by BMI for those TJR patients with an observed preoperative BMI in their record. This was the case for approximately 80%. We made the decision not to use missing data methods (such as multiple imputation) because of a high rate of observed BMI. If preoperative BMI had been available in, for example, 50–60% of cases, we could have reconsidered and use multiple imputation methods. Our BMI-specific projections were in categorical bandings; therefore, fewer concerns were raised about sensitivity to missingness.

The ONS data were split into age- and sex-specific forecasts, by BMI group, prior to applying the estimated incidence rates obtained from the HSE. Two methods were used: BMI proportions fixed at 2010 levels and BMI proportions increasing linearly based on ordinary least squares (OLS) regression estimates derived from the HSE BMI data from 1991 to 2010. A hyperbolic tangent function, similar to the method described in the Foresight report, was used to smooth the proportions over the forecasting time frame.¹³⁰

Analysis projection

Two different projections methods were used on each of the two future UK population scenarios. Hips and knees were analysed separately. The first method used THR/TKR incidence rate estimates held at 2010 levels, applied to the two population scenarios. The second used an exponential extrapolation directly from the log-linear model-estimated rates for THR/TKR. The two population forecast data sets⁷⁶ contained exactly the same population growth estimates by age and sex over time, as forecast by ONS, with a difference that one population data set assumed a static BMI distribution (held fixed at 2010), whereas the other reflected HSE- and CPRD-based estimates of forecast BMI distribution change in the UK.

The results from analysis of the CPRD database contained 50,000 THRs and 45,609 TKRs between 1991 and 2010, and all sets included age, sex and BMI. The average age at time of operation was similar for THR and TKR, and the proportion of women was greater for both TKR and THR (*Table 8*).

Preoperative BMI was slightly higher for TKR than for THR. There was little sex-specific difference in counts when comparing fixed or varying future estimates of BMI category distribution in hips. Knee estimates, however, suggested a 9% higher rate when using the varying BMI distribution.

TABLE 8 Demographic characteristics of CPRD subjects used to construct incidence rates⁹⁸

Variable	Surgery			
	TKR (N = 45,609)		TKR (N = 50,000)	
	Female (n = 26,623)	Male (n = 18,986)	Female (n = 31,148)	Male (n = 18,852)
Sex (%)	58.4	41.6	62.2	37.8
Age (years), mean (SD)	70.3 (9.5)	69.4 (9.2)	69.9 (10.9)	67.8 (10.7)
BMI (kg/m ²), mean (SD)	29.6 (5.4)	28.8 (4.4)	27.2 (5.1)	27.7 (4.2)

Reprinted from *Osteoarthritis and Cartilage*, vol. 23, Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, COASt Study Group, Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink, pp. 594–600, 2015,⁹⁸ with permission from Elsevier.

Hospital Episode Statistics data (1991–2010) were used to estimate future BMI distribution and contained 186,174 subjects with measured BMI. The breakdown of this distribution by age, sex and BMI values is depicted in *Tables 9* and *10*.

TABLE 9 Health Survey for England 1991 to 2010 showing number of female and male subjects by BMI⁹⁸

BMI group (kg/m ²)	Total number of subjects (N = 186,174)			
	Female [n = 100,576 (54.0%)]		Male [n = 85,598 (46.0%)]	
	n	%	n	%
< 20	6117	6.1	2933	3.4
20 to 25	39,261	39.0	27,347	31.9
25 to 29	33,361	33.2	38,681	45.2
30 to 39	19,688	19.6	16,216	18.9
≥ 40	2149	2.1	421	0.5

Reprinted from *Osteoarthritis and Cartilage*, vol. 23, Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, COASt Study Group, Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink, pp. 594–600, 2015,⁹⁸ with permission from Elsevier.

TABLE 10 Health Survey for England 1991 to 2010 showing number of female and male subjects by age⁹⁸

Age group (years)	Total number of subjects (N = 186,174)			
	Female [n = 100,576 (54.0%)]		Male [n = 85,598 (46.0%)]	
	n	%	n	%
18–39	37,664	37.4	32,527	38.0
40–49	18,503	18.4	15,704	18.3
50–59	15,620	15.5	13,640	15.9
60–69	13,813	13.7	12,433	14.5
70–79	10,430	10.4	8504	9.9
≥ 80	4546	4.5	2790	3.3

Reprinted from *Osteoarthritis and Cartilage*, vol. 23, Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, COASt Study Group, Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink, pp. 594–600, 2015,⁹⁸ with permission from Elsevier.

The static rate projection method, with BMI distribution held fixed at levels estimated for 2010, forecasts an annual number of THRs of up to 97,516 and of TKRs of up to 110,306 by 2035. Using the same projection method, but with changing BMI distribution, the estimated rates are expected to grow by up to 95,877 for hips and 118,666 for knees.

Using the log-linear projection method, with BMI distribution held fixed at levels estimated for 2010, the annual number forecast for hips is up to 437,708 and for knees is up to 1,071,790 by 2035. Using the same method, but with changing BMI distribution, the rates are exponential and increase to 439,097 for hip and 1,219,362 for knees by 2035. Five-yearly projections for all four scenarios up to 2035 are shown in *Table 11*.

The results that follow present counts split by sex, BMI and age, all of which are estimated using the static projection method (see *Table 11*).

Hip and knee projected counts by sex are shown in *Table 12*. There is little sex difference in counts at 2035 when we compare projections, with fixed or varying future estimates, of BMI category distribution for hips. Knees results are different, however, especially for women, whose TKR count at 2035 is estimated to be 9% higher when using varying BMI distribution as opposed to fixed.

Discussion

The increasing trends in THR and TKR up to the year 2000 have continued and are more marked in knees than in hips. Although there is a marked increase in the number of knee replacements being carried out per year, the number of TKRs are similar to those for THR. The increase in knee surgery may be because the burden of OA of the knee is more easily identified in radiographs.¹³¹ The number of TKRs per year is similar to the number of THRs, despite the much higher prevalence of OA of the knee.^{75,76} It is possible that the level of provision of THR is appropriate to the burden of OA of the hip, whereas the level for the TKR is still below that required by surgeons operating on patients with lower levels of pain and disability.

TABLE 11 Projected UK counts for total hip and knee replacements in adults to the year 2035⁹⁸

Year	Projection							
	THR incidence rates				TKR incidence rates			
	Estimated rates fixed at the 2010 level		Estimated rates increasing log-linearly		Estimated rates fixed at the 2010 level		Estimated rates increasing log-linearly	
	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time
2015	72,762	72,418	96,314	95,945	82,610	85,019	128,944	133,063
2020	79,716	79,048	141,626	140,945	90,555	94,783	221,653	234,244
2025	85,988	85,026	205,464	204,793	97,780	103,657	376,384	407,400
2030	91,496	90,202	296,354	296,106	103,810	111,015	632,257	701,052
2035	97,516	95,877	437,708	439,097	110,306	118,666	1,071,790	1,219,362

Reprinted from *Osteoarthritis and Cartilage*, vol. 23, Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, COAST Study Group, Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink, pp. 594–600, 2015,⁹⁸ with permission from Elsevier.

TABLE 12 Projected UK counts for total hip and knee replacements respectively by sex to the year 2035, with estimated TKR incidence rates fixed at 2010 level⁹⁸

Year	Surgery							
	THR				TKR			
	Women		Men		Women		Men	
	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time	BMI category proportions fixed at 2010 estimates	BMI category proportions changing over time
2015	45,143	44,905	27,618	27,513	47,703	49,207	34,908	35,812
2020	49,207	48,752	30,509	30,296	51,931	54,638	38,624	40,145
2025	52,949	52,307	33,039	32,719	55,785	59,604	41,995	44,054
2030	56,255	55,426	35,241	34,776	58,919	63,665	44,891	47,350
2035	59,909	58,850	37,607	37,026	62,493	68,082	47,813	50,584

Reprinted from *Osteoarthritis and Cartilage*, vol. 23, Culliford D, Maskell J, Judge A, Cooper C, Prieto-Alhambra D, Arden NK, COAST Study Group, Future projections of total hip and knee arthroplasty in the UK: results from the UK Clinical Practice Research Datalink, pp. 594–600, 2015,⁹⁸ with permission from Elsevier.

The reason for the mismatch between lifetime risk of OA^{75,76} and the established intervention (THR/TKR) is not clear, but could be in part a result of the lack of consensus on the type or severity of symptoms.¹³² The mismatch may also be a result of the difference between the need for and provision of hip and knee arthroplasty in the UK, as described by Judge *et al.*^{107,133} Attempts have been made to understand the geographical and sociodemographic characteristics of different countries,³⁰ and UK-based studies from the late 1990s^{22,23,134} found little mismatch in THR provision but a large mismatch between the estimated need for TKR and provision. Data from the NJR⁷⁰ show that, for England and Wales, growth in provision was much slower over the period 2007–10 than in 1995–2005,⁶⁷ suggesting that the perceived gap between need and provision is unlikely to have been substantially narrowed in the time since the end of our analysis in 2006. This would depend on there being no change in the risk of developing OA over the same period.

Surgical thresholds are also important and hold their own implications for access and provision of care. There have been a number of strategies to cope with this increase in demand, for example national waiting list initiatives, but without accurate and reliable information long-term planning is difficult. It is therefore essential to have up-to-date and accurate disease-specific information available to estimate the future burden of the interventions, as well as the underlying diseases that are indications for hip and knee arthroplasty. This information also needs to be substantiated and validated for an accurate consensual agreement of the changing trends within the surgical community. Estimates from the USA are useful but not always consistent with epidemiological findings from the UK.

Accurate and reliable evidence of the demand and need for hip and knee arthroplasty is necessary for future planning. However, estimating rates for forecasts is difficult because surgical capacity is influenced and limited by governmental planning. Nevertheless, in the absence of supply-side forecasts, future projections based on the extrapolation of observed data⁹⁸ suggest that modest increases in the number of arthroplasties are to be expected in the UK, and that the dramatic increases forecast for the USA up to 2030⁶⁶ are unlikely to be matched in this country.

We estimated the high rates of THR and TKR using the log-linear model. We think that simply using a straight log-linear model (which projects exponentially) to estimate future projections is not helpful as a

substantive analysis, particularly over such a long projection period. Given the rise in the temporal trends in the incidence rates (1991–2010) on which the projections are anchored, it is not surprising that the curvature on a log-linear scale will eventually (by 2035) produce a very large number. Therefore, we did not major on this simple log-linear extrapolation but took a more refined and considered approach.

From the CPRD database we considered the first primary replacement for each subject to be the one of interest, and for the survival study we ignored subsequent primaries, even though the second primary may be a genuine contralateral primary. Likewise, we took the first revision encountered as the one that matches up with the primary. This approach, although arbitrary, is consistent, and is not likely to be biased with respect to any other important factors of interest in our studies (e.g. age, sex and BMI). For survival analyses, taking the first revision recorded is less of an issue, given that multiple revisions on the same-side joint are not only possible but quite common, over and above those revision procedures that are designed to be performed in two separate stages. We found the distributions of certain event coding to be good, considering that the CPRD is derived from routinely recorded data; however, the coding is not always perfect.

The main limitation of the work performed in this work package was the lack of individual validation of events in the data of the CPRD. However, several studies have shown the data to be accurate and complete for the clinical codes corresponding to OA, fracture and other crucial variables. The diagnosis of OA in general practice is often based on clinical symptoms without radiological support. This is because current general practice guidelines do not recommend radiographs to make a diagnosis of uncomplicated OA. Furthermore, we have examined a random sample of patients from general practice with a clinical diagnosis of knee OA to validate the diagnosis. We found that > 75% of GP diagnoses were confirmed using validated criteria. Moreover, there is an increasing need to study the epidemiology and management of the more clinically relevant diagnosis of clinically diagnosed hip OA.

The provisional number of patients identified in the GPRD is consistent with the expected number of patients for hip and knee joint replacements, which are less likely to be a result of misclassification or under-reporting.

The CPRD data are routinely gathered for all contributing practices and are not explicitly censored when requested, including left-censoring. The CPRD has a practice-level requirement that the data delivered to the database by each practice should be 'up to standard', but this does not affect patient-level data. Similar to most users of CPRD data, we used only data from practices after the point at which the CPRD deemed them to be 'up to standard'. Although a subject may have had a primary THR/TKR event after his/her practice began submission of data but before that practice was deemed to be 'up to standard', we used only the up-to-standard data. Therefore, if a subject satisfying the study selection criteria is registered with a valid CPRD practice for any length of registration period, he/she will be in the data set so long as that period coincides with the time during which the practice is supplying up-to-standard data.

Another limitation of our work was that it was possible for us to encounter revisions whereby the matching primary was carried out before the up-to-standard date, before the practice began submission or even before the patient registered with that practice. Once again, our consistent and conservative approach was not to use revision events in our survival or lifetime risk analyses unless we had a valid preceding primary replacement event for the same subject.

Conclusion

In conclusion, rates of hip and knee replacement rose between 1991 and 2006. Projections of future growth in the number of procedures to 2035 suggest a slower increase than that observed since the early 1990s. The long-term risk of revision for hip and knee replacements is slightly higher for subjects with higher BMI, but the effect is small.

Chapter 3 Work package 2: designing the statistical tool to predict surgery outcome

This chapter describes the important risk factors for poor outcome and combine them to produce a clinically relevant instrument (tool) to predict poor outcome and replacement failure.

The chapter contains information from publications that were based on work package 2.

Our objectives for work package 2 were to:

1. describe the risk factors for primary and revision surgery using the data from existing national and hospital prospective arthroplasty cohort studies
2. combine these risk factors to produce a clinically meaningful panel of predictors for poor outcome.

Predictors of outcome

Identification of operational, clinical and biological predictors of poor outcome after TJR is urgently required for a balanced provision of services and avoidance of unjustified use of resources when there is a high risk of implant failure. Identification of risk factors for poor outcome will guide translational research in terms both of the specific interventions and specific patient group selection. In addition, this will provide vital information to clinicians, patients and their carers, as those with a high risk of revision may wish to forgo surgery whereas those with a lower risk would be reassured.

Baker *et al.*³⁶ used a NJR to determine the role of pain and function in postoperative patient satisfaction and found that pain was a significant factor in patients not being satisfied with their operation.

Aims and objectives

In this work package we identify the important operational, clinical, biological and other important risk factors for poor outcome for lower limb joint replacements. We then combine previously described risk factors in order to develop a statistical tool for identification of patients with poor outcomes following THRs and TKRs. To achieve this goal, the work has been done to:

1. initially define the good and bad PROMs
2. identify the role of univariable as well as multivariable risk factors in patient-reported outcomes
3. develop a statistical tool to predict poor outcomes following THR and TKR surgeries.

We have completed further work to confirm the role of individual predictors for hip and knee replacement surgeries, particularly BMI, age and sex, patient's preoperative expectation, premorbidities (such as OA) and the type of implant. We have summarised our findings on the potential risk factors of good or bad patient-reported outcomes after THR and TKR, and revision results from the GPRD, currently known as the CPRD, in a number of publications.^{135,136}

In this chapter we discuss each individual risk factor and its association with surgery outcomes, as described in published papers. At the end of each subsection for risk factors we will summarise the publications when these associations were reported.

Design and setting

Cohorts and databases

The list of databases and cohorts we used in work package 2 is outlined in this section.

European Collaborative Database of Cost and Practice Patterns of Total Hip Replacement^{137,138}

The European Collaborative Database of Cost and Practice Patterns of Total Hip Replacement (EUROHIP) study consists of 327 patients receiving THR treatment across 20 orthopaedic centres in Europe. Patients completed self-administered questionnaires on demographic variables and baseline of pain, stiffness, mobility and quality of life. In addition, they were asked about their expectations of surgery 1 year after the operation. We also collected preoperative radiographs and data on operative procedure, including the prosthesis type used. Patients undergoing primary hip replacement in whom the indication for surgery was OA were included, but those with hip disease other than OA, severe mental conditions and/or dementia were excluded.

Exeter Primary Outcomes Study¹³⁹

In the Exeter Primary Outcomes Study (EPOS), participants were recruited between January 1999 and January 2002 from seven centres in England and Scotland. Patients received primary THR using a cemented Exeter femoral stem component (Stryker Howmedica Osteonics, Mahwah, NJ, USA).¹⁴⁰ A variety of cemented and uncemented acetabular components were used in included patients. Ethics approval was obtained from the North Western Multiple Centre Research Ethics Committee and the local research ethics committees. The cohort was representative of a wider orthopaedic cohort, as the participating hospitals (both teaching and district general hospitals) covered a wide geographical area including urban and rural locations; thus, it covered both affluent and somewhat deprived inner-city suburban areas with an overall population of 1 million.

Elective Orthopaedic Centre¹⁴¹

The Elective Orthopaedic Centre (EOC), also known as the South West London Elective Orthopaedic Centre (SWLEOC), is a purpose-built orthopaedic treatment centre that was opened in 2004. The centre serves a population of 1.5 million people in south-west London and it performs TKR surgeries across four acute NHS trusts: Kingston, St George's, Mayday, and Epsom and St Helier. In this work package we included patients who received either primary UKR or TKR.

Knee Arthroplasty Trial¹⁴²

Participants in the Knee Arthroplasty Trial (KAT) received primary TKR across 34 centres in the UK between July 1999 and January 2003. The primary outcome was the Oxford Knee Score (OKS) at 12 months after surgery. For this work package we used a wide range of preoperative predictors such as patient characteristics, as well as clinical and surgical factors.

Portsmouth and North Staffordshire¹⁴³

We used the data on patients undergoing THR from two districts in England, Portsmouth and North Staffordshire, with a population of approximately 1 million. The districts were selected for our work because, first, they were specialised in the assessment and treatment of hip OA; second, there was much support from the local orthopaedic surgeons; and, third, the district had a diverse socioeconomic profile. The orthopaedic surgeons recorded all men and women aged > 45 years who were listed for primary THA between 1993 and 1995. For our work we excluded the patients who had sustained a hip fracture within the past year, with a diagnosis of rheumatoid arthritis (RA) or ankylosing spondylitis, and those with secondary OA.

Clinical Practice Research Datalink^{77,94}

The CPRD, formerly known as the GPRD, is an English NHS observational data and interventional research service. It is jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and

Healthcare products Regulatory Agency. The database is designed to facilitate the linkage of anonymised patients' clinical data, thus enabling a number of epidemiological studies which would subsequently be beneficial for improved health-care services. The CPRD has become a gold standard in utilising data by observational researches. Currently over 890 clinical reviews and publications have benefited from the service. More information about the database has been previously detailed in *Chapter 2*.

A summary of the cohorts and data sets used in work package 2 is provided in *Table 13*.

Outcome measures

We used two scoring systems as a measure of surgery outcome:

1. PROMs, which are a condition-specific instrument jointly made up of the condition-specific Oxford score, a generic instrument EuroQol-5 Dimensions (EQ-5D) and general patient-specific questions¹⁴⁴
2. the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC).

Oxford Hip Score/Oxford Knee Score

Two of the most commonly used and nationally recommended scores are the Oxford Hip Score (OHS) and OKS,¹⁴⁵ incorporated into PROMs as described but also used as stand-alone questionnaires. They were originally designed as joint-specific scores for use in clinical trials to measure population-based changes,¹⁴⁶ and widely assessed for reliability and validity^{145–147} for this intended purpose. The OHS and OKS consist of 12 questions asking patients about their joint-specific pain and function in the preceding 4 weeks. Questions are scored on a Likert scale from 0 to 4, with the results added up to give a total score. The overall score maximum is 48 units, with 0 the worst possible score, indicating poor function and/or severe pain; and 48 representing the best score, suggesting no adverse symptoms and excellent joint function. Overall satisfaction of Oxford scores is measured by a visual analogue scale (VAS) with scores from 0 to 100 units.

