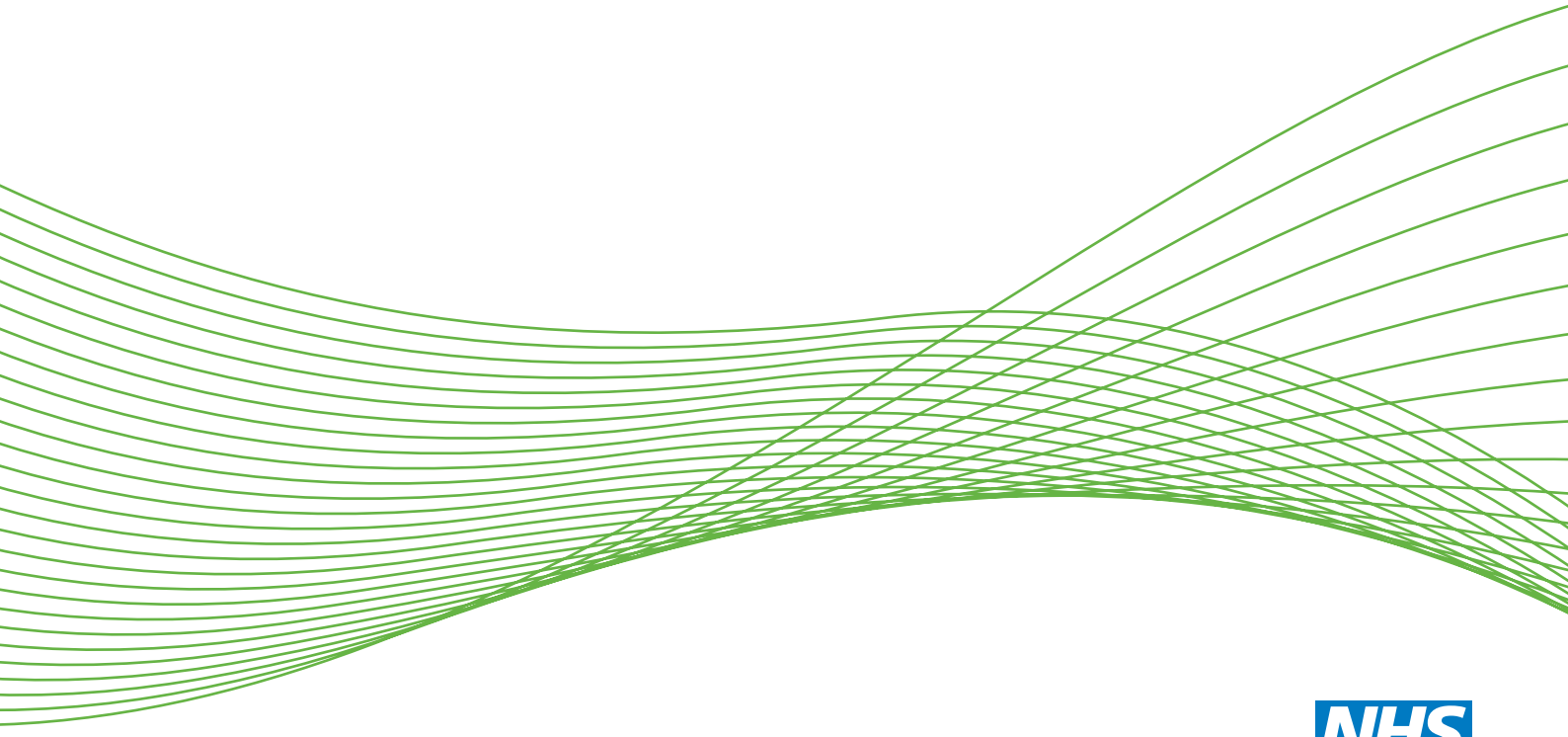
HEALTH TECHNOLOGY ASSESSMENT

VOLUME 19 ISSUE 10 JANUARY 2015
ISSN 1366-5278

Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation

Aileen Clarke, Ruth Pulikottil-Jacob, Amy Grove, Karoline Freeman, Hema Mistry, Alexander Tsertsvadze, Martin Connock, Rachel Court, Ngianga-Bakwin Kandala, Matthew Costa, Gaurav Suri, David Metcalfe, Michael Crowther, Sarah Morrow, Samantha Johnson and Paul Sutcliffe



National Institute for Health Research

DOI 10.3310/hta19100

Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation

Aileen Clarke,[1]* Ruth Pulikottil-Jacob,[1] Amy Grove,[1] Karoline Freeman,[1] Hema Mistry,[1] Alexander Tsertsvadze,[1] Martin Connock,[1] Rachel Court,[1] Ngianga-Bakwin Kandala,[1] Matthew Costa,[1] Gaurav Suri,[1] David Metcalfe,[2] Michael Crowther,[3] Sarah Morrow,[4] Samantha Johnson[1] and Paul Sutcliffe[1]

[1]Warwick Evidence, Warwick Medical School, University of Warwick, Coventry, UK
[2]Warwick Orthopaedics, University Hospitals Coventry and Warwickshire, Coventry, UK
[3]Department of Health Sciences, University of Leicester, Leicester, UK
[4]Oxford Medical School, University of Oxford, Oxford, UK

*Corresponding author

Declared competing interests of authors: Aileen Clarke is a member of the NIHR HTA Editorial Board and the Warwick Medical School receive payment for this work.

Published January 2015
DOI: 10.3310/hta19100

This report should be referenced as follows:

Clarke A, Pulikottil-Jacob R, Grove A, Freeman K, Mistry H, Tsertsvadze A, *et al*. Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation. *Health Technol Assess* 2015;**19**(10).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

NICE TAR and DAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 5.116

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the ISI Science Citation Index and is assessed for inclusion in the Database of Abstracts of Reviews of Effects.