EuroQol-5 Dimensions

The EQ-5D is joint non-specific questionnaire asking patients about their general health state, mobility, self-care, usual activities, pain and anxiety/depression.¹⁴⁸

TABLE 13 Cohorts and data sets used in work package 2

Cohort/database	Year of inception	Arthroplasty	Number of patients	Longest duration of observation (years)	Pain outcome	Joint failure
CPRD	1987	Hip	27,155	18	N	Y
		Knee	23,536			
EPOS	1999	Hip	1411	5	Y	Y
KAT	2000	Knee	2000	5	Y	Y
Portsmouth and North Staffordshire	1993	Hip	643	8	Y	Y
St Helier	1995	Hip	4089	12	Y	Y
EOC	2004	Hip and knee	Still recruiting – > 10,000	3	Y	Y
N, no; Y, yes.						

Satisfaction

In a number of cohorts, patients were asked about their satisfaction with surgery. In many cohorts, the measure of satisfaction is split between satisfaction with the procedure and satisfaction with the service. Satisfaction in some cohorts is a dichotomous variable, whereas in others it measured on a VAS.

Western Ontario and McMaster Universities Osteoarthritis Index

The WOMAC consists of 24 items with three domains: pain, stiffness and physical function. The scores range from 0 (no symptoms) to 100 (symptoms with extreme severity). The total score is created by adding the scores from each domain, multiplying by 100 and then dividing by the maximum score. Combination of these three domains adds up to a total score of 96, which is then converted into a normalised score. In order to classify whether or not patients improved 1 year after their surgery, we used the Outcome Measures in Rheumatology Clinical Trials/Osteoarthritis Research Society International (OARSI) responder criteria. For statistical analysis we used logistic regression to describe association between preoperative expectations and response to surgery.

Short Form questionnaire-12 items

The Short Form questionnaire-12 items (SF-12) consists of 12 questions and is the shorter version of the Short Form questionnaire-36 items (SF-36) health survey. The questionnaire is designed to collect the data on patients' mental and physical functioning, and also overall quality of life related to health. SF-12 is not age or disease specific, but is a general measure of health. The physical and mental component summary scores are combined and range from 0 to 100, in which 0 is indicative of the lowest possible status of health and 100 of the best possible status of health.

Main exposure

Primary and revision hip and knee replacement surgery.

Sample size

For a diagnostic (predictive) model to be of use in predicting outcomes it needs to have a sensitivity of at least 90% and a specificity of at least 75%. The power calculations are based conservatively on having data for 32,500 hip arthroplasties and 27,300 knee arthroplasties. It is assumed that we have complete data for only 80% of the patients and that 16% will experience arthroplasty failure, as defined by poor functional outcome at 2 years. This would result in there being 5200 patients with whom to estimate sensitivity and 27,300 patients with whom to estimate specificity for THR. With 5200 patients, a true sensitivity of 90% can be estimated to within 1% (95% CI 89% to 91%), and a true specificity of 75% can be estimated to within 0.5% (95% CI 74.5% to 75.5%). For TKR, there would be 4370 patients with whom to estimate sensitivity and 22,930 patients with whom to estimate specificity for knee arthroplasty, giving similar precision for the estimated sensitivity of 90% (95% CI 89% to 91%) and specificity of 75% (95% CI 74.4% to 75.6%).

Variables

A detailed literature search was conducted for an up-to-date list of potential risk factors for poor outcome. Our a priori list of factors can be grouped into three main groups: (1) technical factors, including type of prosthesis and cemented compared with uncemented; (2) other non-patient-related factors, such as the hospital where the surgery took place (size, throughput and expertise) and year of surgery; and (3) patient-related factors, such as age, sex, obesity, underlying arthritic condition, comorbid medical problems, radiographic parameters, intraoperative findings, postoperative complications and medication, such as non-steroidal anti-inflammatory drugs, bisphosphonates, hormone replacement therapy, statins and corticosteroids.

Preoperative radiographs were graded for severity using the Kellgren and Lawrence (K/L)¹⁴⁹ grading system for OA and the Sharpe score for RA. Grading of radiographs was done by trained research assistants and in cases of disagreement a formal consensus reading was performed. Radiographs were digitally scanned for a more detailed assessment of bone quality and shape.

The early postoperative radiographs (approximately 6 weeks post surgery) were assessed for joint alignment and the quality of subchondral bone as early predictors of joint failure.

Statistical methods

For identifying and combining predictors we have used various statistical methods. The list of such methods is detailed for each work/publication in *Appendix 2*. Statistical methods for designing predictive tools are detailed in *Developing a predictive model*.

Defining good and bad patient-reported outcome measures

Patient-reported outcome measures have been in development over a number of years for different reasons, and often for use as outcomes in clinical trials. They are now increasingly being used to assess a patient's satisfaction with hip or knee replacements.^{150,151} They are also increasingly considered as a potential tool to prioritise patients for surgery. This has raised some methodological concerns as these scoring systems traditionally accounted for the mean improvement of population group by using pre- and postsurgical scores as a continuous variable, with little validation for use as an outcome measure in individuals.

Some of the most commonly used PROMs include OHSS¹⁴⁶ and OKSs,¹⁴⁷ all validated to assess pain, stiffness and function. However, the scores have historically been used mostly as continuous outcome measures.

The government has introduced the PROM, as the recognised PROM for all patients having a hip or knee replacement,¹⁵² and attempts have been made to use PROMs to prioritise patients for surgery, limiting those with lower (worse) scores. The intention is that all data collected as part of the PROM programme are published and used towards the patient choice agenda. This kind of access to data provides patients and health-care professionals with the opportunity to discuss the information available regarding their care options, and results in shared informed decision-making.¹⁵³

Initially, the routine collection of PROMs has been brought in to clinical trials and national audits followed by the government initiative in 2009 to introduce PROMs throughout the NHS as a measure of improvement of clinical quality within the health-care system.¹⁵⁰ OHSS and OKSs, which form part of PROMs, were not, however, designed to be utilised in this way. Moreover, little work has been done to suggest that Oxford scores can actually predict surgery outcome. We therefore investigated how preoperative Oxford scores can be used for the definition of patient-reported outcomes, a prerequisite for prioritising patients to access surgery.

What is a good patient-reported outcome after total hip replacement?¹⁵⁴

To assess the suitability of the Oxford scores, we investigated the possibility of defining a postoperative OHS threshold anchored on patient satisfaction, as described in Arden *et al.*¹⁵⁴ As OHS is a continuous variable ranging from 0 units, as the worst possible score, to 48 units, as the best possible score, it is unclear which score on this continuous scale would define a cut-off point indicating patient satisfaction with surgery. To explore this, we used the St Helier Hospital Outcome Programme data for defining a postoperative OHS threshold anchored on patient satisfaction. To investigate patients' satisfaction at 12 and 24 months after the surgery, patients were asked 'are you satisfied with the result of your hip replacement?', to which patients would answer yes or no.

From baseline characteristics, we included BMI, sex, age, OHS and duration of pain. *Table 14* shows the median values of these characteristics for the full cohort (799 patients).

A total of 799 patients who had THR were eligible in the period 1986–2007. Of those, 77.5% ($n = 619$) completed the 12-month follow-up questionnaires and 80.0% ($n = 639$) completed the 24-month follow-up questionnaires. The underlying diagnosis was OA in 95.4% of selected patients. Outcome measures included age, height, weight, sex, expectation and satisfaction questions, and OHS.

Indication for surgery was available for only 239 patients, which was a limitation of this data set. Within this group, 95.4% of operations were carried out because of OA/coxarthrosis. The remaining 4.6% of cases were a result of avascular necrosis ($n = 4$), failed postfracture fixation ($n = 3$), acetabular erosion secondary to hemiarthroplasty ($n = 1$), unspecified arthritis ($n = 1$), hip dysplasia ($n = 1$) and joint pain ($n = 1$). Only 487 patients had a baseline BMI measurement and, among these patients, the median BMI was 27 kg/m² [interquartile range (IQR) 24–30 kg/m²]. The duration of pain measurement was available in 630 patients, and sex and baseline OHS were present in all.

TABLE 14 Baseline characteristics of patients¹⁵⁴

Baseline characteristics	Median (full cohort, <i>n</i> = 799)	Non-respondents	
		12-month follow-up (<i>n</i> = 180)	24-month follow-up (<i>n</i> = 160)
BMI (kg/m ²)			
All, median (IQR)	27 (24–30), <i>n</i> = 487	27 (24–30), <i>n</i> = 106	27 (23–30), <i>n</i> = 93
Low tertile, median (range)	23 (15–25), <i>n</i> = 197	23 (15–25), <i>n</i> = 42	22 (15–24), <i>n</i> = 32
Medium tertile, median (range)	27 (26–29), <i>n</i> = 143	27 (26–29), <i>n</i> = 33	27 (25–29), <i>n</i> = 34
High tertile, median (range)	32 (30–43), <i>n</i> = 147	33 (30–42), <i>n</i> = 31	33 (30–39), <i>n</i> = 27
Sex			
Female, frequency (%)	480 (60.1)	107 (59.4)	96 (60.0)
Male, frequency (%)	319 (39.9)	73 (40.6)	64 (40.0)
Age (years)			
All, median (IQR)	68 (58–76), <i>n</i> = 797	65 (54–74), <i>n</i> = 179	67.5 (56–77), <i>n</i> = 158
Low tertile, median (range)	54 (20–62), <i>n</i> = 269	51 (20–60), <i>n</i> = 65	53 (20–60), <i>n</i> = 54
Medium tertile, median (range)	68 (63–73), <i>n</i> = 275	65 (63–71), <i>n</i> = 56	68 (61–74), <i>n</i> = 52
High tertile, median (range)	79 (74–100), <i>n</i> = 253	78.5 (72–100), <i>n</i> = 58	80 (75–100), <i>n</i> = 52
OHS (units)			
All, median (IQR)	17 (11–23), <i>n</i> = 799	17 (12–24), <i>n</i> = 180	17 (12–23), <i>n</i> = 160
Low tertile, median (range)	9 (0–13), <i>n</i> = 278	10 (1–15), <i>n</i> = 71	10 (1–14), <i>n</i> = 57
Medium tertile, median (range)	17 (14–21), <i>n</i> = 277	18 (16–21), <i>n</i> = 51	17 (15–20), <i>n</i> = 50
High tertile, median (range)	27 (22–41), <i>n</i> = 244	26.5 (24–30), <i>n</i> = 58	25 (21–36), <i>n</i> = 53
Duration of pain, median (IQR)	1–3 years (1–3 years, 3–5 years), <i>n</i> = 630	1–3 years (1–3 years, 3–5 years), <i>n</i> = 83	No observations

IQR, interquartile range.

Note

Reprinted from *Osteoarthritis and Cartilage*, vol. 19, Arden NK, Kiran A, Judge A, Biant LC, Javaid MK, Murray DW, *et al.* What is a good patient reported outcome after total hip replacement? pp. 155–162, 2011, with permission from Elsevier.¹⁵⁴

Two different statistical methods were used to identify the cut-off points, which were anchored on patient satisfaction: one was the receiver operating characteristic (ROC) curve technique, which was used to identify the thresholds to maximise sensitivity and specificity; and the other was the 75th percentile approach. The OHS, which is quoted to patient satisfaction, at 12 months was ≥ 38 units using the ROC curve technique and ≥ 38 units by the 75th percentile approach. At 24 months, the figures were 33 units and 40 units, respectively (Figure 9). Using changing OHS as the outcome of choice, the ROC curve revealed that the value of satisfaction was 15 at 12 months and 14 at 24 months.

We took the 75th percentile from the top end of the OHS curve, leaving the cut-off point at 25%. At follow-up, 91.9% were satisfied at 12 months and 92.8% at 24 months, whereas 24 patients were unsatisfied at 24 months.

To assess whether or not these cut-off points varied according to important baseline characteristics, we performed a stratified analysis. The characteristics were sex, age (tertiles), BMI (tertiles), baseline OHS (tertiles), preoperative expectation of pain ('not at all painful' vs. 'any pain') and expectations for function ('not limited at all' vs. 'any limitation').

We demonstrated that the cut-off points, when using the change in OHS between the baseline and 24 months, have greater variation across patients than when using only 24-month OHSs (Figure 10). The value associated with satisfaction was greater in women than men. There were also greater variations accounting for BMI and patients' preoperative expectations. The patients who had the highest preoperative OHS required the lowest change in the score in order to be satisfied.

Using the 12-month data, we demonstrated the heterogeneity in the cut-off points, particularly when stratified for BMI and age. The greatest discrepancy was seen for change and baseline OHS, but with the lesser changes seen in percentage for potential improvement.

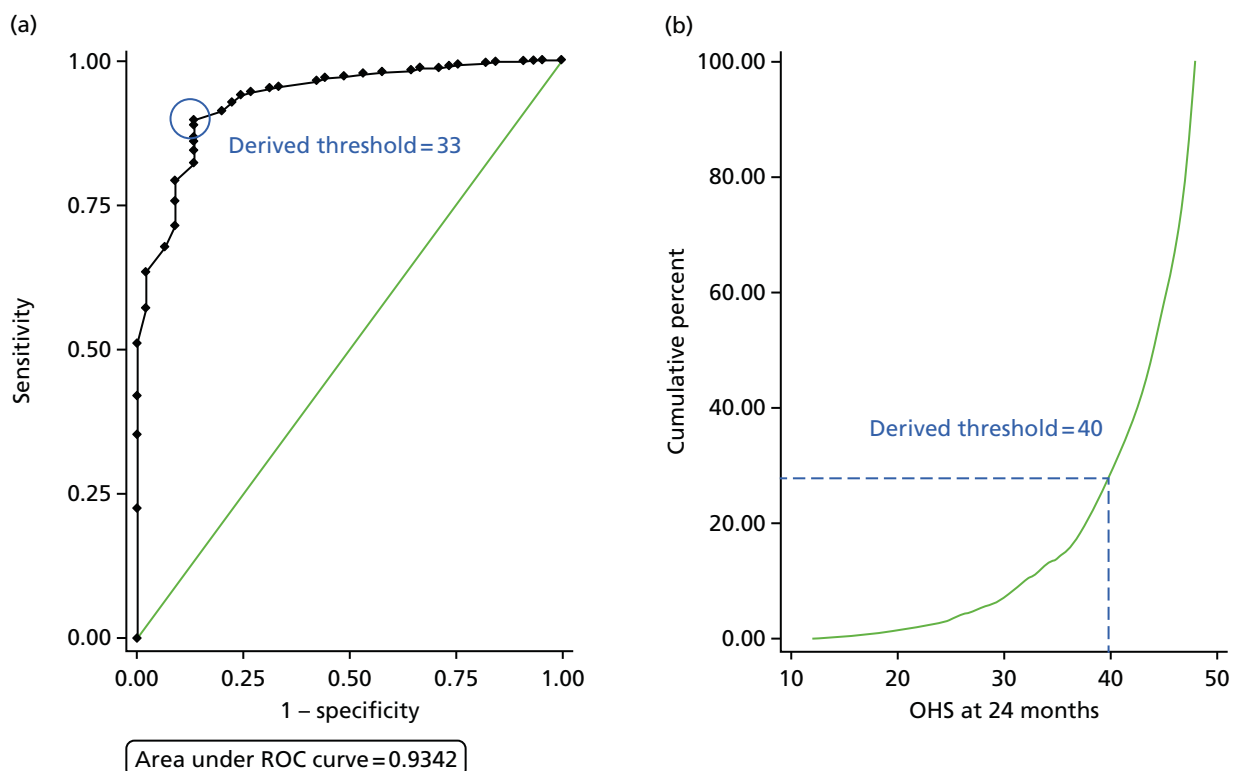


FIGURE 9 Patient satisfaction at 2 years after surgery: (a) ROC plot; and (b) 75th percentile. The 75th percentile shows the cut-off point in which positive responses are within the 75% data. Reprinted from *Osteoarthritis and Cartilage*, vol. 19, Arden NK, Kirna A, Judge A, Biant LC, Javaid MK, Murray DW, *et al.*, What is a good patient-reported outcome after total hip replacement? pp. 155–162, 2011, with permission from Elsevier.¹⁵⁴

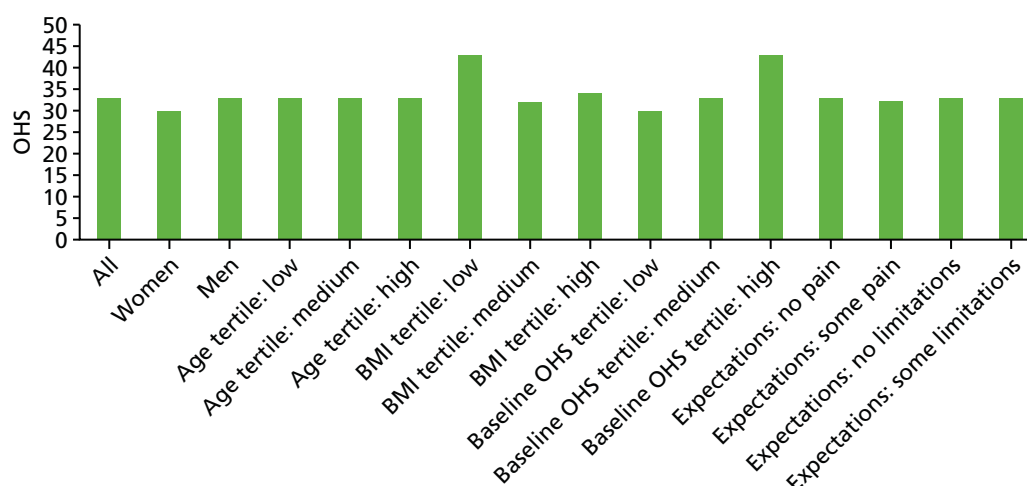


FIGURE 10 The ROC curve cut-off points for satisfaction at 24 months using 24-month OHSs, stratified by baseline clinical variables.¹⁵⁴ Reprinted from *Osteoarthritis and Cartilage*, vol. 19, Arden NK, Kirra A, Judge A, Biant LC, Javaid MK, Murray DW, *et al.*, What is a good patient-reported outcome after total hip replacement? pp. 155–162, 2011, with permission from Elsevier.¹⁵⁴

This study below confirms that we could identify the cut-off point for outcome of hip replacement surgery, which could be used for our research. This was based on the patient-accepted symptom state (PASS) methodology. In view of the heterogeneity discovered on stratification, however, this was not acceptable for clinical decision-making as a single outcome measure and more work would be required to stratify outcome measures if this was to be the case.

Interpretation of patient-reported outcomes for hip and knee replacement surgery¹⁵⁵

Having identified a cut-off score for outcome of hip arthroplasty at 12 months, we then used a second cohort, the EOC, to produce a PASS score for the OHS but also to validate the OHS. This purpose-built EOC performs hip and knee replacement surgeries for four acute NHS trusts (Mayday, Kingston, St George's, and Epsom and St Helier) serving 1.5 million people. The database routinely collected OHS and OKS preoperatively and 6 months after surgery.

We obtained baseline and 6-month postoperative Oxford scores from 1523 patients undergoing hip replacement and 1784 patients undergoing knee replacement. Six months after the surgery, patients were asked to complete their overall satisfaction with surgery using a VAS. On a VAS, 0 depicts no satisfaction and 100 shows a complete satisfaction (very satisfied). We identified a threshold value of ≥ 50 . This cut-off point was observed with 93% patients who had hip arthroplasty and with 89% who had knee arthroplasty.

A ROC curve was used to identify PASS score thresholds for absolute changes in OHS and OKS. For OHS this cut-off point was 14 when 97.6% patients declared satisfaction with surgery, and for OKS the cut-off point was 11 when 95.4% said that they were satisfied with surgery.

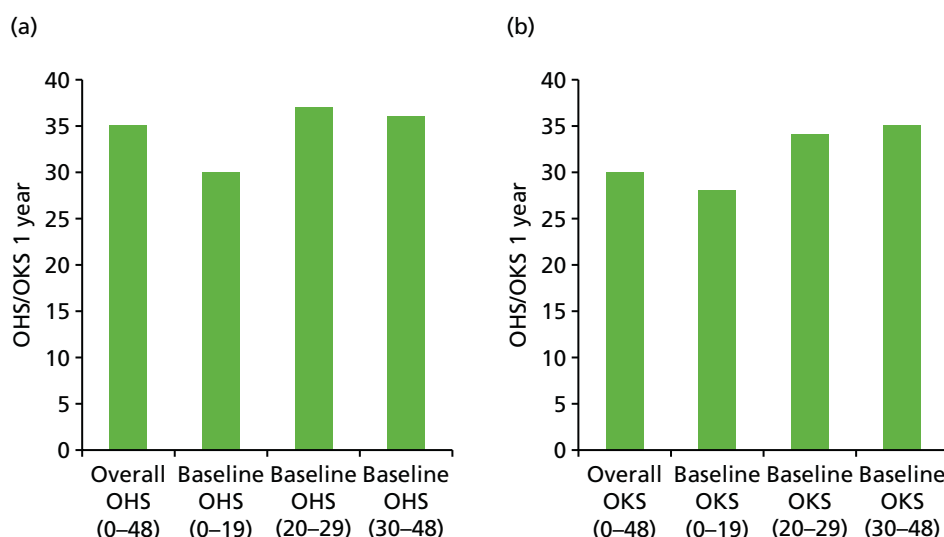
Table 15 shows these results at the baseline and follow-up at 6 months scores in tabular format.

The mean improvement (the absolute change) in OHS was 19 units (10.5 units) and in OKS was 14.5 units (9.8 units). Of interest, 80 patients undergoing THR had no change or worsening of OHS at 6 months, of whom 56.3% still declared themselves satisfied with surgery. Of the 143 patients whose OKS remained unchanged or deteriorated, 54.6% reported that they were satisfied with surgery. Of those whose pain score improved, 94.9% of hip replacement patients and 92.2% of knee replacement patients were satisfied. Using ROCs, the OHS associated with satisfaction at 6 months was 35 units (95% CI 32.9 to 37.1 units) and, of the patients achieving this score, 98% were satisfied with their surgery compared with 78.6% of those not meeting the threshold. The same figures for OKS were 96.7% and 70.1%, respectively. Figure 11 shows the thresholds according to baseline Oxford scores. Overall, it can be seen that patients

TABLE 15 Descriptive statistics of satisfaction with surgery: OKS and OHS preoperatively and at 6 months

Time point	Surgery	
	OKS	OHS
Preoperative		
Mean (SD)	19.9 (8.0)	19.7 (8.8)
Median (IQR)	20 (14–25)	19 (13–26)
6 months		
Mean (SD)	34.5 (9.1)	38.8 (8.7)
Median (IQR)	36 (29–42)	41 (34–46)
Patients satisfied with surgery, <i>n</i> (%)	1591 (89.2)	1415 (92.9)

Note
Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery.¹⁵⁵

**FIGURE 11** Bar charts illustrating the cut-off points of (a) OHS; and (b) OKS in relation to satisfaction 6 months post surgery. Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery.¹⁵⁵

starting with a higher baseline score, not surprisingly, have higher thresholds for satisfaction. Patients with worst pain/function baseline scores require higher change in Oxford scores to achieve the highest level of satisfaction, in contrast to patients with the better preoperative scores. However, patients with low preoperative scores (severe symptoms) require lower 6 months' postoperative Oxford scores to achieve a higher state of satisfaction.

A PROMs score on its own does not translate into a clinically meaningful outcome for either clinicians or patients. We aimed to describe how absolute changes in Oxford scores can relate to patients' satisfaction with surgery. We used PASS¹⁵⁶ to identify the cut-off points for the difference between the scores at baseline and 6 months after surgery. PASS depicts the value of OHS/OKS beyond which a patient's consideration of his or her own health status is good. These cut-off points for the change, as well as 6-month scores, are useful for both patients and clinicians as they improve the understanding of representation of a 'very good outcome' as opposed to a 'good outcome' and of the expectations of the operation.

Overall, this study demonstrated that thresholds for satisfaction could be identified for use in this research programme. The value obtained for the knee was not dissimilar from the previous study and a new value for the OKS was identified. This was substantially, and significantly, lower than the hip score.

Novel methodological approach for measuring symptomatic change following total joint arthroplasty¹⁵⁷

There has been confusion in the orthopaedic literature owing to the different methods used to define patient-reported outcomes. Some papers have used the score as the main outcome with or without adjustments for baseline score. Others have used the changes in score. This has often caused discrepant results in predictors of outcome, most notably when using baseline functional score and pain score as predictors. Harmonising the outcome measure is needed in current research. Both of these functional and pain scores have had limitations. We propose a new score – percentage of potential change (PoPC). PoPC is computed as the actual change divided by the potential improvement multiplied by 100. Thus, 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 (*Figure 12*).

We have used the data from the EOC of patients who underwent THA and TKA between 2004 and 2009. Patients had completed OHS and OKS questionnaires both preoperatively and 6 months postoperatively. For the analysis, 1523 OHS and 1784 OKS completed questionnaires were used. In addition to Oxford scores, patients were also asked to complete a short questionnaire about their satisfaction with surgery on a VAS. A threshold of ≥ 50 units was used to generate a binary variable to identify whether or not patients

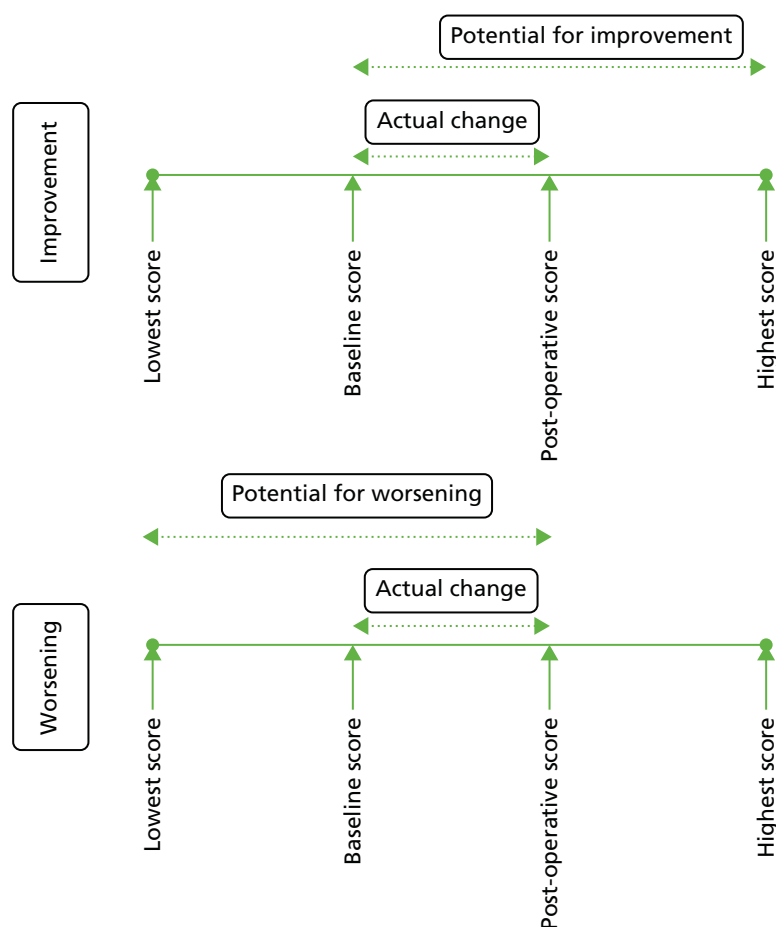


FIGURE 12 Percentage of potential changes. Reprinted from *The Journal of Arthroplasty*, vol. 29, Kiran A, Hunter DJ, Judge A, Field RE, Javaid MK, Cooper C, Arden NK. A novel methodological approach for measuring symptomatic change following total joint arthroplasty, pp. 2140–5, 2014, with permission from Elsevier.¹⁵⁷

were satisfied with surgery. Patients with potential improvement have a PoPC of > 0 units, with a potential worsening PoPC of < 0 units and patients with no actual change have a PoPC value of 0 units. PoPC allows the expression of how much a patient's symptoms have improved or worsened.

Kiran *et al.*,¹⁵⁷ again, demonstrated excellent improvements following knee and hip replacement surgeries. Correlation (Spearman's rank-order correlation) of patient satisfaction score with each of the outcome measures was greatest for PoPC and lowest for the relative change (*Table 16*).

The results showed that the ROC analysis, anchored on satisfaction as the outcome area under the curve (AUC) for hip replacement were follow-up score of 0.86 units, PoPC of 0.86 units, actual change of 0.83 units, relative change of 0.75 units, and for knee replacement were 0.85, 0.85, 0.8 and 0.72 units, respectively.

Kiran *et al.*¹⁵⁷ have demonstrated the importance of defining outcome measures. Different outcome measures identified different numbers of patients as being responders and non-responders, which has major implications in terms of health economics and health-care planning. It also demonstrates that outcomes for individual subjects differ depending on the outcome measures. This is critically important when trying to define outcome in an individual patient, as in this programme.

In summary, so far we have demonstrated that the OHSs and OKSs can be used to define patient-reported outcome following hip and knee replacement surgery. We have defined cut-off points for the OHSs and OKSs that equate to patient satisfaction with the procedure for use in the following sections to identify predictors of outcome.

Value-added publication

Assessing patients for joint replacement: can preoperative Oxford Hip and Knee Scores be used to predict patient satisfaction following joint replacement surgery and to guide patient selection?⁵⁰

In response to increasing moves to use the Oxford scores to ration access to lower limb joint replacement, we investigated the predictive nature of preoperative OHS and OKS in determining patient satisfaction postoperatively. We used the database from the SWLEOC, in which OHS or OKS were routinely collected with the addition of a satisfaction questionnaire both preoperatively and 6 months postoperatively. A total of 1523 THR patients and 1784 TKR patients were selected. Patients were asked routinely to complete the questionnaires with OHSs and OKSs at baseline, preoperatively and 6 months after their surgery. They were also asked to measure their overall satisfaction with surgery using a VAS, with scores from 0 to 100.

We used scatterplots to identify, and describe, the associations between participants' preoperative Oxford scores and satisfaction 6 months postoperatively. We found no such association, as shown in the scatterplots (*Figure 13*).

TABLE 16 Spearman's rank-order correlation of satisfaction with surgery (95% CI)¹⁵⁷

	Surgery	
	OHS	OKS
Follow-up score	0.49 (0.45 to 0.53)	0.57 (0.54 to 0.60)
Actual change	0.43 (0.40 to 0.47)	0.49 (0.46 to 0.53)
Relative change	0.30 (0.25 to 0.34)	0.34 (0.30 to 0.38)
PoPC	0.52 (0.49 to 0.56)	0.58 (0.55 to 0.61)

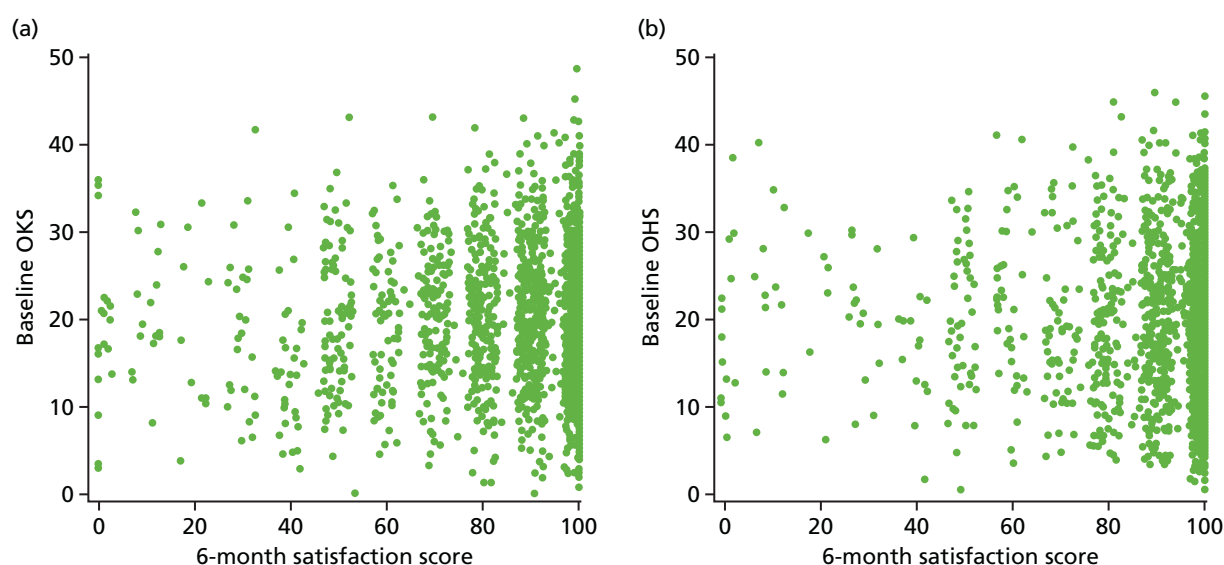


FIGURE 13 Scatterplots showing relationship between (a) baseline OKS; and (b) baseline OHS compared with satisfaction at 6 months after surgery. Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery.⁵⁰

Spearman's rank-correlation coefficients between Oxford scores and satisfaction at 6 months postoperatively were -0.04 units for the OHS (95% CI -0.09 to 0.01 units) and 0.04 for the OKS (95% CI -0.01 to 0.08 units). We also found no differences in median satisfaction scores by baseline OHS (*Table 17*) or OKS (*Table 18*) with the Kruskal–Wallis test. Interestingly, TKR results suggested that TKR patients with the lowest preoperative scores were most dissatisfied with their surgery at 6 months after their operation; conversely, scores were not predictive in patients undergoing THR.

This study suggests that, based on 6-month outcomes, preoperative Oxford scores should not be used to predict patient satisfaction after surgery. We conclude that it is unlikely that PROMs can be used on their own to predict patient satisfaction, but it is likely that a combination of multifactorial elements would play an important factor. Such a multidimensional instrument would incorporate a wide range of risk factors in the assessment of pain, function, satisfaction and health-related quality of life (HRQoL).

TABLE 17 Patient satisfaction with surgery 6 months after THR, stratified by baseline scores

OHS	Number of patients	Number of satisfied patients at 6 months after surgery (IQR)	Kruskal–Wallis <i>p</i> -value	Patients satisfied at 6 months after surgery (%)	Chi-squared <i>p</i> -value
Total	1523	100 (90–100)		1415 (92.9)	
Baseline OHS of ≤ 26 units	1170	100 (90–100)	0.45	1085 (92.7)	0.63
Baseline OHS of > 26 units	353	100 (90–100)		330 (93.5)	
Baseline OHS of 0–15 units (low)	533	100 (90–100)	0.36	496 (93.1)	0.97
Baseline OHS of 16–24 units (medium)	546	100 (90–100)		506 (92.7)	
Baseline OHS of 25–46 units (high)	444	100 (90–100)		413 (93.0)	

Note

Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery.⁵⁰

TABLE 18 Patient satisfaction with surgery 6 months after TKR, stratified by baseline scores

OKS	Number of patients	Number of satisfied patients at 6 months after surgery (IQR)	Kruskal–Wallis <i>p</i> -value	Patients satisfied at 6 months after surgery (%)	Chi-squared <i>p</i> -value
Total	1784	90 (80–100)		1591 (89.2)	
Baseline OKS of ≤ 20 units	954	90 (75–100)	0.079	834 (87.4)	0.01
Baseline OKS of > 20 units	830	90 (80–100)		757 (91.2)	
Baseline OKS of 0–16 units (low)	623	90 (75–100)	0.36	540 (86.7)	0.037
Baseline OKS of 17–23 units (medium)	568	90 (80–100)		511 (90.0)	
Baseline OKS of 24–47 units (high)	593	90 (80–100)		540 (91.1)	
Note Reproduced with permission and copyright © of the British Editorial Society of Bone and Joint Surgery. ⁵⁰					

Identifying the role of univariable as well as multivariable risk factors in patient-reported outcomes

In this work package we report on the identification of the operational, clinical, biological and other important risk factors for poor outcome, and on the combining of risk factors, with the aim of producing a clinically relevant instrument to predict poor outcome. We have summarised our findings on the potential risk factors of good or bad patient-reported outcomes after THR and TKR, and revision results from GPRD and the new CPRD.^{135,136}

Data from a number of cohorts were used to inform the analysis. In particular, four data sets, which have been reported in detail elsewhere, were used extensively:

1. EUROHIP¹³⁷
2. EPOS¹³⁹
3. EOC database¹⁵⁵
4. St Helier Hospital outcome programme.¹⁵⁴

In this section, we will discuss each individual risk factors and its association with outcomes, as described in published studies.

Preoperative Oxford Hip Score and Oxford Knee Score

Introduction

In the UK, within the NHS, PROMs have been routinely used as the outcome measure. PROMs comprise the OHS, the OKS, the EQ-5D and several questions asking patients about their satisfaction with surgery, services and expectations. The OHS was developed in 1996 mainly for use in clinical trials.¹⁴⁶ The OHS and OKS are joint-specific measures,^{145,158} whereas EQ-5D is non-specific, asking patients about their general health state, mobility, self-care, usual activities, pain and anxiety/depression.¹⁴⁸

As previously described, the OHS and OKR score patients' satisfaction between 0 and 48 units, with 0 describing the worst possible pain and function, and 48 indicating the best possible outcomes. These scales have been used to identify patients with surgery failure; however, scores alone are not sufficient to inform clinicians and patients of the outcome. For this reason, we introduced a dichotomous variable, the

PASS score.¹⁵⁶ The PASS score is a necessary measure to translate the continuous Oxford score variable into a binary variable, that is, to calculate a cut-off point as an indicative score for surgery satisfaction or dissatisfaction. It is a useful variable that provides clinicians and patients with more meaningful information about outcomes of surgery. In addition to PROMs, we have also used the WOMAC, which assesses pain, stiffness and function.¹⁵⁹

Findings

Knee

We investigated the baseline OKS from the EOC database. We included patients undergoing total and UKR surgeries, but excluded those with previous knee surgeries and bilateral operations. Judge *et al.*¹³⁸ investigated the OKS at baseline and 6 months after surgery, as well as the absolute change in OKS.¹³⁸ The histogram shows a relatively normal distribution of the baseline OKS (Figure 14a). The OKS distribution after 6 months is skewed to the right, indicating that pain and function improved in the majority of patients (Figure 14b).