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

Editorial contact: nihredit@southampton.ac.uk

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

Criteria for inclusion in the *Health Technology Assessment* journal

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

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

HTA programme

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

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

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

This report

The research reported in this issue of the journal was commissioned and funded by the HTA programme on behalf of NICE as project number 11/118/01. The protocol was agreed in November 2012. The assessment report began editorial review in July 2013 and was accepted for publication in January 2014. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

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

Abstract

Total hip replacement and surface replacement for the treatment of pain and disability resulting from end-stage arthritis of the hip (review of technology appraisal guidance 2 and 44): systematic review and economic evaluation

Aileen Clarke,[1]* Ruth Pulikottil-Jacob,[1] Amy Grove,[1] Karoline Freeman,[1] Hema Mistry,[1] Alexander Tsertsvadze,[1] Martin Connock,[1] Rachel Court,[1] Ngianga-Bakwin Kandala,[1] Matthew Costa,[1] Gaurav Suri,[1] David Metcalfe,[2] Michael Crowther,[3] Sarah Morrow,[4] Samantha Johnson[1] and Paul Sutcliffe[1]

[1]Warwick Evidence, Warwick Medical School, University of Warwick, Coventry, UK
[2]Warwick Orthopaedics, University Hospitals Coventry and Warwickshire, Coventry, UK
[3]Department of Health Sciences, University of Leicester, Leicester, UK
[4]Oxford Medical School, University of Oxford, Oxford, UK

*Corresponding author Aileen.Clarke@warwick.ac.uk

Background: Total hip replacement (THR) involves the replacement of a damaged hip joint with an artificial hip prosthesis. Resurfacing arthroplasty (RS) involves replacement of the joint surface of the femoral head with a metal surface covering.

Objectives: To undertake clinical effectiveness and cost-effectiveness analysis of different types of THR and RS for the treatment of pain and disability in people with end-stage arthritis of the hip, in particular to compare the clinical effectiveness and cost-effectiveness of (1) different types of primary THR and RS for people in whom both procedures are suitable and (2) different types of primary THR for people who are not suitable for hip RS.

Data sources: Electronic databases including MEDLINE, EMBASE, The Cochrane Library, Current Controlled Trials and UK Clinical Research Network (UKCRN) Portfolio Database were searched in December 2012, with searches limited to publications from 2008 and sample sizes of $\geq$ 100 participants. Reference lists and websites of manufacturers and professional organisations were also screened.

Review methods: Systematic reviews of the literature were undertaken to appraise the clinical effectiveness and cost-effectiveness of different types of THR and RS for people with end-stage arthritis of the hip. Included randomised controlled trials (RCTs) and systematic reviews were data extracted and risk of bias and methodological quality were independently assessed by two reviewers using the Cochrane Collaboration risk of bias tool and the Assessment of Multiple Systematic Reviews (AMSTAR) tool. A Markov multistate model was developed for the economic evaluation of the technologies. Sensitivity analyses stratified by sex and controlled for age were carried out to assess the robustness of the results.

Results: A total of 2469 records were screened of which 37 were included, representing 16 RCTs and eight systematic reviews. The mean post-THR Harris Hip Score measured at different follow-up times (from 6 months to 10 years) did not differ between THR groups, including between cross-linked polyethylene and traditional polyethylene cup liners (pooled mean difference 2.29, 95% confidence interval –0.88 to 5.45). Five systematic reviews reported evidence on different types of THR (cemented vs. cementless cup fixation and implant articulation materials) but these reviews were inconclusive. Eleven cost-effectiveness studies were included; four provided relevant cost and utility data for the model. Thirty registry studies were included, with no studies reporting better implant survival for RS than for all types of THR. For all analyses, mean costs for RS were higher than those for THR and mean quality-adjusted life-years (QALYs) were lower. The incremental cost-effectiveness ratio for RS was dominated by THR, that is, THR was cheaper and more effective than RS (for a lifetime horizon in the base-case analysis, the incremental cost of RS was £11,284 and the incremental QALYs were –0.0879). For all age and sex groups RS remained clearly dominated by THR. Cost-effectiveness acceptability curves showed that, for all patients, THR was almost 100% cost-effective at any willingness-to-pay level. There were age and sex differences in the populations with different types of THR and variations in revision rates (from 1.6% to 3.5% at 9 years). For the base-case analysis, for all age and sex groups and a lifetime horizon, mean costs for category E (cemented components with a polyethylene-on-ceramic articulation) were slightly lower and mean QALYs for category E were slightly higher than those for all other THR categories in both deterministic and probabilistic analyses. Hence, category E dominated the other four categories. Sensitivity analysis using an age- and sex-adjusted log-normal model demonstrated that, over a lifetime horizon and at a willingness-to-pay threshold of £20,000 per QALY, categories A and E were equally likely (50%) to be cost-effective.

Limitations: A large proportion of the included studies were inconclusive because of poor reporting, missing data, inconsistent results and/or great uncertainty in the treatment effect estimates. This warrants cautious interpretation of the findings. The evidence on complications was scarce, which may be because of the absence or rarity of these events or because of under-reporting. The poor reporting meant that it was not possible to explore contextual factors that might have influenced study results and also reduced the applicability of the findings to routine clinical practice in the UK. The scope of the review was limited to evidence published in English in 2008 or later, which could be interpreted as a weakness; however, systematic reviews would provide summary evidence for studies published before 2008.