We established a PASS score using the ROC curve in order to produce the outcome for this study. A score of 30 units was used as a PASS score, which identified 71.7% of the patients ($\text{OKS} \geq 30$ units) as satisfied

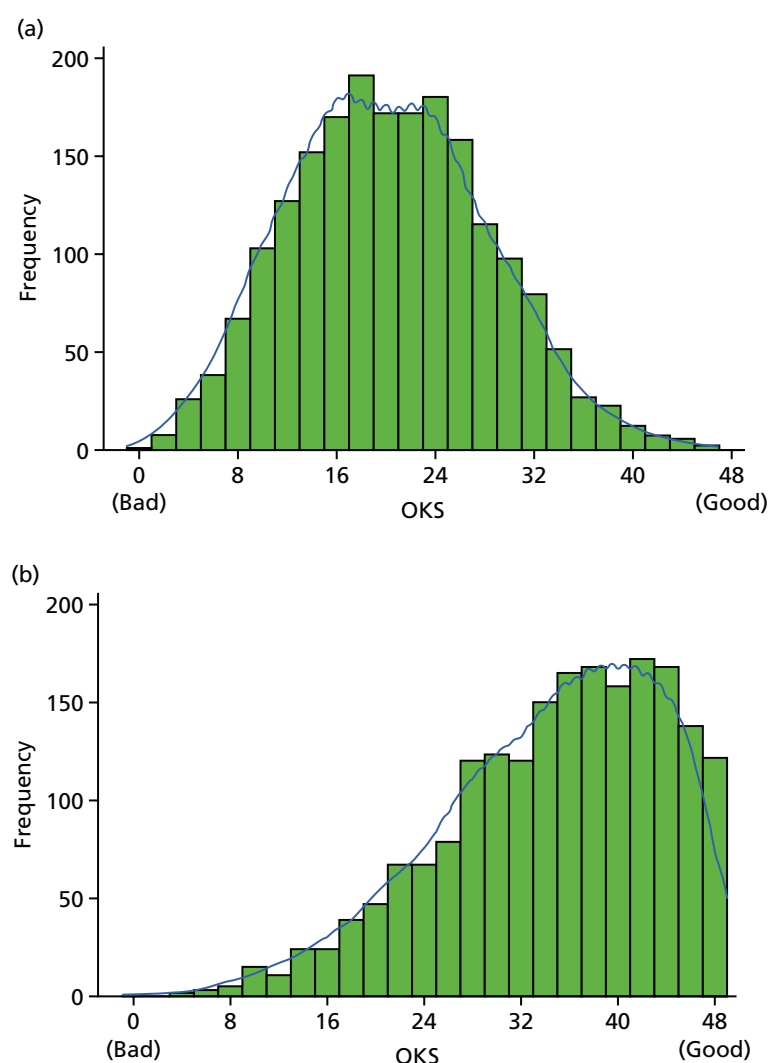


FIGURE 14 Distributions of OKS at (a) baseline; and (b) 6 months after surgery. Reproduced from Judge A, Arden NK, Cooper C, Kassim Javaid M, Carr AJ, Field RE, Dieppe PA. Predictors of outcomes of total knee replacement surgery. *Rheumatology*, 2012, vol. 51, pp. 1804–13, with permission of Oxford University Press.¹⁴¹

at 6 months postoperatively. A higher baseline OKS predicts better outcome, defined by the PASS score, at the 6-month follow-up [odds ratio (OR) 1.52, 95% CI 1.40 to 1.66]. It also predicted a higher follow-up OKS ($\beta = 1.70$, 95% CI 1.43 to 1.96).

In another study, Sánchez-Santos *et al.*¹⁴² used the data from 1967 patients from the KAT across 34 UK centres in the UK.¹⁴² The results showed a baseline OKS mean of 18.2 units (SD 7.5 units) among patients who completed both pre- and postoperative questionnaires. The study found an association between the OKS at baseline and the OKS 12 months after surgery; patients with a better preoperative OKS achieved better pain and functional outcome following TKR ($\beta = 0.35$, 95% CI 0.29 to 0.42).

Hip

Our work confirmed that a higher PROMs score was associated with a better outcome.^{139,160} The mean baseline OHS was 16.5 units (SD 7.6 units) in patients from the EPOS cohort¹³⁹ and 16.49 units (SD 7.7 units), 15.67 units (SD 8.61 units), 19.51 units (SD 8.77 units) and 17.52 units (SD 8.30 units) in a meta-analysis combining responders at 12 months of the four cohorts.¹⁶⁰ Baseline SF-36 physical function score, collected from responders of two England health districts, was 20 units (SD 5.35 units).

Primary hip replacement for OA was assessed by using the EPOS prospective cohort.¹³⁹ The study showed that better preoperative pain/function was associated with a higher postoperative score. *Figure 15a* shows the left-skewed histogram of baseline OHS. *Figure 15b* depicts the 1-year OKS distribution, indicating

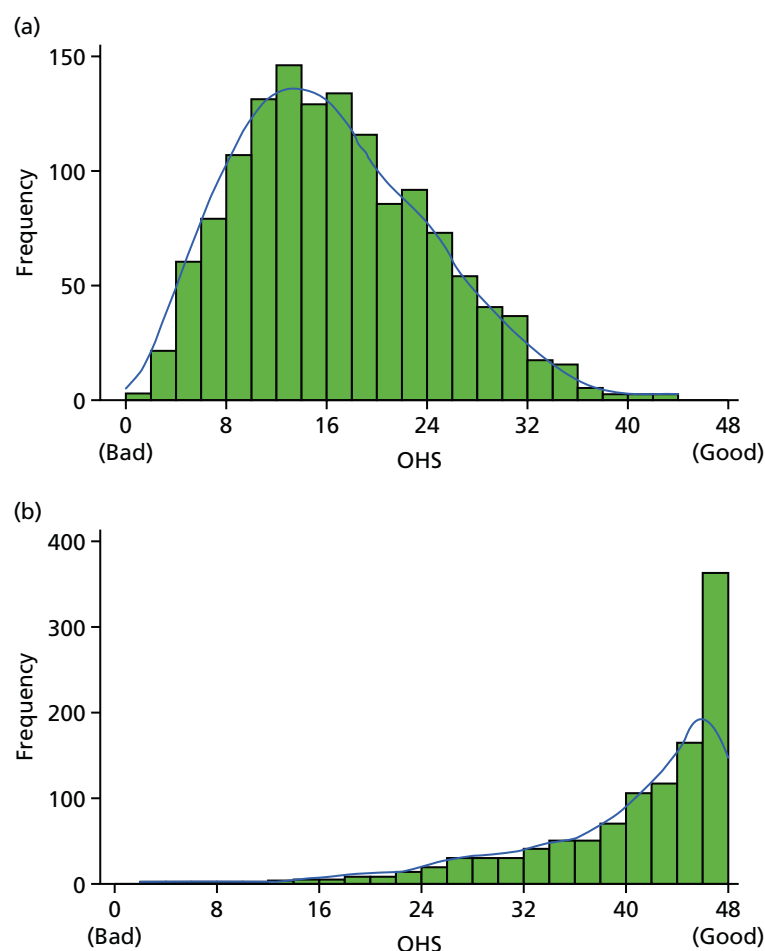


FIGURE 15 The distribution of OHSs at (a) baseline; and (b) 1 year postoperatively. Reproduced from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

that the majority of patients exhibited better outcomes at 1–5 years using a repeated measures linear regression model (baseline OHS 10 units; $\beta = 2.68$, 95% CI 2.16 to 3.21).

More importantly, this study found that, regardless of the preoperative OHS, participants attained a statistically significant improvement in pain and function after THR.¹³⁹ We observed that the patients with the worst baseline scores attained the greatest improvements in pain and function (patients with a preoperative OHS of < 5 achieved a 28.8-point change in the score). However, the patients with the best baseline scores still attained a substantial improvement (patients with a baseline OHS of > 30 achieved a 10.6-point change) (Figure 16).

We have investigated the association of the baseline OHS and BMI.¹⁶⁰ We observed that the baseline OHS decreased as BMI increased; that is, patients of normal weight (BMI 18.5–25 kg/m²) had an OHS of 17.02 units (IQR 15.69–18.34 units), whereas patients with the highest weight (BMI ≥ 40 kg/m²) had an OHS of 12.25 units (IQR 9.02–15.49 units). Underweight patients (< 17.02 kg/m²) also had a low OHS (14.01 units, IQR 9.54–18.48 units). A higher baseline OHS was associated with a better outcome.

Interestingly, a further study investigating patients from two England public health districts (Portsmouth and North Staffordshire districts) demonstrated that patients reporting better baseline physical function were at higher risk of a bad postoperative outcome.¹⁴³ This can be explained by the fact that the study used a different measure from the OHS, that is, the SF-36. Functional improvement was classified as change in SF-36 physical function score in the upper quartile. The cut-off point for improved outcome on the SF-36 was defined as ≥ 30 units, which falls in the upper quartile. Patients with better preoperative scores had also improved outcomes, although they achieved a lesser change between pre- and postoperative scores. Finally, we observed that the increased baseline OHS was associated with a greater OHS.¹³⁸ This study emphasises the importance of classifying outcome similarly across all studies.

Conclusion

Preoperative pain/function was one of the strongest predictors of outcome. A high baseline preoperative score was related to a better postoperative score. It has been established that patients with lower

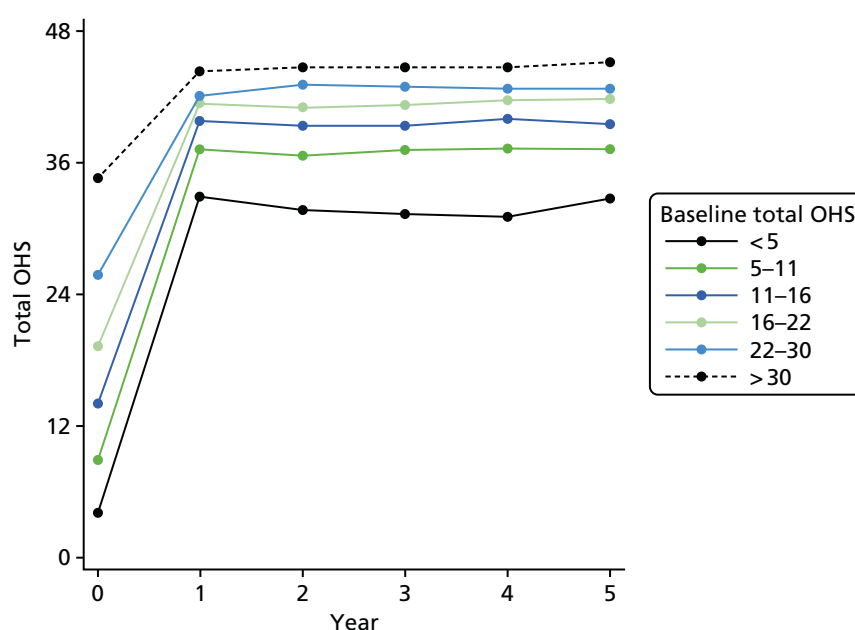


FIGURE 16 Change in OHS over time, stratified by baseline score, from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

preoperative pain and better preoperative function attain the highest postoperative pain and function, whereas patients with the worst baseline scores achieve the highest change between baseline and follow-up.^{32,44–47} We also need to consider floor and ceiling effects when we estimate change because the level of satisfaction attained differs according to the baseline degree of functional disability. It is also known that patients with the lower preoperative scores never obtain the original functional and pain levels, although this outcome is possible in patients with better preoperative scores. *Table 19* summarises our findings.

TABLE 19 Associations found between pre- and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	High OS predicts better outcome	Adjusted OR _{total OKS} : 1.52 (95% CI 1.40 to 1.66) ^a	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	High OS predicts better outcome	Linear model coefficient _{total OKS} : 1.70 (95% CI 1.43 to 1.96) ^a	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	High OS predicts better outcome	Linear model coefficient _{log total OKS} : 5.6 (95% CI 4.4 to 6.7) ^b	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS	Hip	PROMs	1- to 5-year OHS	High OS predicts better outcome	Δ linear model coefficient _{OHS (10 units)} : 2.68 (95% CI 2.16 to 3.21) ^c	Judge <i>et al.</i> ¹³⁹
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	High OS predicts better outcome	Linear model coefficient _{OHS (10 units)} : 1.48 (95% CI 0.62 to 2.34) ^d	Judge <i>et al.</i> ¹⁶⁰
Portsmouth and North Staffordshire	Hip	PROMs	6 months, with ≥ 30 points in the SF-36	High OS predicts lower risk of good functional outcome	Adjusted OR _{OHS (10 units)} : 0.73 (95% CI 0.60 to 0.89) ^e	Judge <i>et al.</i> ¹⁴³
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	High OS predicts better outcome	Linear model coefficient _{OHS (10 units)} : 2.23 (95% CI 1.68 to 2.79) ^f	Judge <i>et al.</i> ¹⁶¹

OS, Oxford Score.

a Predictor variables were age, sex, BMI, Index of Multiple Deprivation 2004, side (left or right), diagnosis (primary OA, RA or other), operation type (TKR or UKR), American Society of Anesthesiologists grade (1, 2, 3 or 4), preoperative EQ-5D anxiety/depression question, year of surgery and aged < 60 years.

b Predictor variables were age, sex, age × sex, Index of Multiple Deprivation 2004, BMI, SF-12 mental component summary score, American Society of Anesthesiologists grade (1, 2, 3 or 4), other conditions affecting mobility, previous knee surgery, fixed flexion deformity, preoperative valgus/varus deformity and preoperative anterior cruciate ligament.

c Predictor variables were 1- to 5-year postoperative OHS, age, sex, BMI, occupation, comorbidities, SF-36 mental component summary score and femoral component offset size.

d Confounding variables included age, sex, SF-36 mental component summary score, comorbidities, fixed flexion, analgesic use, college education, OA in other joints, expectation of less pain, radiographic K/L grade, American Society of Anesthesiologists grade (1, 2, 3 or 4) and years of hip pain.

e Predictor variables were age, sex, BMI, Index of Multiple Deprivation 2004, side of surgery, primary diagnosis (OA, RA, other), operation type (TKR or UKR), American Society of Anesthesiologists grade (1, 2, 3 or 4), preoperative OKS, preoperative EQ-5D anxiety/depression question and year of surgery.

f Predictor variables were age, sex, BMI, education, SF-36 mental component summary score, pattern of OA (superolateral, superomedial/medial/concentric) no reduction, number of joints with OA, number of joints with surgery, surgical approach (anterolateral or posterior) and femoral component offset size.

Age

Introduction

Previous research on the association of age with the outcome of TJR has been heterogeneous. Some authors have found an association between age and outcome, but others have found no such evidence.^{44–49,51,162} Literature reviews suggest that age is not a strong predictor of outcome.⁴³

Findings

Knee

Data from the EOC database showed no association with age overall but a minor effect on OKS function, with younger patients having a better outcome.¹⁴¹ However, despite the statistical significance, the effect size observed was small. We further explored this association using the KAT data and found that patients aged < 60 years and ≥ 80 years presented a worse pain and functional status at 12 months after knee surgery.¹⁴² We observed that younger women (aged < 60 years) had a better outcome than men, whereas in the oldest age group (aged ≥ 80 years) women had a worse outcome than men.

Age at TJR was also a significant predictor of revision for knee. Importantly, Culliford *et al.*⁹⁷ found that the patients undergoing TKR had a 4.3% reduction in revision rates for each extra year of age.⁹⁷ Figure 17 illustrates that patients aged between 55 and 65 years had up to 10% revision rates at 15 years after primary knee replacement; patients aged > 85 years had a < 2% revision rate; interestingly, the youngest group of patients had the highest revision rate.

Hip

Judge *et al.*³⁰ investigated the association between age and patient outcomes following THR surgery in the EUROHIP cohort. Age was grouped at baseline as < 50, 50–69 or ≥ 70 years. The results showed that there were no statistically significant differences across the age groups, although the youngest responders presented better outcomes.³⁰ In another study, Judge *et al.*¹⁴³ collected data from 282 patients from two England health districts: Portsmouth and North Staffordshire. The primary outcome was the long-term functional improvement after THR. Patients included in the study were aged ≥ 45 years and were listed for THR for primary OA. To identify the risk factors to predict functional improvement in the long term (≈ 8 years), we used the logistic regression modelling. The results showed that older patients were less likely to have an improvement in

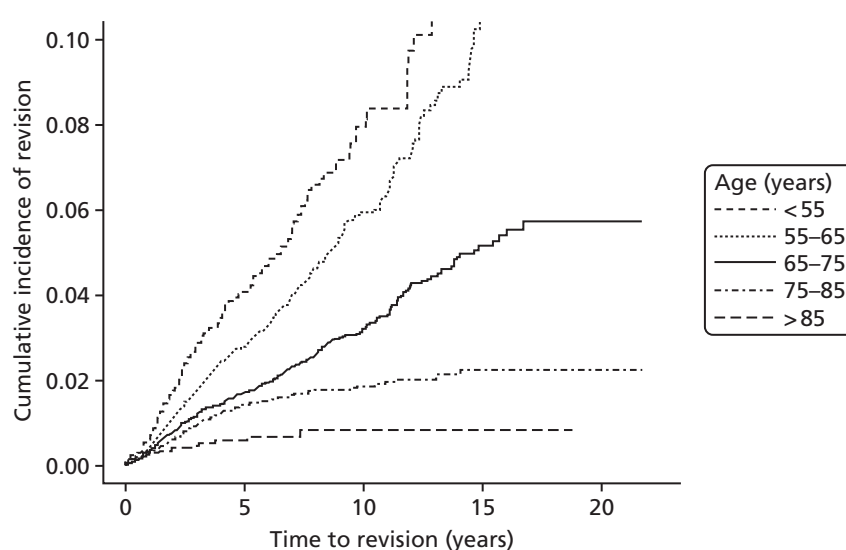


FIGURE 17 Cumulative incidence estimate for revision of TKR by age. Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution-Noncommercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

physical function; however, the association found in the study was weak.¹⁴³ Furthermore, data analysis from EPOS and EUROHIP cohorts revealed a non-linear association between age and outcome in patients undergoing THR.¹³⁸ In those patients aged ≥ 75 years, increasing age was associated with worse outcomes. In addition, a worse outcome was found in those patients aged 50–60 years than in those who are younger (aged < 50 years) and older (aged > 60 years),¹³⁹ although there was a small, statistically significant change in achieved postoperative outcome associated with patient age (Figure 18).

Among patients aged > 65 years, revision rates 15 years after the primary hip replacement were up to 10%⁹⁷ (Figure 19). Among THR patients, the revision rate fell by 3% for each extra year of age, after adjusting for the competing risk of death.

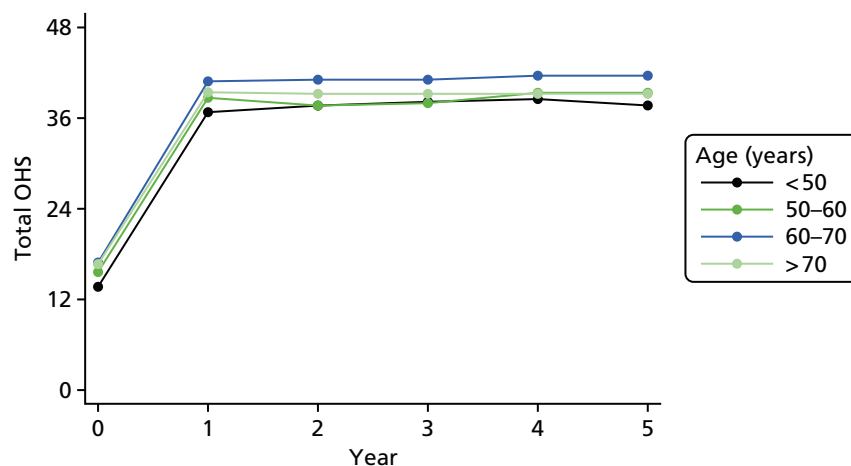


FIGURE 18 The change in OHS over time, layered by age, from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited and the use is non-commercial and is otherwise in compliance with the licence.

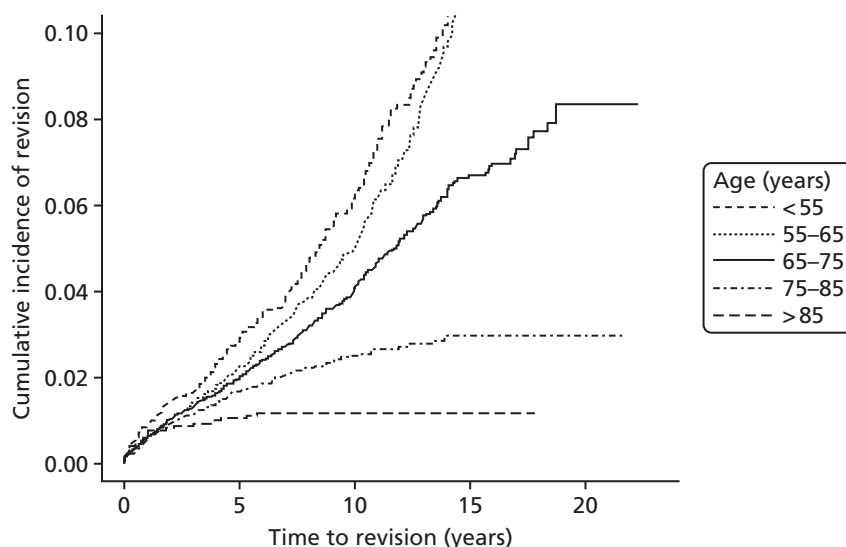


FIGURE 19 Cumulative incidence estimate for revision of THR by age. Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution-Noncommercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

Conclusion

We have demonstrated that increasing age reduces the risk of revision surgery.⁹⁷ Different, and heterogeneous, results were seen when using PROMs as the primary outcome. Some of our work showed that the effect of age on joint replacement surgery outcome is not significant,^{30,141,160} whereas other work showed the effect of age to be non-linear, with the youngest and oldest patients having the worst outcomes.^{138,139,142} Age was associated with functional outcome following TKR and THR, although size effect of the association was small.

Overall, the small statistically significant differences relating to patient age at the time of surgery were greatly outweighed by the substantial change in PROMs achieved by these patients.¹³⁹ *Table 20* summarises associations found between age and outcomes.

Sex

Introduction

The work carried out in work package 1 showed that the lifetime risk of THR in the UK is higher among women than among men.^{98,99} It was observed that, at the age of 50 years in 2005, the risk of THR was 11.6% (95% CI 11.1% to 12.1%) for women and 7.1% (95% CI 6.7% to 7.5%) for men. Similarly, the risk of TKR was also greater among women (10.8%, 95% CI 10.3% to 11.3%) than among men (8.1%, 95% CI 7.6% to 8.5%).⁹⁹ In work package 2 we aimed to confirm sex, as previously described in the literature, as a predictor of surgery outcome following THR and TKR.

Findings

Knee

Judge *et al.*¹⁴¹ found that, among the EOC cohort patients undergoing TKA, functional outcomes were worse in women. The attained 6-month OKS was 0.88 units lower in women than in men (95% CI 0.08 to 1.68 units). In the KAT cohort, the responders showed strong evidence of an interaction between age and sex: younger women (aged < 60 years) and older men (aged ≥ 80 years) had better outcomes.¹⁴² This difference of sex was not found on OKS outcome in the middle age groups (60–80 years).

Conversely, Culliford *et al.*⁹⁷ showed that the revision risks were significantly higher among men than among women following TKR. The adjusted overall SHR was greater in men than in women in the adjusted competing risk analysis (SHR 1.51, 95% CI 1.32 to 1.73).

Hip

In hip arthroplasty patients we did not find an association between sex and surgery outcomes.^{30,139,143,160} Slight improvements were observed in the EUROHIP and Portsmouth and North Staffordshire Health districts' cohorts, although this difference was found to be non-significant.^{30,143} On the other hand, Culliford *et al.*⁹⁷ described a significantly higher risk of revision THR in men than in women: men had a 23% higher risk of revision arthroplasty than women (1.23, 95% CI 1.10 to 1.38). Although women and men did not present differences in the OHS, WOMAC and SF-36 scores at 12 months, 1–5 years and close to 8 years after surgery,^{30,139,143,160} it was found that men had a greater risk of revision THR.⁹⁷

Conclusion

The results from the knee and hip studies show that sex may have small effects in knee arthroplasty patients with little effect on hip arthroplasty when using PROMs as the outcome. In general, women had a worse PROMs outcome (of minor clinical significance),^{141,142} whereas men were at a higher risk for prosthesis failure resulting in the revision arthroplasties.⁹⁷ Our results are summarised in *Table 21*.

TABLE 20 Associations found between age and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	NS pain effect. Weak function effect: younger Higher likelihood of improvement	Adjusted OR _{pain} : 0.98 (95% CI 0.92 to 1.05) Adjusted OR _{function} : 0.93 (95% CI 0.87 to 0.99)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	NS pain effect. Weak function effect: younger Higher likelihood of improvement	Linear model coefficient _{pain} : 0.01 (95% CI -0.10 to 0.12) Linear model coefficient _{function} : -0.21 (95% CI -0.3 to -0.08)	Judge <i>et al.</i> ¹⁴¹
GPRD	Knee	Revision	15 years	Increasing age reduces risk	SHR: 0.957 (95% CI 0.951 to 0.962) ^a	Culliford <i>et al.</i> ⁹⁷
KAT	Knee	PROMs	12-month OKS	Younger (< 60 years) and older (> 80 years) worst outcome	Linear model coefficient _{>80} : -2.8 (95% CI -5.6 to 0.1)	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS	Hip	PROMs	1- to 5-year OHS	Very old (> 80 years) worst outcome, but NS worse for young	Δ linear model coefficient _{>80} : -3.81 (95% CI -5.29 to -2.33) Δ linear model coefficient ₅₀₋₆₀ : -1.87 (95% CI -3.22 to -0.53)	Judge <i>et al.</i> ¹³⁹
GPRD	Hip	Revision	15 years	Increasing age reduces risk	Adjusted OR _(1 year) : 0.971 (95% CI 0.966 to 0.975)	Culliford <i>et al.</i> ⁹⁷
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	NS	Linear model coefficient: -0.28 (95% CI -1.12 to 0.57)	Judge <i>et al.</i> ¹⁶⁰
EUROHIP	Hip	PROMs	12-month WOMAC score	NS effect, but young did better	OR _{return to normal (<50)} : 1.7 (95% CI 0.6 to 4.6)	Judge <i>et al.</i> ³⁰
Portsmouth and North Staffordshire	Hip	PROMs	6 months, with ≥ 30 points in the SF-36	Weak significant effect, younger age higher likelihood of improvement	Adjusted OR _(10 unit) : 0.94 (95% CI 0.90 to 0.98)	Judge <i>et al.</i> ¹⁴³
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	Increasing age. When aged > 75 years worst outcome	Linear model coefficient _{≥75} : -2.00 (95% CI -3.55 to -0.45)	Judge <i>et al.</i> ¹⁶¹

NS, not significant.

^a Risk of THR and TKR revision associated with BMI, age and sex, after adjusting for the competing risk of death.

TABLE 21 Associations found between sex and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	NS	Adjusted OR _{female} : 0.92 (95% CI 0.72 to 1.17)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	Female worse	Linear model coefficient _{female} : −0.88 (95% CI −1.68 to −0.08)	Judge <i>et al.</i> ¹⁴¹
GPRD	Knee	Revision	15 years	Higher risk in men	Adjusted OR _{male} : 1.54 (95% CI 1.37 to 1.72)	Culliford <i>et al.</i> ⁹⁷
KAT	Knee	PROMs	12-month OKS	High predicts better outcome. Younger women and older men had a better outcome	Linear model coefficient _{male} : −4.6 (95% CI −7.7 to −1.4)	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS	Hip	PROMs	1- to 5-year OHS	NS	–	Judge <i>et al.</i> ¹³⁹
GPRD	Hip	Revision	15 years	Higher risk in men	Adjusted OR _{male} : 1.35 (95% CI 1.23 to 1.48)	Culliford <i>et al.</i> ⁹⁷
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	NS	Linear model coefficient: −0.88 (95% CI −0.67 to 2.43)	Judge <i>et al.</i> ¹⁶⁰
EUROHIP	Hip	PROMs	12-month WOMAC SCORE	No significant difference, but women had a slightly better improvement	Adjusted OR _{OMERACT/OARSI (female)} : 1.6 (95% CI 0.9 to 2.8)	Judge <i>et al.</i> ³⁰
Portsmouth and North Staffordshire	Hip	PROMs	6 months, with ≥ 30 points in the SF-36	NS	Adjusted OR _(female) : 0.37 (95% CI 0.19 to 0.72)	Judge <i>et al.</i> ¹⁴³
NS, not significant.						

Body mass index

Introduction

The World Health Organization recommends that BMI is classified into four categories: underweight (< 18.5 kg/m²), normal (between 18 and 25 kg/m²), overweight (> 25 to 30 kg/m²) and obese (class I, > 30 to 35 kg/m²; class II, > 35 to 40 kg/m²; and class III, > 40 kg/m²). BMI is widely recognised as an important predictor for many conditions, including OA, and, as such, a number of patients referred for lower limb arthroplasty will have a raised BMI. Although some studies identified a positive relationship between increased BMI and susceptibility to knee and hip OA, and the need for replacement surgeries,^{163–165} there is increasing concern that obesity could be seen as an obstacle to accessing replacement surgeries. We aimed to confirm an association between BMI and patient-reported outcomes of THR surgeries. A number of detailed studies investigating the BMI effects on surgery outcomes have been carried out. This section will detail the results from several publications.

Knee

We used the GPRD data to describe the association of BMI with revision rates.⁹⁷ The data were collected from patients who had undergone hip and knee replacement surgeries between 1998 and 2011. From this cohort we then identified those with subsequent revision surgeries. We investigated the effects of BMI on the time of revision surgery. We estimated cumulative incidences of TKR revisions at 1, 5, 10 and 15 years. The results showed that at 5 years the estimated cumulative revision rate for TKR was 1.9% (95% CI 1.8% to 2.1%). The cumulative incidences across all BMI groups are shown in *Figure 20*. For each increased unit of BMI the estimated risk of TKR revision increased by 1.015 (95% CI 1.002 to 1.028).

Our results suggested that BMI appears to be a significant risk factor of time to revision of TKR. The risk of revision for morbidly obese TKR patients was found as high as 6% after 10 years following surgery. Up to approximately 7 years there was a more even distribution across all BMI categories; however, there was a much higher risk for the morbidly obese patients between 7 and 10 years.

In Judge *et al.*,¹⁴¹ a high BMI was related to a worse outcome at 6 months after TKR (coefficient total OKS -0.44 units, 95% CI -0.86 to -0.01 units), although BMI was not found to be a clinically important predictor of outcome taking into account the PASS score binary variable (OR total OKS 0.90 units, 95% CI 0.80 to 1.01 units) (*Table 22*). Patients who had a higher BMI showed worse functional outcomes (coefficient function OKS -0.33 units, 95% CI -0.57 to -0.09 units). Cohort analysis by Sánchez-Santos *et al.*¹⁴² showed that a high BMI was a determinant factor associated with a decreased OKS at 12 months' follow-up. Although statistically significant, the effect sizes are small and below the minimal clinically important difference of the OKS.

Wallace *et al.*⁷⁷ further investigated the association between BMI and the risks of complications 6 months following TKR. We used the CPRD to collect baseline BMI measurements, as recorded by GP practices, on patients who had undergone primary TKR between 1995 and 2011. We selected 32,485 TKR patients (including those who died within 6 months of their surgery) and, of those, < 1% were underweight, 17% were of normal weight, 38% were overweight, 29% were obese class I, 12% were obese class II and 4% were obese class III. The following outcomes were recorded following their surgeries: myocardial infarction, stroke, deep-vein thrombosis (DVT) or pulmonary embolism (PE), respiratory infection, anaemia, urinary tract infection, wound infection and death.

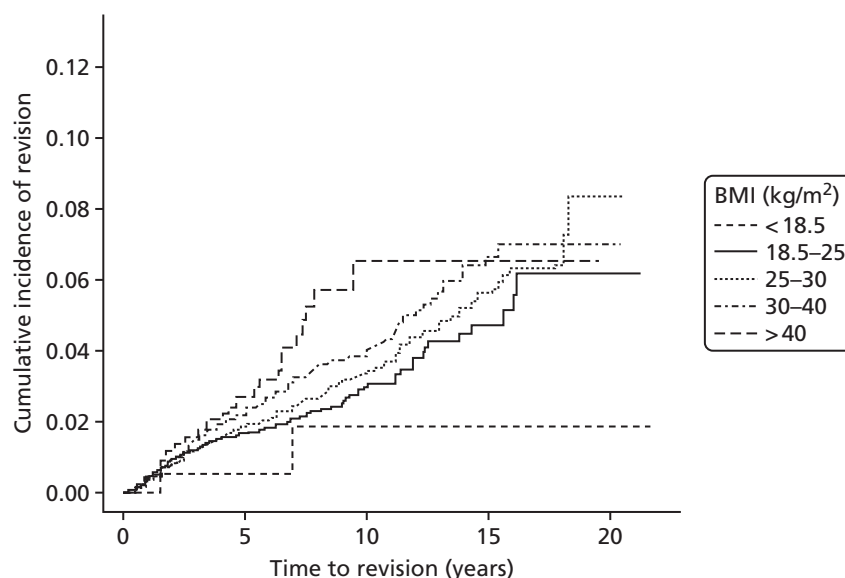


FIGURE 20 Estimated cumulative incidence for TKRs by BMI. Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution-Noncommercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

TABLE 22 Associations found between BMI and postoperative knee score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	NS	Adjusted OR _{total OKS} : 0.90 (95% CI 0.80 to 1.01)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	High BMI, worse outcome	Linear model coefficient _{total OKS} : -0.44 (95% CI -0.86 to -0.01)	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	High BMI, worse outcome	Linear model coefficient _{BMI (10 units)} : -1.5 (95% CI -2.4 to -0.6)	Sánchez-Santos <i>et al.</i> ¹⁴²
GPRD	Knee	Revision	15 years	Increasing BMI increases risk (small)	Adjusted OR _{TKR} : 1.015 (95% CI 1.002 to 1.028)	Culliford <i>et al.</i> ⁹⁷
CPRD	Knee	PROMs	DVT/PE 6 months after surgery	Increasing BMI increases risk	Adjusted OR _{TKR (obese)} : 1.59 (95% CI 1.26 to 1.99)	Wallace <i>et al.</i> ⁷⁷
CPRD	Knee	PROMs	Anaemia 6 months after surgery	Obesity decreases TKR risk	Adjusted OR _{TKR (obese)} : 0.74 (95% CI 0.58 to 0.94)	Wallace <i>et al.</i> ⁷⁷
CPRD	Knee	PROMs	A wound infection 6 months after surgery	Increasing BMI increases risk	Adjusted OR _{TKR (obese)} : 1.23 (95% CI 1.01 to 1.50)	Wallace <i>et al.</i> ⁷⁷
CPRD	Knee	PROMs	A UTI 6 months after surgery	TKR NS	Adjusted OR _{TKR (obese)} : 0.93 (95% CI 0.74 to 1.17)	Wallace <i>et al.</i> ⁷⁷
CPRD	Knee	PROMs	Death 6 months after surgery	Underweight increases risk	Adjusted OR _{TKR (underweight)} : 4.61 (95% CI 1.64 to 13.01)	Wallace <i>et al.</i> ⁷⁷

DVT, deep-vein thrombosis; NS, not significant; PE, pulmonary embolism; UTI, urinary tract infection.

The study analysis showed that a higher BMI was associated with a significantly higher risk of developing wound infections, up from 3% to 4.1% (adjusted $p < 0.05$), in TKR patients. An association was also found between increased BMI and DVT/PE risk, up from 2.0% to 3.3% (adjusted $p < 0.01$), in TKR patients. Interestingly, no association was found between BMI and other confounders, particularly myocardial infarction, stroke and mortality.

Hip

We have published several studies investigating associations between BMI and THR outcomes.^{97,160,166}

Table 23 shows these associations. The results of the study by Culliford *et al.*⁹⁷ show that at 5 years the estimated cumulative rate of revision surgery for THR was 2% (95% CI 1.8% to 2.1%). Figure 21 depicts cumulative incidences across all BMI groups for patients undergoing THR. BMI was a significant predictor of revision. Severely obese THR patients seem to have an increased risk of revision surgery in the first year. For each additional unit of BMI, the estimated risk of THR revision rises by 1.02 units (95% CI 1.009 to 1.032 units).

Batra *et al.*¹⁶⁶ used the analysis from four prospective cohort studies of patients who had undergone THR for OA: EPOS, SWLEOC, St Helier and EUROHIP. We determined the relationship between BMI and OHS at 1 year following THR. This was adjusted for the baseline OHS. We used a meta-analysis to combine the results from separately built models in all four cohorts. All models were adjusted for common variables such as sex and age. The analysis showed that, for every 5-unit rise in BMI, the 1-year OHS fell by 0.81 units (95% CI 0.55 to 1.08 units)¹⁶⁶ (Figure 22).

TABLE 23 Associations found between BMI and postoperative hip score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
GPRD	Hip	Revision	15 years	Increasing BMI increases risk (small)	Adjusted OR _{THR} : 1.020 (95% CI 1.009 to 1.032)	Culliford <i>et al.</i> ⁹⁷
EPOS	Hip	PROMs	1- to 5-year OHS	High BMI worse outcome but small effect	Δ linear model coefficient _{BMI (10 units)} : -1.54 (95% CI -2.45 to -0.64)	Judge <i>et al.</i> ¹³⁹
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	Higher BMI worse outcome	Linear model coefficient _{BMI (5 units)} : -0.81 (95% CI -1.08 to -0.54)	Batra <i>et al.</i> ¹⁶⁶
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	Higher BMI worse outcome but not clinically significant	Linear model coefficient _{BMI (5 units)} : -0.78 (95% CI -1.28 to -0.27)	Judge <i>et al.</i> ¹⁶⁰
EUROHIP	Hip	PROMs	12-month WOMAC score	The obese improved less than the non-obese	OR _{return to normal (obese)} : 0.8 (95% CI 0.5 to 1.3)	Judge <i>et al.</i> ³⁰
Portsmouth and North Staffordshire	Hip	PROMs	6 months, with ≥ 30 points in the SF-36	NS. No influence of BMI on functional outcome	Crude OR: 1.00 (95% CI 0.93 to 1.07)	Judge <i>et al.</i> ¹⁴³
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	High BMI, worse outcome	Linear model coefficient _{BMI (5 units)} : -0.66 (95% CI -1.10 to -0.22)	Judge <i>et al.</i> ¹⁶¹
CPRD	Hip	PROMs	DVT/PE 6 months after surgery	Increasing BMI increases risk	OR _{THR (obese)} : 1.64 (95% CI 1.34 to 2.00)	Wallace <i>et al.</i> ⁷⁷
CPRD	Hip	PROMs	Anaemia 6 months after surgery	THR NS	OR _{THR (obese)} : 1.03 (95% CI 0.83 to 1.28)	Wallace <i>et al.</i> ⁷⁷
CPRD	Hip	PROMs	A wound infection 6 months after surgery	Increasing BMI increases risk	OR _{THR (obese)} : 1.52 (95% CI 1.21 to 1.90)	Wallace <i>et al.</i> ⁷⁷
CPRD	Hip	PROMs	A UTI 6 months after surgery	Obesity increases THR risk	OR _{THR (obese)} : 1.25 (95% CI 1.02 to 1.55)	Wallace <i>et al.</i> ⁷⁷
CPRD	Hip	PROMs	Death 6 months after surgery	Underweight increases risk	OR _{THR (underweight)} : 2.71 (95% CI 1.67 to 4.39)	Wallace <i>et al.</i> ⁷⁷

NS, not significant; UTI, urinary tract infection.