Conclusions: Compared with THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was dominated by THR. Similar results were obtained in the deterministic and probabilistic analyses and for all age and sex groups THR was almost 100% cost-effective at any willingness-to-pay level. Revision rates for all types of THR were low. Category A THR (cemented components with a polyethylene-on-metal articulation) was more cost-effective for older age groups. However, across all age–sex groups combined, the mean cost for category E THR (cemented components with a polyethylene-on-ceramic articulation) was slightly lower and the mean QALYs gained were slightly higher. Category E therefore dominated the other four categories. Certain types of THR appeared to confer some benefit, including larger femoral head sizes, use of a cemented cup, use of a cross-linked polyethylene cup liner and a ceramic-on-ceramic as opposed to a metal-on-polyethylene articulation. Further RCTs with long-term follow-up are needed.

Study registration: This study is registered as PROSPERO CRD42013003924.

Funding: The National Institute for Health Research Health Technology Assessment programme.

Contents

List of tables xv

List of figures xxiii

List of boxes xxix

List of abbreviations xxxi

Plain English summary xxxiii

Scientific summary xxxv

Chapter 1 Background 1
Description of the health problem 1
Aetiology, pathology and prognosis 1
Impact of the health problem 3
Current service provision 4
Relevant national guidance 5
Description of the technology under assessment 8
Summary of total hip replacement 8
Summary of hip resurfacing arthroplasty 17
Failure of hip replacement 18
Revision of hip arthroplasty 19
Current usage in the NHS 20
General statistics 20
Hip replacement surgery 20
Type of procedure 20
Background summary 21

Chapter 2 Definition of the decision problem 23
Decision problem 23
Overall aims and objectives 23

Chapter 3 Joint registries 25
Description of the three largest international registries 25
Australian Orthopaedic Association National Joint Replacement Registry 25
The Swedish Hip Arthroplasty Register 26
National Joint Registry for England and Wales 26
Summary of national registries 27

Chapter 4 Assessment of evidence 29
Methods for the review of clinical effectiveness 29
Identification of studies 29
Inclusion criteria 30
Exclusion criteria 31
Study selection process 31
Quality assessment strategy 31
Grading the overall quality of clinical effectiveness evidence 32

Data extraction strategy 32
Data management 33
Meta-analysis 33
Publication bias 33
Analysis to explore heterogeneity 33
Data synthesis and interpretation 34
Industry submissions regarding effectiveness of treatments 34
Results of the review of clinical effectiveness 34
Search results 34
Comparison of different types of total hip replacement 36
Grading the overall quality of the evidence 61
Comparison between total hip replacement and resurfacing arthroplasty 64
Grading the overall quality of the evidence 75
Overall summary of the clinical effectiveness findings 77
Limitations of the reviewed evidence and pitfalls in interpretation 77
Scenario analysis around revision rates 77
Comparison of the results from randomised controlled trials and systematic reviews 78
Strengths and limitations of the review 78
Future research 79
Methods for the review of cost-effectiveness 79
Identification of studies 79
Inclusion and exclusion criteria 80
Assessment of eligibility 81
Data extraction 81
Quality assessment 82
Results of the review of cost-effectiveness 82
Identification of studies 82
Description of included studies 83
Summary of the cost-effectiveness evidence 92
Methods for the review of registries 92
Identification of studies 92
Inclusion and exclusion criteria 92
Assessment of eligibility 93
Data extraction 93
Results of the registry review 94
Identification of studies 94
Review of included studies following stage two exclusion 95
Summary of the registry evidence 104

Chapter 5 Individual patient data set **105**
Introduction to individual patient data analysis 105
Method 105
Selection of patients 105
Structure of the database 105
Contents of the database 106
Results 106
Hip resurfacing arthroplasty 106
Total hip replacement 106
Total hip replacement category development 108
Matching 110

Assessment of utility and quality of the National Joint Registry for England and Wales database 112
Strengths 112
Weaknesses 112
Summary of the individual patient data set 112

Chapter 6 Patient-reported outcome measures **113**
Quality of life and utilities 113
Background 113
Methods 113
Results 113
Summary of patient-reported outcome measure data 114

Chapter 7 Methods for modelling revision rates **115**
Introduction 115
Revision rates 116
Categories of total hip replacement 116
Patient populations to be compared 116
Overall revision rates, competing risks and rationale for analysis 119
Results 122
Proportional hazards tests 122
Comparison of resurfacing arthroplasty with total hip replacement 123
Comparison of the total hip replacement categories 131
Comparison between resurfacing arthroplasty and total hip replacement: subgroup analyses according to sex (women) 139
Comparison between resurfacing arthroplasty and total hip replacement: subgroup analyses according to sex (men) 142
Comparison of total hip replacement revision rates according to sex and age: men aged more than 65 years 143
Comparison of total hip replacement revision rates according to sex and age: women aged more than 65 years 145
Comparison of total hip replacement revision rates according to sex and age: men aged less than 65 years 145
Comparison of total hip replacement revision rates according to sex and age: women aged less than 65 years 148
Comparison of revision rates with the National Institute for Health and Care Excellence benchmark 148
Flexible parametric modelling 150
Discussion of methods of modelling revision rates 151
Summary 153

Chapter 8 Warwick economic assessment **155**
Methods 155
De novo analysis 155
Model structure 155
Base-case analysis 157
Structural model assumptions 157
Estimation of model parameters 158
Base-case cost inputs: resurfacing arthroplasty compared with total hip replacement 159
A summary of the transition probabilities, utilities and cost inputs for the cost–utility model 163
Cost-effectiveness analysis 165
Sensitivity analyses 165

Results of the cost-effectiveness analysis 170
Base-case results 170
Sensitivity analysis results 175
Discussion of the economic assessment 210
Summary of results 210
Conclusion of the cost-effectiveness analysis 212
Comparison of the results with technology appraisal guidance 2, technology appraisal guidance 44, the manufacturer's submission and international registries 213
Summary and critique of the manufacturers' submissions 217
DePuy International Ltd 217
Smith & Nephew, Inc. 230
Stryker 231
JRI Orthopaedics Ltd 232

Chapter 9 Discussion **235**
Decision problem and objectives 235
Methods and summary of findings 235
Strengths and limitations 237

Chapter 10 Conclusions and implications for practice **241**
Resurfacing arthroplasty compared with total hip replacement 241
Comparison between different types of total hip replacement 241
Recommendations for research 241

Acknowledgements **243**

References **245**

Appendix 1 Search strategies for the reviews of clinical effectiveness, cost-effectiveness and registry data **277**

Appendix 2 Quality assessment of included randomised controlled trials and systematic reviews **293**

Appendix 3 Study details and patient characteristics of included randomised controlled trials and systematic reviews **327**

Appendix 4 Full data extraction of included randomised controlled trials and systematic reviews **351**

Appendix 5 Summary of manufacturer submissions **465**

Appendix 6 Excluded papers and reasons for exclusion **475**

Appendix 7 Clinical trials and health services research identified **487**

Appendix 8 Overview of outcomes of relevance in included randomised controlled trials and systematic reviews **493**

Appendix 9 Table of functional/clinical and quality of life measures **499**

Appendix 10 Data extraction table of characteristics of eligible cost-effectiveness total hip replacement and resurfacing arthroplasty studies **501**

Appendix 11 Key studies reporting costs for resurfacing arthroplasty 539

Appendix 12 Key studies reporting costs for total hip replacement 541

Appendix 13 Excluded cost-effectiveness papers and reasons for exclusion 545

Appendix 14 Key studies reporting utilities for total hip replacement 547

Appendix 15 Key studies reporting utilities for resurfacing arthroplasty 551

Appendix 16 Excluded studies for registry searches 553

Appendix 17 Catalogue of modelled parametric fits to observed time to revision 557

Appendix 18 Results from competing risk and Kaplan–Meier analyses of revision rates for patients receiving total hip replacement and resurfacing arthroplasty interventions 621

Appendix 19 Information criteria scores for parametric models 623

Appendix 20 Plots of Kaplan–Meier-estimated cumulative hazards compared with modelled cumulative hazards 629

Appendix 21 Comparison of resurfacing arthroplasty and total hip replacement according to sex 641

Appendix 22 Comparison of total hip replacement revision rates according to sex and age 645

Appendix 23 Flexible parametric fits to observed time to revision 653

Appendix 24 Revision rates in studies with extended follow-up 665

List of tables

TABLE 1 Total hip replacement and RS articulation and fixation type combinations 12

TABLE 2 Percentage spread of fixation type and bearing/articulation surface for primary hip replacements reported in the 2011 NJR report (2010/11) 13

TABLE 3 Combination of bearing/articulation surface and fixation method by frequency as reported in the 2011 NJR report (2010/11) 13

TABLE 4 Revision procedures by type and year as reported in the 2011 NJR report 19

TABLE 5 Percentage of procedures by organisation type reported in the NJR 20

TABLE 6 Percentage of interventions by fixation method across NHS hospitals and treatment centres reported in the 2011 NJR report 21

TABLE 7 Population, intervention, comparator/control and outcome table 24

TABLE 8 The three longest-established joint (hip) replacement registries available worldwide 25

TABLE 9 Overall study characteristics across the 13 RCTs comparing different types of THR 36

TABLE 10 Distribution of 13 RCTs according to basis of THR comparison 37

TABLE 11 Risk of bias summary for RCTs: review authors' judgements about each risk of bias item – THR vs. THR 38

TABLE 12 Methodological quality assessment summary for systematic reviews: THR vs. THR 40

TABLE 13 Harris Hip Score (range 0–100): RCTs 42

TABLE 14 Harris Hip Score (range 0–100): systematic reviews 45

TABLE 15 Western Ontario and McMaster University Osteoarthritis Index (range 0–100): RCTs 45

TABLE 16 McMaster Toronto Arthritis Patient Preference Questionnaire scores (range 0–30): RCTs 46

TABLE 17 Merle d'Aubigné and Postel scores (range 0–18): RCTs 47

TABLE 18 University of California Los Angeles activity scores (range 1–10): RCTs 47

TABLE 19 Oxford Hip Score (range 0–48): systematic review 47

TABLE 20 Short Form questionnaire-12 items (range 0–100): RCTs 48

TABLE 21 Short Form questionnaire-12 items (range 0–100): systematic review 49

TABLE 22 Revision rate: RCTs 50

TABLE 23 Revision rate: systematic reviews 51

TABLE 24 Mortality rate: RCTs 52

TABLE 25 Femoral head penetration rate: RCTs 54

TABLE 26 Implant dislocation rate: RCTs 55

TABLE 27 Implant dislocation rate: systematic reviews 57

TABLE 28 Osteolysis: RCTs 58

TABLE 29 Osteolysis: systematic reviews 58

TABLE 30 Aseptic loosening: RCTs 59

TABLE 31 Femoral fracture: RCTs 59

TABLE 32 Infection: RCTs 60

TABLE 33 Deep-vein thrombosis: RCTs 60

TABLE 34 Aseptic loosening: systematic review 60

TABLE 35 Grading of Recommendations, Assessment, Development and Evaluation evidence profile for gradable outcomes reported in RCTs of THR 62

TABLE 36 Summary of evidence regarding the differences between the different types of THR for each reported outcome: RCTs 63

TABLE 37 Summary of evidence regarding the differences between the different types of THR for each reported outcome: systematic reviews 64

TABLE 38 Overall study characteristics across the three RCTs comparing THR and RS 64

TABLE 39 Risk of bias summary for RCTs: review authors' judgements about each risk of bias item – THR vs. RS 66

TABLE 40 Methodological quality assessment summary for systematic reviews: THR vs. RS 68

TABLE 41 Summary of the results for THR vs. RS: RCTs 72

TABLE 42 Summary results for RS vs. THR: systematic reviews 74

TABLE 43 Grading of Recommendations, Assessment, Development and Evaluation evidence profile for gradable outcomes reported in RCTs of THR vs. RS 75