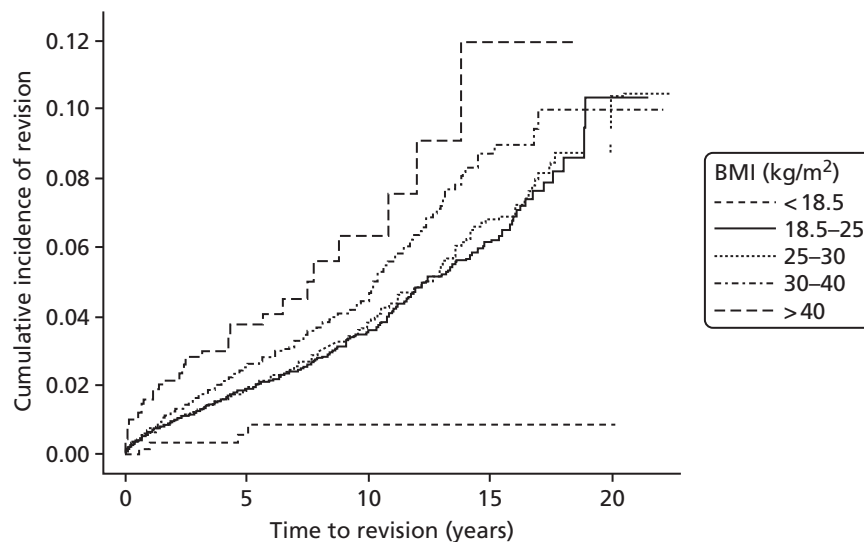


FIGURE 21 Figure showing estimated cumulative incidence for THRs by BMI. Reproduced from Culliford *et al.*⁹⁷ This is an Open Access article distributed in accordance with the Creative Commons Attribution-Noncommercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial.

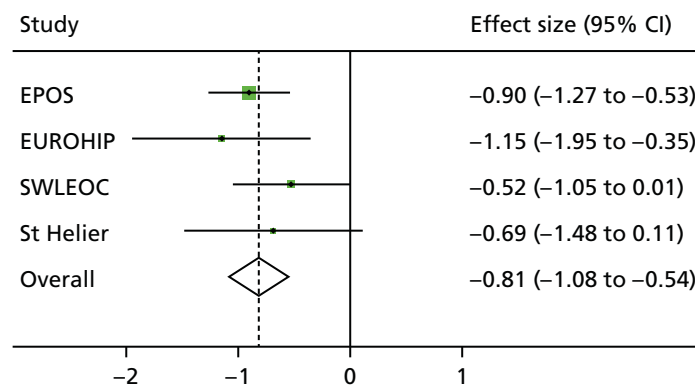


FIGURE 22 Fixed-effects meta-analysis from Batra *et al.*¹⁶⁶ *Osteoarthritis and Cartilage*, vol. 20, Batra RN, Judge A, Javaid MK, Thomas GE, Beard D, Murray D, *et al.* Pre-operative BMI as a predictor of patient reported outcomes of primary hip replacement surgery: a combined analysis of 4 prospective cohort study, pp. S152–3, 2012, with permission from Elsevier.

We then combined the data from all cohorts and used multiple imputations for missing data. With this analysis, when the sex and age variables were adjusted, every 5-unit rise in BMI was associated with a 1-year fall in OHS of 0.72 units (95% CI 0.46 to 0.99 units); when adjusting for all potential confounders, this reduction was decreased to 0.51 units (95% CI 0.09 to 0.92 units). The results suggest that, for each 5-unit increase in BMI, the difference in 1-year OHS becomes more significant.

Following these preliminary results, Judge *et al.*¹⁶⁰ further expanded the investigation and used the same cohorts, exposure, primary outcome OHS and confounding variables to reinvestigate whether or not BMI is a clinically significant predictor of patient-reported outcomes in patients with THR. *Tables 24 and 25* show that patients achieved significant improvement in their OHS, regardless of their baseline BMI value.

The results confirmed the following associations between BMI and OHS: for each 5-kg/m² increase in BMI, the OHS at 1 year decreased by 0.78 units (95% CI 0.27 to 1.28 units; $p < 0.001$). Obese class II patients would have a 1-year OHS 2.34 units lower than that of people with a normal BMI.

TABLE 24 Repeated measures regression model: estimates of pre- and postoperative OHS adjusted for age and sex

BMI categories (kg/m ²)	OHS, mean (95% CI)	
	Baseline	12 months
Underweight (< 18.5)	14.01 (9.54 to 18.48)	39.31 (34.93 to 43.68)
Normal (18.5–25)	17.02 (15.69 to 18.34)	40.04 (38.72 to 41.36)
Overweight (25–30)	16.65 (15.38 to 17.91)	39.01 (37.75 to 40.28)
Obese class I (30–35)	14.23 (12.81 to 15.64)	36.95 (35.54 to 38.37)
Obese class II (35–40)	13.69 (11.82 to 15.57)	35.90 (34.01 to 37.79)
Obese class III (> 40)	12.25 (9.02 to 15.49)	36.43 (33.10 to 39.76)

Reprinted from *Osteoarthritis and Cartilage*, vol. 22, Judge A, Batra RN, Thomas GE, Beard D, Javaid MK, Murray DW, *et al.*, Body mass index is not a clinically meaningful predictor of patient reported outcomes of primary hip replacement surgery: prospective cohort study, pp. 431–9, 2014,¹⁶⁰ with permission from Elsevier.

TABLE 25 Repeated measures regression model: estimates of pre- and postoperative OHS adjusted for all confounders

BMI categories (kg/m ²)	OHS, mean (95% CI)	
	Baseline	12 months
Underweight (< 18.5)	14.04 (9.56 to 18.52)	39.34 (34.97 to 43.71)
Normal (18.5–25)	16.83 (15.25 to 18.40)	39.85 (38.25 to 41.45)
Overweight (25–30)	16.79 (15.22 to 18.36)	39.15 (37.56 to 40.75)
Obese class I (30–35)	14.93 (13.13 to 16.72)	37.66 (35.93 to 39.39)
Obese class II (35–40)	14.71 (12.51 to 16.91)	36.92 (34.72 to 39.11)
Obese class III (> 40)	13.66 (10.24 to 17.07)	37.83 (34.25 to 41.41)

Reprinted from *Osteoarthritis and Cartilage*, vol. 22, Judge A, Batra RN, Thomas GE, Beard D, Javaid MK, Murray DW, *et al.*, Body mass index is not a clinically meaningful predictor of patient reported outcomes of primary hip replacement surgery: prospective cohort study, pp. 431–9, 2014,¹⁶⁰ with permission from Elsevier.

Our results confirmed that the effect of BMI on 1-year postoperative OHS is statistically significant, although the degree of significance is low: patients who are obese class II would have an OHS 2.34 units lower than patients with a normal BMI. In addition, patients who are rated as obese class II achieved a 22.2-unit change in their OHS following THR. As suggested by Murray *et al.*,¹⁴⁵ the smallest change in OHS that can be regarded as clinically important is approximately 5 units. Therefore, although a difference of 2.34 units is statistically significant, a difference of this magnitude in OHS at 1 year across all categories of BMI will have clinical significance only in patients who are rated as obese class II or III. This effect is greatly outweighed by a significant change in OHS in obese class II people (22.2-unit change), indicating substantial improvement in outcomes in this patients over the year. Thus, we conclude that BMI should not be indicative to deny patients access to hip replacement surgery.

Judge *et al.*¹³⁸ analysed the data from the EPOS cohort preoperatively and every year up to 5 years after the patients' THR surgery. The analysis showed that a 10-kg/m² increase in BMI was associated with decrease in OHS of 1.54 units (95% CI 0.64 to 2.45 units) averaged between the 1- and 5-year follow-ups, although these differences were small compared with the overall benefit of the operation (*Figure 23*).¹³⁹ Similarly, patients from the EUROHIP cohort presented similar results at 1 year after their THR [OR return to normal (obese) 0.8 units, 95% CI 0.5 to 1.3 units].³⁰ In contrast, there was no association found between BMI and functional outcomes in THA patients from Portsmouth and North Staffordshire districts.¹⁴³

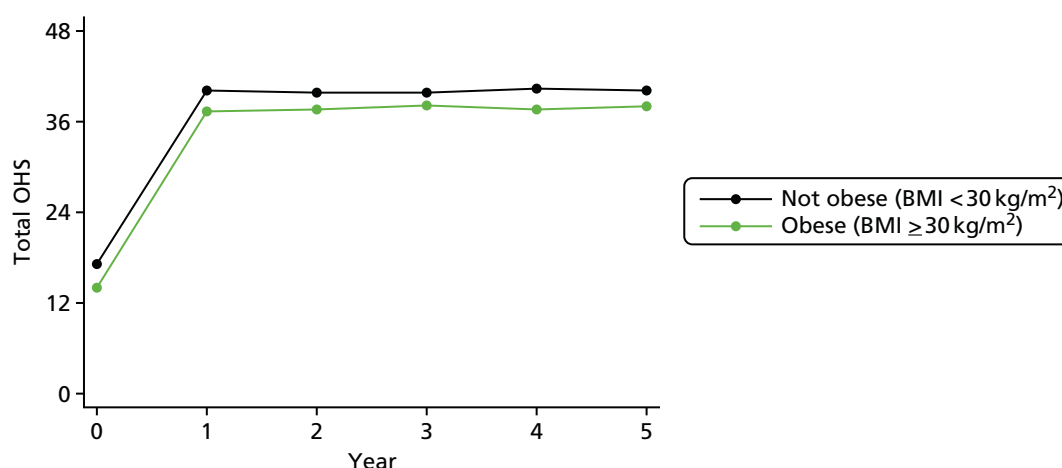


FIGURE 23 Change in OHS over time, stratified by BMI from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

Wallace *et al.*⁷⁷ investigated the association between BMI and the risks of complications 6 months after THR surgery using the CPRD.⁷⁷ We collected the baseline BMI measurements from the CPRD for THR patients between 1995 and 2011. For THR we selected 31,817 patients. From this cohort, BMI distribution was as follows: 1.5% underweight, 28% normal weight, 40% overweight, 21% obese class I, 7% obese class II and 2% obese class III. The results showed that in THR patients increased BMI was associated with a significantly higher risk of wound infections, ranging from 1.6% to 3.5% (adjusted $p < 0.01$). The association between increased BMI and DVT/PE risk was significant, with increased BMI from 2.2% to 3.3% (adjusted $p < 0.01$) in THR patients.

Conclusion

We found weak associations between increasing BMI and worse PROMs outcome. In accordance with other studies,¹⁶⁷ however, the effect sizes were small and often below the minimal clinically important difference for the Oxford scores. BMI was found not to be a strong predictor of functional outcomes. Therefore, high preoperative BMI should not be a deterrent to knee or hip replacement surgeries.^{26,43}

Obese patients have a high risk of developing of DVT, PE, wound infection and urinary tract infection following knee and hip replacement surgeries.

We cannot advocate selecting patients for joint replacement surgeries without consideration of BMI, but we do suggest that denial of surgery based on high BMI is unwarranted.

Deprivation

Introduction

Historically, deprivation was measured using the Index of Multiple Deprivation (IMD) 2004.⁵³ This index was extracted from the residence area where the patients lived at the time of surgery. The index is a compound of seven deprivation indices, employing the indicated weightings: income (22.5%), employment (22.5%), health deprivation and disability (13.5%), education, skills and training (13.5%), barriers to housing and services (9.3%), crime (9.3%) and living environment (9.3%). Poorer areas attract a higher deprivation score and more prosperous areas have a lower score. In addition, educational level was considered in some of our studies as a proxy of socioeconomic,^{30,161} as well as employment, status.³⁰ Table 26 shows postoperative scores in relation to deprivation indices.

TABLE 26 Associations found between deprivation and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	Deprived worse outcome	Adjusted OR _{total OKS} : 0.73 (95% CI 0.62 to 0.87)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	Deprived worse outcome	Linear model coefficient _{total OKS} : -1.40 (95% CI -1.96 to -0.85)	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	Deprived worse outcome	Linear model coefficient: -0.5 (95% CI -0.9 to -0.1)	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	College/university better outcome	Linear model coefficient: 3.39 (95% CI 0.12 to 6.67)	Judge <i>et al.</i> ¹⁶⁰
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	College/university better outcome	Linear model coefficient: 2.08 (95% CI 0.59 to 3.57)	Judge <i>et al.</i> ¹⁶¹
EUROHIP	Hip	PROMs	12-month WOMAC score	Employment no significant effect, but employed not significantly better	Adjusted OR _{return to normal (employed)} : 0.8 (95% CI 0.5 to 1.5)	Judge <i>et al.</i> ³⁰
EUROHIP	Hip	PROMs	12-month WOMAC score	Education better for dichotomous outcomes but not continuous	Adjusted OR _{return to normal (university degree)} : 2.9 (95% CI 1.4 to 5.9)	Judge <i>et al.</i> ³⁰

Findings

Knee

The data analysed from the EOC and KAT cohorts showed that deprivation is one of the main predictors of worse outcome in knee patients.^{141,142} The IMD 2004⁵³ score in the KAT cohort showed that, for each 10-unit increase in deprivation index, the 12-month OKS was reduced by 0.5 units (95% CI 0.1 to 0.9 units), whereas increased deprivation was associated with a lower OR of achieving a 6-month PASS score (OR 0.73, 95% CI 0.62 to 0.87) (see *Table 26*).

Hip

We did not identify any THR cohorts for which IMD data were available,⁵³ and so we used data on educational level and occupation as imperfect surrogates. The data showed that patients who had higher levels of education had better outcomes.^{30,160,161} There was no significant difference between employed and retired patients.³⁰

Conclusion

Higher levels of deprivation were associated with worse patient outcomes following TKR. A higher attained educational level was associated with better postoperative reported outcomes following THR.

Indication for surgery

Introduction

The most frequent indication for THR and TKR in the UK is OA, which is the most common type of arthritis in developed countries and for which TJR is the only effective therapy in severe cases.² The total number of

hip procedures reported in the UK in 2013 was 89,945. Of these, 80,194 were primary procedures and 9751 were revision replacements.¹⁶⁸ The indication for primary hip replacement was OA in 91% of cases. The total number of knee procedures reported in the UK in 2013 was 91,703. Of these, 85,920 were primary procedures and 5783 were revision replacements. The indication for primary knee replacement was OA in 97% of cases.

Another important disease for THR and TKR indication is RA. RA is a chronic autoimmune disease affecting joints. Severe RA also requires surgical intervention. RA indication for THR and TKR in 2013 was 1%.¹⁶⁸ Since the licensing of biological agents for the treatment of RA, the number of arthroplasties in patients with RA is declining. Association between the indication for lower limb joint replacement (OA or RA) and postoperative score is shown in *Table 27*.

Findings

Knee

The data analysis from EOC showed that patients diagnosed with RA had a better outcome than those diagnosed with OA.¹⁴¹ Clinical outcomes in patients with RA were more than twice as better than in those

TABLE 27 Associations found between indication for OA or RA and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	RA better outcome	Adjusted OR _{total OKS} : 2.17 (95% CI 1.02 to 4.60)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	Not significantly (RA better for pain)	Linear model coefficient _{total OKS pain} : 1.75 (95% CI 0.61 to 2.89)	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	No model differences in OA vs. OA + RA analysis	–	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	More joints with OA, worse improvement	Linear model coefficient _(number of joints with OA) : –1.24 (95% CI –1.71 to –0.77)	Judge <i>et al.</i> ¹⁶⁰
Portsmouth and North Staffordshire	Hip	PROMs	6 months with ≥ 30 points in the SF-36	OA worse preoperative radiograph grade, higher improvement	Adjusted OR: 2.15 (95% CI 1.17 to 3.93)	Judge <i>et al.</i> ¹⁴³
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	OA, superomedial, medial or concentric disease, had worse outcomes	Linear model coefficient _(superomedial, medial or concentric) : –1.44 (95% CI –2.79 to –0.09)	Judge <i>et al.</i> ¹⁶¹
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	More joints with OA, worse improvement	Linear model coefficient: –1.11 (95% CI –1.48 to –0.74)	Judge <i>et al.</i> ¹⁶¹
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	More joints with surgery, worse improvement	Linear model coefficient: –0.78 (95% CI –1.50 to –0.06)	Judge <i>et al.</i> ¹⁶¹

with primary OA, as indicated by the 6-month PASS score (OR 2.17, 95% CI 1.02 to 4.60), and patients with RA had a better PASS pain score than those with OA (OR 2.33, 95% CI 1.03 to 5.29). There was no difference in PASS function scores between patients with RA and OA (OR 1.56, 95% CI 0.73 to 3.31). However, the result using the continuous OKS was not statistically significant (see *Table 27*).

Hip

We collected data on the number of joints affected by OA, apart from the hip.¹⁶⁰ The adjusted multivariable analysis showed that, for each additional joint affected by OA, the 12-month OHS fell by 1.24 units (95% CI 0.77 to 1.71 units) (see *Table 27*).

Baseline data on preoperative radiographic severity were collected from OA patients in two health districts in England (Portsmouth and North Staffordshire).¹⁴³ Radiographic grades (Croft grading system) of hip for surgery were grouped as 0–3, 4 and 5. Patients with worse preoperative radiographic grades had greater improvement (OR 2.15, 95% CI 1.17 to 3.93). In another study we collected data from preoperative anteroposterior radiographs of the pelvis and its intra-articular pattern of disease distribution (superolateral, superomedial, medial or concentric).¹⁶¹ The pattern of OA was strongly associated with the outcome. Patients with superomedial, medial or concentric disease presented with a lower improvement than those with a superolateral pattern of OA or no reduction in joint space. Furthermore, the number of additional joints affected by arthritis and the number of joint replacements was also related to worse outcomes. Therefore, for each extra joint affected by OA, the 12-month OHS was reduced by 1.11 units (95% CI 0.74 to 1.48 units), and for each extra joint with surgery the OHS was reduced by 0.78 units (95% CI 0.06 to 1.50 units).

Conclusion

Most of the patients undergoing lower limb joint replacements have been diagnosed with OA. In only one study did patients diagnosed with RA have better outcomes than those diagnosed with OA.¹⁴¹ We adjusted the model by the confounding factor age, as the patient population diagnosed with RA is younger than that diagnosed with OA.^{141,169} Similar results were also observed for OKS pain score. The other studies were related to OA patterns. These studies unveiled better TKR outcomes in patients with a superolateral pattern of OA or no reduction in joint space, any or a low number of other joints affected by OA, as well as any or a low number of previous interventions on other joints. In summary, fewer baseline OA complications were related to better outcomes.

American Society of Anesthesiologists/comorbidities

Introduction

The American Society of Anesthesiologists (ASA) status classification system is a standard measure of fitness for surgery, scored from 1 (normal, healthy) to 6 (brain-dead patients whose organs are being removed for donor purposes). Our studies included patients up to ASA 4 (severe systemic disease that is a constant threat to life).