TABLE 44 Summary of evidence regarding the differences between THR and RS for each reported outcome in the RCTs 76

TABLE 45 Summary of evidence regarding the differences between THR and RS for each reported outcome in the systematic reviews 76

TABLE 46 Critical appraisal of the economic evaluation studies using the CHEC-list 86

TABLE 47 Critical appraisal of the economic models using an adapted checklist from Philips *et al.* 87

TABLE 48 Characteristics of key cost-effectiveness studies informing the Markov model 89

TABLE 49 Summary table of registry studies on RS 96

TABLE 50 Summary table of registry studies on THR 99

TABLE 51 Subdivision of male RS patients by age and head size (excludes RA patients) 107

TABLE 52 Subdivision of female RS patients by age and head size (excludes RA patients) 107

TABLE 53 Characteristics of the five THR categories 110

TABLE 54 Composition of the five THR categories by age and sex 110

TABLE 55 Prosthesis characteristics for the five THR categories 111

TABLE 56 European Quality of Life-5 Dimensions utility index scores for patients who completed the EQ-5D-3L questionnaire after THR (analysis 1) 113

TABLE 57 European Quality of Life-5 Dimensions index scores for patients who completed the EQ-5D-3L questionnaire after THR and who required further surgery (analysis 1) 114

TABLE 58 European Quality of Life-5 Dimensions index scores for patients who completed the EQ-5D-3L questionnaire after THR by age band and sex (analysis 2) 114

TABLE 59 Age and sex of RS and THR recipients 118

TABLE 60 Age and sex characteristics of patient groups receiving THR prostheses 119

TABLE 61 Age and sex mix of RS and THR populations 127

TABLE 62 Information criteria scores for models of revision rates (RS and matched THR) 129

TABLE 63 Bathtub-modelled percentage of patients requiring revision 131

TABLE 64 Information criteria scores for models of revision rates: THR categories 133

TABLE 65 Age and sex of recipients of THR categories A–E 136

TABLE 66 Bathtub-modelled percentage of patients requiring revision 137

TABLE 67 Log-normal-modelled percentage of patients requiring revision 138

TABLE 68 Bathtub-modelled percentage of patients requiring revision (women aged 53.5 years) 142

TABLE 69 Bathtub-modelled percentage of patients requiring revision (men aged 57.1 years) 143

TABLE 70 Log-normal-modelled percentage of patients requiring revision (men aged > 65 years) 145

TABLE 71 Log-normal-modelled percentage of patients requiring revision (women aged > 65 years) 147

TABLE 72 Bathtub-modelled percentage of patients requiring revision (men aged < 65 years) 148

TABLE 73 Modelled percentage of patients requiring revision (women aged < 65 years) 150

TABLE 74 Estimated percentage of patients requiring revision at 10 years 150

TABLE 75 Key features of the analysis 156

TABLE 76 Summary of assumptions 158

TABLE 77 Resurfacing arthroplasty prosthesis costs as reported by the NHS Supply Chain 159

TABLE 78 Total hip replacement prosthesis costs as reported by the NHS Supply Chain 159

TABLE 79 Total cost of surgery 160

TABLE 80 Costs of revision 160

TABLE 81 Cost of a successful revision procedure for THR and RS 161

TABLE 82 Prosthesis costs for the comparison between different types of THR 162

TABLE 83 Cost of a bone cement pack 163

TABLE 84 Bathtub parameters for the comparison between RS and THR and the comparison between different types of THR 163

TABLE 85 Summary of the transition probabilities, utilities and cost inputs used in the base-case analysis 164

TABLE 86 Subgroup analysis: time to revision for RS vs. THR and THR vs. THR 166

TABLE 87 Summary of utility inputs for subgroup analysis 167

TABLE 88 Sensitivity analysis: time to revision for RS vs. THR 167

TABLE 89 Sensitivity analysis: time to revision for THR vs. THR 168

TABLE 90 Prosthesis costs for sensitivity analysis for the comparison between RS and THR: highest and lowest list prices for both RS and THR (weighted average of all categories) 169

TABLE 91 Prosthesis costs for sensitivity analysis for the comparison between different types of THR: highest and lowest unit costs for each category of prosthesis and a 20% reduction in prosthesis list price for each category 169

TABLE 92 Summary of utility inputs for sensitivity analysis 170

TABLE 93 Base-case deterministic and probabilistic results for all patients using the bathtub model 171

TABLE 94 Base-case deterministic and probabilistic results for all THR patients using the bathtub model 173

TABLE 95 Deterministic results for 40-, 50- and 60-year-old male and female patients 175

TABLE 96 Probabilistic results for 40-, 50- and 60-year-old male and female patients 176

TABLE 97 Deterministic results for men and women aged > 65 years for a 10-year time horizon 179

TABLE 98 Probabilistic results for men and women aged > 65 years for a 10-year time horizon 180

TABLE 99 Deterministic results for men and women aged > 65 years for a lifetime horizon 181

TABLE 100 Probabilistic results for men and women aged > 65 years for a lifetime horizon 182

TABLE 101 Deterministic results for men aged < 65 years for a 10-year time horizon 184

TABLE 102 Probabilistic results for men aged < 65 years for a 10-year time horizon 185

TABLE 103 Deterministic results for men aged < 65 years for a lifetime horizon 186

TABLE 104 Probabilistic results for men aged < 65 years for a lifetime horizon 187

TABLE 105 Deterministic results for women aged < 65 years for a 10-year time horizon 189

TABLE 106 Probabilistic results for women aged < 65 years for a 10-year time horizon 190

TABLE 107 Deterministic results for women aged < 65 years for a lifetime horizon 191

TABLE 108 Probabilistic results for women aged < 65 years for a lifetime horizon 192

TABLE 109 Deterministic and probabilistic results for all patients using the bathtub model adjusted for age and sex 193

TABLE 110 Deterministic and probabilistic results for all THR patients using the bathtub model adjusted for age and sex 195

TABLE 111 Deterministic and probabilistic results for all THR patients using the log-normal model 197

TABLE 112 Deterministic and probabilistic results for all THR patients using the log-normal model adjusted for age and sex 199

TABLE 113 Deterministic and probabilistic results for the lowest and highest cost estimates for prostheses for THR vs. RS patients 201

TABLE 114 Deterministic and probabilistic results for all THR patients using the highest cost estimates for prostheses using the bathtub model 203

TABLE 115 Deterministic and probabilistic results for all THR patients using the lowest cost estimates for prostheses using the bathtub model 205

TABLE 116 Deterministic and probabilistic results for all THR patients assuming a price de-escalator of 20% using the bathtub model 206

TABLE 117 Deterministic and probabilistic results using utility values from Rolfson *et al.* 207

TABLE 118 Patient populations considered in the manufacturer's model based on NJR data 219

TABLE 119 Utility values used in the manufacturer's model 219

TABLE 120 Resource use costs used in the manufacturer's model 220

TABLE 121 Base-case results for the comparison between THR and RS in patients suitable for both procedures 222

TABLE 122 Base-case results for the comparison between different categories of THR and RS in patients suitable for both procedures 223

TABLE 123 Base-case results for the comparison of different THR categories in patients not suitable for RS 224

TABLE 124 Results reported in the DePuy submission for the scenario analyses for patients suitable for THR and hip RS 225

TABLE 125 Results reported in the DePuy submission for the scenario analyses for patients not suitable for hip RS 226

TABLE 126 Results reported in the DePuy submission for the probabilistic sensitivity analysis for patients suitable for THR and hip RS 227

TABLE 127 Results reported in the DePuy submission for the probabilistic sensitivity analysis for patients not suitable for hip RS 228

TABLE 128 Mean LOS for patients receiving primary THR or hip RS 229

TABLE 129 Age and sex of patients receiving primary hip replacements in 2011 230

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR 328

TABLE 131 Study details of included systematic reviews: studies comparing different types of THR 341

TABLE 132 Study details and patient characteristics of included RCTs: THR vs. RS 345

TABLE 133 Study details of included systematic reviews: THR vs. RS 347

TABLE 134 Overview of outcomes of relevance in included RCTs: studies comparing different types of THR ($n = 14$) 494

TABLE 135 Overview of outcomes of relevance in included systematic reviews: studies comparing different types of THR ($n = 5$) 496

TABLE 136 Overview of outcomes of relevance in included RCTs: THR vs. RS ($n = 3$) 497

TABLE 137 Overview of outcomes of relevance in included systematic reviews: THR vs. RS ($n = 3$) 497

TABLE 138 Predicted percentage of patients requiring revision at 20 and 30 years according to each model 642

TABLE 139 Predicted percentage of patients requiring revision at 20 and 30 years according to each model 644

TABLE 140 Predicted revision rates for well-fitting models (% revised) 646

TABLE 141 Predicted revision rates for well-fitting models (% revised) 648

TABLE 142 Predicted revision rates for well-fitting models (% revised) 650

TABLE 143 Predicted revision rates for well-fitting models (% revised) 652

List of figures

FIGURE 1 Example pathway for patient with arthritis in primary care 9

FIGURE 2 Example hip replacement care pathway in secondary care 10

FIGURE 3 Overview of four different fixation options for the femoral stem and acetabular cup in THR 11

FIGURE 4 Cemented metal stem, metal femoral head and polyethylene acetabular cup 14

FIGURE 5 Cemented metal stem, ceramic femoral head and polyethylene acetabular cup 14

FIGURE 6 Cemented metal stem, ceramic femoral head and ceramic acetabular cup 15

FIGURE 7 Cemented metal stem, metal femoral head and metal acetabular cup 15

FIGURE 8 Diagrammatic representation of a hip RS 17

FIGURE 9 Flow diagram of study identification for the clinical effectiveness review 35

FIGURE 10 Risk of bias graph for RCTs: review authors' judgements about each risk of bias item – THR vs. THR 39

FIGURE 11 Harris Hip Score 44

FIGURE 12 Mortality 53

FIGURE 13 Implant dislocation 56

FIGURE 14 Risk of bias graph for RCTs: review authors' judgements about each risk of bias item – THR vs. RS 67

FIGURE 15 Risk of infection 71

FIGURE 16 Risk of deep-vein thrombosis 71

FIGURE 17 Risk of implant dislocation 71

FIGURE 18 Flow diagram of study identification for the cost-effectiveness review 82

FIGURE 19 Flow diagram of study identification for the registry review 94

FIGURE 20 End points for all RS patients included in the analysis (excludes RA patients) 106

FIGURE 21 End points for all THR patients included in the analysis (includes RA patients) 108

FIGURE 22 Density of patients among the most frequently used combinations of THR (RA patients excluded) 109

FIGURE 23 Numbers of patients in each of the selected THR categories (RA patients excluded) 109

FIGURE 24 Age distribution of NJR patients who received THR or RS and age distribution by sex for those who received RS 116

FIGURE 25 Kernel density diagram of the two distributions 117

FIGURE 26 Differing modelled hazard on extrapolation beyond observation for THR HyMoP (category D) female patients aged < 65 years 120

FIGURE 27 Kaplan–Meier analysis for (a) death; and (b) revision for THR CeMoP (category A) female patients aged > 85 years 121

FIGURE 28 Log-Kaplan–Meier estimated cumulative hazard vs. log-time for different THR categories 122

FIGURE 29 Log-Kaplan–Meier estimated cumulative hazard vs. log-time for different THR category populations stratified by sex and age 123