In addition, we considered some coexisting diseases that could affect surgery outcomes. We take into account DVT and PE, urinary tract infection, other musculoskeletal diseases and neurological, respiratory, cardiovascular, renal or hepatic disease. ASA status and number of comorbidities were analysed in several cohorts, listed in *Table 28*. Some of these coexistent diseases were also related to BMI (see *Tables 18* and *20*).

Findings

Knee

The data analysis in EOC cohort study showed that the ASA grade was not statistically significant when normal healthy patients (ASA 1) or those with severe systemic disease (ASA 3) were compared with patients with mild systemic disease (ASA 2).¹⁴¹ On the other hand, the KAT cohort study¹⁴² showed that a worse ASA grade was linked to a worse OKS outcome (see *Table 28*).¹⁴²

TABLE 28 Associations found between ASA score and number of extra comorbidities and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	ASA NS	Adjusted OR _{total OKS (ASA 1 vs. 2)} : 1.30 (95% CI 0.81 to 2.08)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	ASA NS	Linear model coefficient _{total OKS (ASA 1)} : 1.00 (95% CI -0.29 to 2.29)	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	Worse ASA grade worse outcome	Linear model coefficient _(ASA 3/4 vs. ASA 1) : -2.6 (95% CI -4.1 to -1.1)	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS	Hip	PROMs	1- to 5-year OHS	Increasing number of comorbidities associated with worse outcome	Δ linear model coefficient _(number of comorbidities) : -0.90 (95% CI -1.27 to -0.54)	Judge <i>et al.</i> ¹³⁹
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	ASA NS	Linear model coefficient _(number of joints with OA) : -0.56 (95% CI -3.24 to 2.12)	Judge <i>et al.</i> ¹⁶⁰
EUROHIP	Hip	PROMs	12-month WOMAC score	Continuous no effect	$p = 0.13$	Judge <i>et al.</i> ³⁰

NS, not significant.

Note

ASA status: a standard measure of fitness for surgery, scored from 1 (normal, healthy) to 4 (severe systemic disease that is a constant threat to life).

Hip

In the EUROHIP cohort, there was no statically significant association between ASA grade and change in WOMAC score at 12 months: the median OHS for ASA 1 patients was 40.6 (95% CI 35.2 to 46.0), for ASA 2 patients was 35.4 (95% CI 32.8 to 38.1), for ASA 3 patients was 33.3 (95% CI 27.8 to 38.9) and for ASA 4 patients was 38.5 (95% CI 30.9 to 46.2) ($p = 0.13$).³⁰ We did not find differences between preoperative ASA grades and the OHS 12 months after hip surgery.¹⁶⁰

In EPOS, the number of coexistent diseases was statistically significantly associated with outcome, that is, for each additional preoperative disease the OHS between 1 year and 5 years was reduced by 0.90 units (95% CI 0.54 to 1.27 units)¹³⁹ (Figure 24).

Conclusion

We can conclude that ASA grade is not an important predictor of outcomes, although healthier preoperative conditions could relate to better outcome. On the other hand, the number of other coexisting diseases is a predictor of worse outcome, although the effect size was small and it would be necessary to identify the weight for each illness in the final outcome.

Anxiety/depression**Introduction**

We employed several scoring systems to assess mental health. For instance, patients from EUROHIP, the EOC, St Helier and EPOS cohorts completed a SF-36 questionnaire. This tool measures quality of life through eight health concepts, from which we selected mental health (psychological distress and psychological well-being).^{170,171} In the KAT cohort, quality of life was assessed using a subset of those in

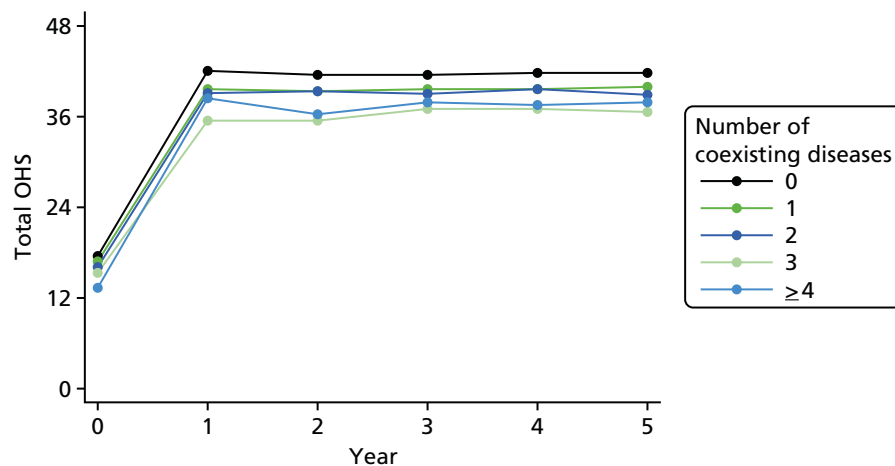


FIGURE 24 Change in OHS over time, stratified by number of coexisting diseases, from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

the SF-36 (SF-12). Both questionnaires score from 0 (the worst possible health) to 100 (best possible health). In addition, we estimated anxiety and depression among EOC participants using the EQ-5D.¹⁴⁸

Findings

Knee

In the EOC cohort, those patients with moderate or extreme anxiety and/or depression presented a worse TKR outcome, that is, worse EQ-5D scores were related to worse outcome using the continuous OKS as outcome.¹⁴¹ We observed lower OKS in anxious or depressed patients, with moderate anxiety or depression being associated with a reduction of 0.85 units (95% CI 0.03 to 1.68 units), and extreme anxiety or depression a reduction of 2.21 units (95% CI 0.09 to 4.34 units), compared with non-anxious or depressed patients. In the KAT cohort, the SF-12 questionnaire was used and demonstrated that worse mental health is found in those with poor outcomes.¹⁴² These results highlight the clinical relevance of mental health in relation to outcome.

Hip

In the EPOS cohort, lower preoperative SF-36 mental health scores were associated with reduced postoperative improvement in OHS between 1 and 5 years (*Table 29*).¹³⁹ The differences in achieved postoperative OHS among categorised mental health levels were not substantial, although they were statistically significant when we followed up between 1 and 5 years (*Figure 25*). We observed the same results when we followed up patients at 12 months and we used the same questionnaire (coefficient 0.76 units, 95% CI 0.18 to 1.33 units).¹⁶⁰ Finally, we obtained a similar increase in the OHS outcome (coefficient 0.59 units, 95% CI 0.24 to 0.94 units) in patients from the EPOS and EUROHIP cohorts followed up at 12 months.¹⁶¹

Conclusion

Our results show an association between preoperative mental health and the PROM. Therefore, patients having worse preoperative mental health were more likely to have worse postoperative outcome scores. This is concordant with other studies in the literature which used detailed measures of mental health^{28,29,172,173} or more specific ones, for example the Beck Depression Inventory.¹⁷⁴ However, there were no clinically important differences among mental health categories and the outcome found after surgery.

TABLE 29 Associations found between mental health and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
EOC	Knee	PROMs	6-month PASS score	NS. Anxiety/ depression worse outcome	Adjusted OR _{total OKS (extremely anxious/depressed)} : 0.70 (95% CI 0.42 to 1.18)	Judge <i>et al.</i> ¹⁴¹
EOC	Knee	PROMs	6-month OKS	Anxiety/ depression worse outcome	Linear model coefficient _{total OKS (extremely anxious/depressed)} : -2.21 (95% CI -4.34 to -0.09)	Judge <i>et al.</i> ¹⁴¹
KAT	Knee	PROMs	12-month OKS	Worse mental health score associated with poor outcome	Linear model coefficient _{SF-12 (10 units)} : 0.9 (95% CI 0.4 to 1.3)	Sánchez-Santos <i>et al.</i> ¹⁴²
EPOS	Hip	PROMs	1- to 5- year OHS	Lower SF-36 mental health score associated with poor outcome	Δ linear model coefficient _{SF-36 (10 units)} : 0.76 (95% CI 0.46 to 1.07)	Judge <i>et al.</i> ¹³⁹
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	Lower SF-36 mental health score associated with poor outcome	Linear model coefficient _{SF-36 (10 units)} : 0.76 (95% CI 0.18 to 1.33)	Judge <i>et al.</i> ¹⁶⁰
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	Lower SF-36 mental health score associated with poor outcome	Linear model coefficient _{SF-36 (10 units)} : 0.59 (95% CI 0.24 to 0.94)	Judge <i>et al.</i> ¹⁶¹

NS, not significant.

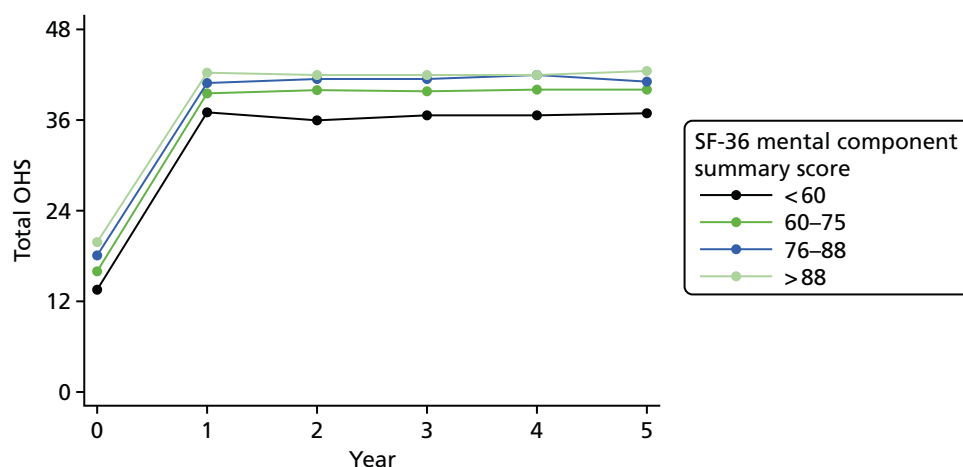


FIGURE 25 Change in OHS over time, stratified by SF-36 mental health categories from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

Clinical, surgical and drug predictors

Introduction

Surgical variables are not usually included in models assessing lower limb joint replacement. We considered clinical examination findings, that is, the presence of fixed flexion deformity, joint deformity (valgus, varus, no deformity) and presence of anterior cruciate ligaments (ACLs) and posterior cruciate ligaments. We also considered variables extrinsic to the patient, that is grade of operating surgeon (consultant, associate specialist staff, registrar and senior house officer) and the grade of senior surgeon present at the operation (consultant, associate specialist staff and registrar).¹⁴² Furthermore, information about the surgical approach (anterolateral and posterior) and patient position (supine and lateral) was also collected.¹³⁹

Findings

Knee

Clinical variables were included in a general linear model to predict OKS patient outcome.¹⁴² We found better outcomes in patients with fixed flexion deformity than in those with no deformity (linear model coefficient 1.5, 95% CI 0.6 to 2.4). In addition, we found better outcomes in patients with varus deformity than in those with no deformity and those with an absent preoperative ACL than in those with an intact ACL (linear model coefficient 1.5, 95% CI 0.0 to 3.0).

Hip

In EUROHIP, we observed increasing differences between the 12-month follow-up and baseline scores from K/L grade 2 to K/L grade 4, that is, the median WOMAC score for K/L grade 2 was 29.7 units (95% CI 22.6 to 36.8 units), K/L grade 3 was 34.4 units (95% CI 31.3 to 37.4 units) and K/L grade 4 was 38.5 units (95% CI 35.4 to 41.7 units); differences among groups were statistically significant ($p = 0.03$). Although this effect seems to increase also from grade 2 to grade 0, the absolute numbers of patients in groups 0 and 1 (six and five patients, respectively) did not let us affirm that K/L grade was a U-shaped curve, with group 2 being the worst. We also analysed K/L grade in patients from four cohorts, being K/L grades 1 and 2 versus K/L grade 4, and found this not to be statistically significant, that is 0.43 units (95% CI -2.75 to 3.62 units) and 0.86 units (95% CI -4.73 to 6.44 units), respectively (Table 30). We observed better outcomes in hip replacement associated with a larger offset size (offset of ≥ 44 mm).

Change among femoral offset size categories had a small statistically significant difference in the postoperative OHS achieved (Figure 26). We found a significant interaction between offset size and sex, with the effect limited to women.¹³⁹

We also identified the effect of surgical approach as significant, with the posterior approach having better outcomes than anterolateral; that is, there is a difference in the OHS at 12 months of 2.2 units (95% CI 1.1 to 3.30 units).¹³⁹ This result was very similar to those presented by Judge *et al.*;¹⁶¹ that is, the difference at 12 months for the OHS was 2.42 units (95% CI 0.44 to 4.39 units).

Finally, we found that patients from GPRD receiving hormone replacement therapy had a lower risk of joint (hip and knee replacements) revision surgery after 6 months (adjusted HR 0.62, 95% CI 0.41 to 0.94) and after 12 months (adjusted HR 0.48, 95% CI 0.29 to 0.78)¹⁷⁵ (Figure 27 and see Table 30).

Conclusion

Clinical factors have demonstrated their importance in TKR as predictors of OKS outcomes.¹⁴² Among them, we identified fixed flexion deformity, preoperative valgus/varus deformity and preoperative damaged ACL. We unveiled an association between femoral offset size and THR, with femoral offsets > 44 mm related to better outcomes,^{139,161} although this finding may happen only in women.¹³⁹ In addition, we discovered that a posterior approach may have better outcomes than an anterolateral approach.^{139,161} Furthermore, K/L status was not a good predictor of the outcome.^{30,160}

TABLE 30 Associations found between surgical and drug variables and postoperative score

Cohort	Hip/knee	PROMs/ revision	Outcome	Association found	For example	Reference
KAT	Knee	PROMs	12-month OKS	Fixed flexion deformity, varus deformity, absent preoperative ACL were associated with better outcome	Linear model coefficient _{fixed flexion deformity} : 1.5 (95% CI 0.6 to 2.4) Linear model coefficient _{no valgus/varus deformity} : -1.6 (95% CI -3.0 to -0.3) Linear model coefficient _{preoperative ACL absent} : 1.5 (95% CI 0.0 to 3.0)	Sánchez-Santos <i>et al.</i> ¹⁴²
EUROHIP	Hip	PROMs	12-month WOMAC score	K/L grade change increases from 2 to 4	K/L grade 2 worse outcome; $p = 0.03$	Judge <i>et al.</i> ³⁰
EPOS, EUROHIP, EOC and St Helier	Hip	PROMs	12-month OHS	NS K/L grade	Linear model coefficient _(K/L 1) : 0.43 (95% CI to 2.75 to 3.62) Linear model coefficient _(K/L 2) : 0.86 (95% CI -4.73 to 6.44)	Judge <i>et al.</i> ¹⁶⁰
EPOS	Hip	PROMs	1- to 5-year OHS	Femoral offset size larger = better outcome	Δ linear model coefficient _(offset) : 0.17 (95% CI 0.06 to 0.28)	Judge <i>et al.</i> ¹³⁹
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	Femoral offset size larger = better outcome	Linear model coefficient _(offset) : 0.18 (95% CI 0.03 to 0.32)	Judge <i>et al.</i> ¹⁶¹
EPOS	Hip	PROMs	12-month OHS	Posterior approach = better outcome	Linear model coefficient: 2.2 (95% CI 1.1 to 3.3)	Judge <i>et al.</i> ¹³⁹
EPOS and EUROHIP	Hip	PROMs	12-month OHS and WOMAC score	Posterior approach = better outcome	Linear model coefficient: 2.42 (95% CI 0.44 to 4.39)	Judge <i>et al.</i> ¹⁶¹
GPRD	Hip and knee	Revision	10-year implant survival	HRT reduces revision rates	HR _{HRT \geq 6 months} : 0.62 (95% CI 0.41 to 0.94)	Prieto- Alhambra <i>et al.</i> ¹⁷⁵

HRT, hormone replacement therapy; NS, not significant.

Conclusion

We have identified a number of predictors of outcome for both THR and TKR. With the exception of baseline Oxford score, the majority of the predictors have statistically significant, but clinically small, effects in isolation. We have identified that the predictors of PROMs may be different from the predictors of revision and complications of surgery, this was best evidenced by the data relating to age and sex. We found that increasing age was associated with a lower risk of knee and hip revision,⁹⁷ whereas in knee studies younger patients had a higher likelihood of improvement and the older patients a worse outcome.¹⁴² Furthermore, men had a higher risk of TKR revision,⁹⁷ whereas women had worse PROMs after TKR.¹⁴¹ Therefore, in order to understand the risks and benefits of lower limb arthroplasty, results from PROMs outcomes would complement those obtained from revision studies.

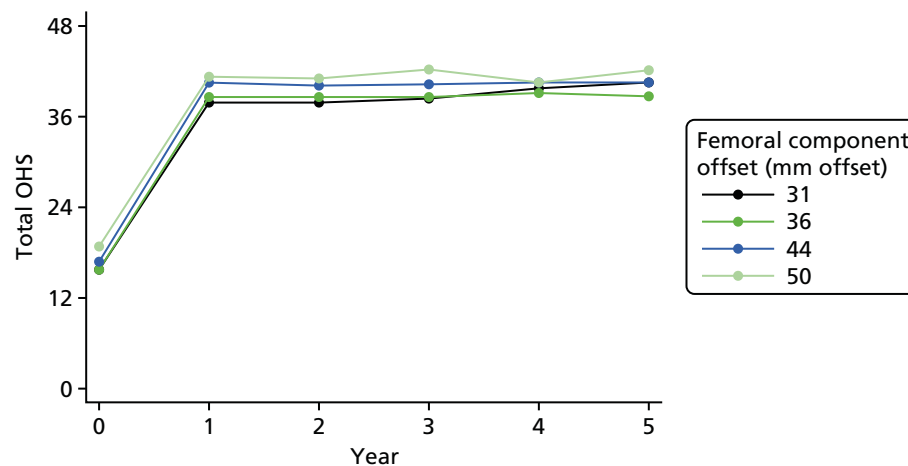


FIGURE 26 Change in OHS over time, layered by femoral component offset (mm offset) categories, from Judge *et al.*¹³⁹ This is an open-access article distributed under the terms of the Creative Commons Attribution Noncommercial license, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the licence.

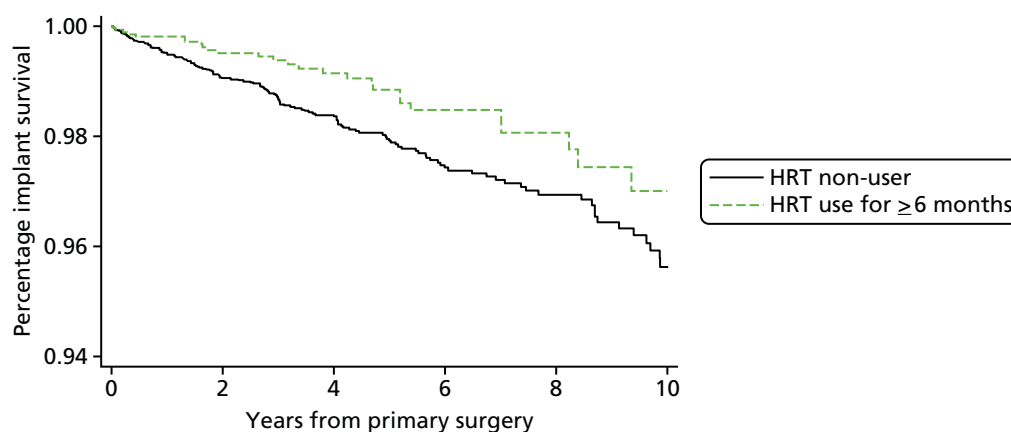


FIGURE 27 Kaplan–Meier estimates of probability of revision surgery according to hormone replacement therapy use. Reproduced from *Annals of the Rheumatic Diseases*, Hormone replacement therapy and mid-term implant survival following knee or hip arthroplasty for osteoarthritis: a population-based cohort study, Prieto-Alhambra D, Javadi MK, Judge A, Maskell J, Cooper C, Arden NK, COAST Study Group. vol. 74, pp. 557–63, 2015, with permission from BMJ Publishing Group Ltd.¹⁷⁵ HRT, hormone replacement therapy.