FIGURE 30 Time to revision: all RS patients and all THR patients 123

FIGURE 31 Time to revision: all RS and all THR patients according to sex 124

FIGURE 32 Time to revision: all RS and all THR patients (categories A–E) 124

FIGURE 33 Time to revision: all RS and all THR patients (categories A–E) by sex 125

FIGURE 34 Time to revision: all male RS and all male THR patients (categories A–E) 126

FIGURE 35 Time to revision: all female RS and all female THR patients (categories A–E) 126

FIGURE 36 (a) Age distribution; (b) and time to revision for RS and THR matched populations 128

FIGURE 37 Revision rates for matched and whole THR populations 128

FIGURE 38 Bathtub fits and extrapolations for matched THR and RS populations 129

FIGURE 39 Kaplan–Meier vs. modelled cumulative hazard 130

FIGURE 40 Revision estimated for all THR patients and those receiving category A–E THRs 131

FIGURE 41 Kaplan–Meier analyses across the five categories of THR 131

FIGURE 42 Observed time to revision with 95% CIs for compared THR categories 132

FIGURE 43 Modelled vs. Kaplan–Meier-estimated cumulative hazards 134

FIGURE 44 Bathtub parametric fits to observed time to revision for THR categories A–E 135

FIGURE 45 Extrapolation of bathtub models of revision for THR categories A–E 136

FIGURE 46 Total hip replacement revision rates observed for men and women 137

FIGURE 47 Kernel density plots of age at primary THR for category A–E THR prostheses 138

FIGURE 48 Log-normal parametric fits to observed revision rates for THR categories A–E 139

FIGURE 49 Log-normal-modelled revision 140

FIGURE 50 Kernel density plot for age distribution in matched RS and THR female groups 141

FIGURE 51 Observed revision (95% CI) and bathtub models for RS and THR female groups 141

FIGURE 52 Kernel density plot for age distribution in matched RS and THR male groups 142

FIGURE 53 Observed revision (95% CI) and bathtub models for RS and THR male groups 143

FIGURE 54 Observed revision (95% CI) for men aged > 65 years and log-normal models for THR categories 144

FIGURE 55 Observed revision (95% CI) for women aged > 65 years and log-normal models for THR categories 146

FIGURE 56 Observed revision (95% CI) for men aged < 65 years and bathtub models for THR categories 147

FIGURE 57 Observed revision (95% CI) for women aged < 65 years and bathtub models for THR categories 149

FIGURE 58 Markov model 155

FIGURE 59 Cost-effectiveness planes and CEACs for THR vs. RS 172

FIGURE 60 Cost-effectiveness planes and CEACs for the comparison between different types of THR 174

FIGURE 61 Cost-effectiveness planes and CEACs for the comparison between THR and RS for female patients by age group (lifetime horizon) 177

FIGURE 62 Cost-effectiveness planes and CEACs for the comparison between THR and RS for male patients by age group (lifetime horizon) 178

FIGURE 63 Cost-effectiveness acceptability curves for the comparison between different types of THR (patients aged > 65 years) 183

FIGURE 64 Cost-effectiveness acceptability curves for the comparison between different types of THR for women and men by age group (lifetime horizon) 188

FIGURE 65 Cost-effectiveness planes and CEACs for the comparison between RS and THR using the bathtub model adjusted for age and sex 194

FIGURE 66 Cost-effectiveness planes and CEACs for the comparison between different types of THR using the bathtub model adjusted for age and sex 196

FIGURE 67 Cost-effectiveness planes and CEACs for the comparison between different types of THR using a log-normal model 198

FIGURE 68 Cost-effectiveness planes and CEACs for the comparison between different types of THR using the log-normal model adjusted for age and sex 200

FIGURE 69 Cost-effectiveness acceptability curves for RS vs. THR using the lowest and highest cost estimates for prostheses 202

FIGURE 70 Cost-effectiveness acceptability curves for all THR patients using the lowest and highest costs for prostheses and assuming a 20% price de-escalator 204

FIGURE 71 Cost-effectiveness acceptability curves for the comparison of different types of THR using utility values from Rolfson *et al.* 208

FIGURE 72 Tornado diagram illustrating sensitivity analysis for net monetary benefit: CeCoP (category E) vs. CeMoP (category A) 209

FIGURE 73 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010) 214

FIGURE 74 Time to revision results reproduced with permission from (a) Schulte *et al.*; (b) Madey *et al.*; and (c) Callaghan *et al.* 216

FIGURE 75 Competing risk and Kaplan–Meier analyses of revision rates after (a) RS; and (b) THR 622

FIGURE 76 Female RS vs. non-RS (matched THR) cumulative hazard plots 630

FIGURE 77 Male RS vs. non-RS (matched THR) cumulative hazard plots 631

FIGURE 78 Cumulative hazard plots for men aged > 65 years 632

FIGURE 79 Cumulative hazard plots for women aged > 65 years 634

FIGURE 80 Cumulative hazard plots for men aged < 65 years 636

FIGURE 81 Cumulative hazard plots for women aged < 65 years 638

FIGURE 82 Age distribution in matched RS and THR female groups 641

FIGURE 83 Observed revision rate for women receiving RS and THR (note difference in *y*-axis scales) 641

FIGURE 84 Observed revision (95% CI) for RS (women) with model fits and extrapolation 642

FIGURE 85 Age distribution in matched RS and THR male groups 643

FIGURE 86 Observed revision rate for men receiving RS and THR (note the difference in the *y*-axis scales) 644

FIGURE 87 (a) Bathtub fit to observed time to revision for male THR patients; and (b) extrapolation beyond observed data 644

FIGURE 88 Kaplan–Meier plots: time to revision for THR categories A–E for men aged > 65 years 645