When comparing predictors of THR and TKR, we found that the results were broadly similar for the majority of the risk factors assessed. We therefore surmise that the features of patients affecting TKR and THR outcomes would be similar.

Having identified a number of predictors of outcome, it is now important to combine them into a model to assess their independence and to search for interactions. This will be described in the next section.

Developing a predictive model

Statistical tool for predicting poor outcomes following total hip or knee replacement surgeries

The final work of work package 2 was to combine the identified risk factors to develop a prognostic tool to predict poor outcome following TKR and THR surgeries. We published a number of studies and combined

them into the statistical tools (for both hip and knee) for each surgery category. The papers are currently under internal review and waiting for external validation in the pragmatic cohort study Clinical Outcomes in Arthroplasty Study (COAST). Although the previous research provides information on different type of predictors, it is important to understand how the interplay of these predictors can play a role in the development of poor functional and pain outcome after THR/TKR.

We aimed to combine risk factors into clinically meaningful tools for stratifying patients for THR and TKR surgeries. For this purpose we collected and used the data from EPOS, EUROHIP and KAT.

Hip predictive tool

The work for the development of the prognostic tool to predict the outcome following THR was done and summarised in the article by Judge *et al.*,¹³⁸ as described in the next section.

Patient-reported outcomes following primary hip replacement surgery: development and internal validation of a prognostic tool¹³⁸

Following identification and confirmation of risk factors within this work package we aimed to develop a clinical risk prediction tool for both hip and knee arthroplasty patients. Such tools will be beneficial for clinical decision-making and patients' expectation of their surgery.

The recent literature describes combining of data from multiple risk factors and, thus, including the broad range of predictors in a prognostic model.^{80,81} This includes patient-related risk factors such as age, sex, education, obesity and mental health status; clinical predictors such as preoperative level of pain and function, indication for surgery, coexisting conditions and radiographic (grade) variables; and surgery-related risk factors, such as femoral component offset. In this work we aimed to develop a similar prognostic model allowing inclusion of a number of risk factors to predict pain and function following hip replacement surgery.

We collected data from prospective cohorts of patients receiving primary THR for OA: EUROHIP and EPOS. From EUROHIP we used the data from 845 patients and from EPOS we used the data from 1247 patients. By combining the data from these large cohorts we took into account a comprehensive range of both already-investigated risk factors and novel predictors.

As OHS was one of the predictors included for the analysis, we collected OHS questionnaires preoperatively and 1 year postoperatively. OHS was used as a primary outcome measure. In the EUROHIP cohort, outcome measures were predominantly collected via WOMAC, in contrast to EPOS, which collected OHS. To derive OHSs from WOMAC we used the truncated regression model for 110 patients in whom both OHS and WOMAC data are complete. Using this model we could predict WOMAC score at baseline ($R^2 = 75.5\%$) and at 1 year ($R^2 = 63.4\%$).

Predictor variables we included in this work are represented in *Table 31*.

In the EPOS study, patients completed the SF-36. The SF-36 is an instrument that measures quality of life in eight domains. One of these domains is the mental health component, which we have selected for this work. In SF-36 the lowest score, 0, is indicative of the worst possible health and the highest score, 100, of the best possible health. We obtained the fixed flexion range of motion variable in degrees from the Charnley Modification of D'Aubigné–Postel Grade questionnaire.¹⁷⁶ From comorbidities we collected information on coexistent diseases, such as DVT, PE, urinary tract infection, other musculoskeletal disorders, neurological, renal, cardiovascular, respiratory or hepatic disease or treatment for other medical conditions. We have also collected detailed intraoperative information: grade of operating surgeon and patient position. From the implant information we obtained and used data such as implant material (stainless steel or ceramic), femoral head size in millimetres and the femoral component size (offset).

TABLE 31 List of predictors available within the EUROHIP and EPOS cohorts, and distribution of the extent of missing data in each study

Variable	Cohort, data available (%)	
	EUROHIP (<i>n</i> = 845)	EPOS (<i>n</i> = 1247)
Patient variables		
Preoperative and 12-month follow-up OHS ^a	100	100
Age (years)	99	99
Sex	95	100
BMI (kg/m ²)	92	95
Employment/occupation	98	100
Education	88	0
Mental health ^b	98	68
ASA grade	87	0
Years of hip pain	99	0
Care for someone else	99	0
Fixed flexion	0	92
Preoperative expectations	100	0
Preoperative comorbidity	0	69
Preoperative medication use	100	0
Analgesic/NSAIDs	99	92
Radiographic variables		
K/L grade	93	0
Pattern of OA	87	0
Prosthesis type	97	0
Number of joints OA	100	0
Number of joints surgery	100	0
Number of sites osteophytes	85	0
Surgical variables		
Grade of operator	0	100
Surgical approach	0	73
Patient's position	0	100
Femoral component size (mm offset)	0	100
Femoral head	0	100
Head size	0	100
Duration of operation	0	95

NSAID, non-steroidal anti-inflammatory drug.

a The WOMAC score from EUROHIP was used.

b The EQ-5D anxiety/depression score was used in EUROHIP and the SF-36 mental component summary score was used in EPOS.

Note

ASA status: a standard measure of fitness for surgery, scored from 1 (normal, healthy) to 4 (severe systemic disease that is a constant threat to life).

From the EUROHIP cohort we obtained data on home circumstances, employment, education, duration of pain in the affected hip, the number of preoperative expectations of surgery¹⁶² and other joints affected by OA, as well as other surgeries in other joints. Medications (prior to surgery) that were considered relevant to the study included analgesic/non-steroidal anti-inflammatory drugs (NSAIDs), bisphosphonates, medications for heart disease, anticoagulants, antidepressants, bronchodilators and antidiabetic drugs. We obtained EQ-5D scores for the patient's health state today, mobility, self-care, usual activities and pain and anxiety. We obtained the data collected from baseline radiographs of the pelvis taken in the anteroposterior view. Radiographic variables included the standard K/L grade (0–4)^{162,176–179} and the intra-articular pattern of disease distribution (superolateral, superomedial, medial or concentric). Osteophyte size in the superior–femoral, superior–acetabular, inferior–femoral and inferior–acetabular regions was recorded. For osteophytes with moderate sizes, we created an ordinal variable of the number of sites. From intraoperative data we also obtained records from surgical teams on patients' height and weight, prosthesis type and ASA status, the last being a standard measure of fitness for surgery, scored from 1 (normal, healthy) to 4 (severe systemic disease that is a constant threat to life).

Methodology for combining variables

Because cohorts contained a different set of confounders, combining the two studies resulted in a high proportion of missing data. *Appendix 3* shows the variables available in all cohorts.

Previously methodology has been developed to combine data from multiple sources.^{80,81,178,179} For instance, Heymans *et al.*¹⁷⁸ described the method multivariate imputation by chained equations (MICE), based on a Bayesian approach,¹⁷⁶ in which data from three controlled trials were combined to identify risk factors of chronic low back pain. Jackson *et al.*⁸⁰ used the multiple imputation methods within Bayesian graphical modelling and described the associations between low birthweight and air pollution, thus incorporating adjustments for crucial confounders that were not available within the individual data set. These two approaches have been further compared in both simulation¹⁸⁰ and cases studies.¹⁸¹ It was concluded that, even when the proportion of missing data is high, the performance of both approaches is similar. Moreover, they both correct most of the bias with a non-hierarchical (simple) data structure.

In our work we used MICE to combine the data sources as we were more familiar with this methodology, it is easier to implement in standard statistical software and it takes less time to run the models.

Use of the multiple imputation method allows the assumption that the data are missing at random. This is plausible, as the reason behind the missing data is that not all covariates were collected in each study.¹⁷⁸ At first, we created 25 copies of the data set. In each copy missing values were replaced by imputed values,^{176,182,183} with 20 cycles of regression switching. Imputations were made by drawing from the posterior predictive distribution of each variable that required imputation. We included all of the covariates together with the outcome variable (12-month postoperative OHS) in the imputation model, as this carries information about the missing values of predictors.¹⁷⁶ Prior to imputation we transformed continuous variables so that they were approximately normally distributed. For the imputation of continuous variables we used the linear regression method, whereas for categorical variables we used logistic, ordinal and multinomial regression. At the second stage we fitted a statistical model to each of the imputed data sets separately. The results were then averaged to obtain a single estimate of the association. We calculated standard errors using Rubin's rule, as it accounts for the variability between results of imputed data sets and reflects the uncertainty associated with imputing missing data.¹⁷⁶

Although the methodology we used has been tested and used before and proved to be transparent, and it is relatively easy to implement within the standard statistical software, its use required us to make the assumption that data were missing at random. This was plausible in the context of this study because the reason for the missing data was that the variables were not collected. This was further supported by inclusion of a wide range of covariates to ensure that sufficient predictors were included to recover missing data for missing information. We are aware, however, that there were too many missed data in this study. Even if the missing at random assumption is valid, with increased proportion of missing variables the

reliability of the regression coefficient may diminish.¹⁸² However, we are reassured by the previous simulations and case studies in which the methods have been shown to be effective, even with extreme numbers (> 90%) of missing data.^{180,181}

Statistical methods used

For statistical analysis we used Stata version 12.1 (StataCorp LP, College Station, TX, USA). We used analysis of covariance (ANCOVA) to identify predictors of the 12-month follow-up OHS, adjusting for preoperative OHS. In order to model non-linear relationships for continuous variables we used linear splines. We use MICE in order to combine data from two studies and adjust for a wider range of variables.^{80,81} In the imputation model we included all of the listed covariates together with the outcome variable (12-month postoperative OHS). Before imputation we log-transformed continuous variables so that they were approximately normally distributed. We created 25 imputed data sets for missing data by using MICE in Stata. The final regression model included all predictor variables and was fitted to each imputed data set. This was then averaged for overall estimated associations. We applied the automatic backward selection per 200 bootstrap samples of imputed data sets. The variables retained were those consistently selected for at least 70% of the time.

We assessed the performance of the tool by using the calibration and discrimination methods.^{184,185} By calibration we could judge how close the predicted OHS was to the observed OHS for each tenth of predicted score in 10 equally sized groups. By discrimination we assessed the measure of variation and for this we used R^2 . In total, 200 bootstrap samples with replacement and combined with multiple imputations were used to estimate overoptimism in the predictive ability of the model, and obtain bias-corrected estimates of R^2 . We compared R^2 from models developed in 200 bootstrap samples with R^2 in the same models applied to the original sample. As previously described in this chapter, a change in OHS by fewer than 5 units is described as clinically not significant. We used this knowledge to finally test the ability of the predictive tool to identify patients with the worst possible scores after their THR (i.e. identify the patients with < 5-unit changes in OHS). We then used a logistic regression model and assessed the discriminatory ability of the tool by calculating the area under the ROC curve.

Results

From the selected patients, only 63.7% in EUROHIP and 87.1% in EPOS had completed both a baseline and 1-year follow-up OHS. We observed discrepancies between the patients who did not answer the 1-year follow-up questionnaire and those who responded (*Table 32*). In the EUROHIP cohort, responders had better baseline pain and function and EQ-5D mental health scores, lower educational level and higher levels of obesity as well as greater ASA scores. In the EPOS study we found differences in younger responders and also in those employed and with a better preoperative SF-36 score.

The mean age of patients in EUROHIP was 65.7 years (SD 10.6 years) and 68 years (SD 10.7 years) in EPOS. There were 55.2% and 62.8% women in EUROHIP and EPOS, respectively. BMI was similar across the two cohorts, with a mean BMI of 27.8 kg/m² (SD 4.4 kg/m²) in EUROHIP and 27.3 kg/m² (SD 4.9 kg/m²) in EPOS.

Table 33 displays the predictors that have been found statistically significant.

For example, higher baseline OHS (better pain and function) was associated with an increased follow-up OHS, which is indicative of better pain and function outcomes. The effect of age was non-linear and there was a threshold effect. Patients aged > 75 years had worse outcomes. Worse outcomes were also observed in patients with an increased BMI, lower levels of education and lower baseline SF-36 scores. The pattern of OA was found to be an important predictor of outcome as a radiographic variable. Patients with superomedial, medial or concentric OA had worse outcomes than those with a superolateral pattern of disease or with no reduction in joint space. Worse outcomes were also linked with having had previous surgeries in other joints and having had arthritis in other joints. A posterior surgical approach and femoral component offset of ≥ 44 mm were associated with significantly better outcomes.

TABLE 32 Clinical baseline data between responders and non-responders

Variables	Cohort			
	EUROHIP		EPOS	
	Responders (n = 908)	Non-responders (n = 419)	Responders (n = 987)	Non-responders (n = 437)
ASA grade, n (%)				
Fit and healthy	123 (15.5)	86 (22.5)		
Asymptomatic no restriction	505 (63.7)	214 (56.0)		
Symptomatic minimal	160 (20.2)	77 (20.2)		
Severe restriction	5 (0.6)	5 (1.3)		
K/L grade, n (%)				
0	6 (0.7)	0 (–)		
1	5 (0.6)	1 (0.4)		
2	26 (3.2)	6 (2.4)		
3	394 (49.2)	116 (46.4)		
4	370 (46.2)	127 (50.8)		
Number of coexisting diseases, n (%)				
0			136 (31)	295 (30)
1			147 (33)	351 (35)
2			99 (22)	216 (22)
3			43 (10)	97 (10)
4			16 (4)	31 (3)
	WOMAC score		OHS	
Pain score, mean (SD)	54.5 (17.6)	57.7 (18.1)	16.1 (8.2)	16.5 (7.6)

Note

ASA status: a standard measure of fitness for surgery, scored from 1 (normal, healthy) to 4 (severe systemic disease that is a constant threat to life).

Internal validation and model performance

To validate the model internally we used bootstrapping to the imputed data sets. To ensure that the risk factors are replicated in other external validation studies we ensured that all predictors identified were those consistently selected across the 200 bootstrap resamples at least 70% of the time. The performance of the model was assessed by calibration and discrimination (see *Table 34*). Calibration of the predicted 12-month postoperative OHS was good, except in the lowest deciles of OHS, in which case the model overestimated the predicted score. Calibration of this predicted change in OHS was very good across all deciles of change in OHS. The model showed a discriminatory ability with a bias-corrected R^2 of 23.1%. We assessed the performance of the model and found the calibration of the predicted 12-month OHS to be good. The exception was the lowest deciles of OHS, where we found that the developed model overestimated the predicted score. The performance of the model is described in *Table 34*.

We also calculated the predicted the absolute change in OHS by subtracting the predicted 12-month score from the observed preoperative score. It was observed that the calibration of this absolute predicted change was very good. Importantly, the model also showed good discriminatory ability, with a bias-corrected R^2 of 23.1%.

TABLE 33 Results of linear regression (ANCOVA) model to identify predictors of OHS at 12 months' follow-up

Variable	% included	Coefficient (95%)	p-value
Patient variables			
Baseline total OHS (10 units)	100.00	2.23 (1.68 to 2.79)	< 0.001
Age < 75 years (per 10 years)	75.00	0.26 (−0.19 to 0.71)	0.252
Age ≥ 75 years (per 10 years)	82.50	−2.00 (−3.55 to −0.45)	0.011
BMI (kg/m ²) (5 units)	90.00	−0.66 (−1.10 to −0.22)	0.003
Education			
None		0	
College/university	100.00	2.08 (0.59 to 3.57)	0.006
SF-36 mental component summary score preoperative (10 units)	100.00	0.59 (0.24 to 0.94)	0.001
Radiographic variables			
Pattern of OA			
Superolateral		0	
Superomedial/medial/concentric	89.00	−1.44 (−2.79 to −0.09)	0.037
No reduction	2.50	−1.72 (−4.80 to 1.35)	0.272
Number of joints with OA	100.00	−1.11 (−1.48 to −0.74)	< 0.001
Number of joints with surgery	85.00	−0.78 (−1.50 to −0.06)	0.034
Surgical variables			
Surgical approach			
Anterolateral		0	
Posterior	100.00	2.42 (0.44 to 4.39)	0.017
Femoral component size (mm offset)	82.50	0.18 (0.03 to 0.32)	0.016

Notes

Variables included in the final regression model are those that are retained in at least 70% of the 200 bootstrap backward selection regression models. Age is represented as a linear spline, with a knot at 75 years of age.

Development of a risk prediction tool

We have developed a web-based tool that can be used to inform patients of the likely outcomes following their surgery. This web-based tool (*Figure 28*) uses the predictive equation:

$$\begin{aligned}
 \text{OHS}_{12 \text{ months}} = & [2.23 \times \text{OHS}_{\text{baseline}(10 \text{ units})}] + [0.26 \times \text{age}_{<75 \text{ years}(10 \text{ units})}] - [2.00 \times \text{age}_{\geq 75 \text{ years}(10 \text{ units})}] \\
 & - [0.66 \times \text{BMI}_{(5 \text{ units})}] + (2.08 \times \text{education}_{\text{college/university}}) \\
 & + [0.59 \times \text{mental health}_{(10 \text{ units})}] - (1.44 \times \text{OA pattern}_{\text{superomedial/medial/concentric}}) \\
 & - (1.72 \times \text{OA pattern}_{\text{no reduction}}) - (1.11 \times \text{OA joints}) - (0.78 \times \text{surgery joints}) \\
 & + (2.42 \times \text{posterior surgical approach}) + (0.18 \times \text{femoral component size}).
 \end{aligned} \tag{1}$$

We tested the ability of the tool to identify the patients with the worst possible outcomes of surgery. The worst outcomes are defined as those THR patients whose preoperative score does not improve for at least 5 units over the 12 months. For 89 patients who met the criteria of the worst possible outcome, the discriminatory ability of the tool was good, with an area under the ROC curve of 0.78 (95% CI 0.73 to 0.82) (*Figure 29*). The tool had a sensitivity of 72% and a specificity of 73%. The calibration was also reasonable, as indicated by the Hosmer–Lemeshow *p*-value of 0.13.

TABLE 34 Results of calibration and discrimination of the predictive model

Calibration	OHS					
	12 months			Change		
	Observed	Predicted	Ratio	Observed	Predicted	Ratio
Deciles						
1	26.78	31.79	0.84	10.53	11.49	0.92
2	33.88	34.90	0.97	15.42	16.36	0.94
3	35.85	36.30	0.99	18.25	19.08	0.96
4	38.19	37.50	1.02	20.74	20.95	0.99
5	39.55	38.49	1.03	22.77	22.51	1.01
6	41.02	39.41	1.04	23.00	24.02	0.96
7	42.14	40.36	1.04	26.16	25.54	1.02
8	42.74	41.53	1.03	27.98	27.20	1.03
9	43.67	42.90	1.02	29.42	29.00	1.01
10	44.57	45.22	0.99	33.79	31.97	1.06
Discrimination (%)						
R^2	24.00					
Optimism	0.80					
Corrected R^2	23.10					