FIGURE 89 Kaplan–Meier plots: time to revision for THR categories A–E for women aged > 65 years 647

FIGURE 90 Kaplan–Meier plots: time to revision for THR categories A–E for men aged < 65 years 649

FIGURE 91 Kaplan–Meier plots: time to revision for THR categories A–E for women aged < 65 years 651

FIGURE 92 Resurfacing arthroplasty and non-RS, men and women 653

FIGURE 93 Total hip replacement comparisons: men (a–e) and women (f–j) aged > 65 years 654

FIGURE 94 Total hip replacement comparisons: men (a–e) and women (f–j) aged < 65 years 657

FIGURE 95 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010) 666

FIGURE 96 Time to revision results reproduced with permission from (a) Schulte *et al.*; (b) Madey *et al.*; and (c) Callaghan *et al.* 668

List of boxes

BOX 1 Limitations and strengths of the clinical effectiveness review by DePuy 466

BOX 2 Limitations and strengths of the clinical effectiveness review by Smith & Nephew, Inc. 470

BOX 3 Limitations and strengths of clinical effectiveness review by Stryker 471

BOX 4 Limitations and strengths of the clinical effectiveness review by JRI Orthopaedics Ltd 473

List of abbreviations

AAOS	American Academy of Orthopedic Surgeons
AIC	Akaike information criterion
AIMS	Arthritis Impact Measurement Scale
AMSTAR	Assessment of Multiple Systematic Reviews
AOANJRR	Australian Orthopaedic Association National Joint Replacement Registry
ASA	American Society of Anesthesiologists
BIC	Bayesian information criterion
BMI	body mass index
CDSR	Cochrane Database of Systematic Reviews
CEAC	cost-effectiveness acceptability curve
CEA Registry	Cost-effectiveness Analysis Registry
CeCoP	ceramic head (cemented stem) on cemented polyethylene cup
CeLCoC	ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)
CeLMoP	metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)
CeMoP	metal head (cemented stem) on cemented polyethylene cup
CENTRAL	Cochrane Central Register of Controlled Trials
CG	clinical guideline
CHEC	Consensus on Health Economic Criteria
CI	confidence interval
CRD	Centre for Reviews and Dissemination
DARE	Database of Abstracts of Reviews of Effects
DSU	Decision Support Unit
EQ-5D	European Quality of Life-5 Dimensions
EQ-5D-3	European Quality of Life-5 Dimensions 3 Levels
FDA	US Food and Drug Administration
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
HHS	Harris Hip Score
HOOS	Hip Disability and Osteoarthritis Outcome Score
HR	hazard ratio
HRG4	Healthcare Resource Group v4
HSCI	Health Service Cost Index
HSRProj	Health Services Research Projects in Progress
HTA	Health Technology Assessment
HUI	Health Utilities Index
HyMoP	hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)
ICER	incremental cost-effectiveness ratio
IPD	individual patient data
LISOH	Lequesne Index of Severity for Osteoarthritis of the Hip
LOS	length of stay
MACTAR	McMaster Toronto Arthritis Patient Preference Questionnaire
MCID	minimal clinically important difference
MD	mean difference
MHRA	Medicines and Healthcare products Regulatory Agency
MRI	magnetic resonance imaging

NCC-CC	National Collaborating Centre for Chronic Conditions
NHP	Nottingham Health Profile
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
NJR	National Joint Registry for England and Wales
NLM	National Library of Medicine
NSAID	non-steroidal anti-inflammatory drug
OA	osteoarthritis
ODEP	Orthopaedic Data Evaluation Panel
OHS	Oxford Hip Score
OR	odds ratio
PbR	payment by results
PICO	population, intervention, comparator/control and outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	patient-reported outcome measure
PSS	personal and social services
QALY	quality-adjusted life-year
RA	rheumatoid arthritis
RCT	randomised controlled trial
RR	risk ratio (relative risk)
RS	resurfacing arthroplasty
SD	standard deviation
SE	standard error
SF-12	Short Form questionnaire-12 items
SF-36	Short Form questionnaire-36 items
SHAR	Swedish Hip Arthroplasty Register
SSI	surgical site infection
TA	technology appraisal
THR	total hip replacement
UCLA	University of California Los Angeles
UKCRN	UK Clinical Research Network
VAS	visual analogue scale
WOMAC	Western Ontario and McMaster University Osteoarthritis Index
WTP	willingness to pay

Note

This monograph is based on the Technology Assessment Report produced for NICE. The full report contained a considerable number of data that were deemed commercial-in-confidence. The full report was used by the Appraisal Committee at NICE in their deliberations. The full report with each piece of commercial-in-confidence data removed and replaced by the statement 'commercial-in-confidence information (or data) removed' is available on the NICE website: www.nice.org.uk.

The present monograph presents as full a version of the report as is possible while retaining readability, but some sections, sentences, tables and figures have been removed. Readers should bear in mind that the discussion, conclusions and implications for practice and research are based on all the data considered in the original full NICE report.

Plain English summary

Osteoarthritis (OA) is a leading cause of both pain and disability in the UK. People with OA can be treated with a total hip replacement (THR). This operation involves the replacement of a damaged hip joint with an artificial hip joint. Hip resurfacing arthroplasty (RS) (surgical repair of the joint) is a different operation. The hip joint is not removed but instead the joint surface of the leg bone (femur) is covered with a metal surface. The aim of this work was to review the clinical effectiveness and cost-effectiveness of both treatment options for patients who are suitable for either one. In addition, we aimed to review the different types of THR available.

Our review found that patients' hip function and their general quality of life improved after having either operation. We found that more hip RS had to be revised than THR. The economic model showed that costs for RS were also higher than those for THR. Very similar results were found for all age and sex groups and THR was almost 100% cost-effective each time. When we investigated the differences between the different types of THR we found small but clear differences in the costs and effectiveness of particular types of THR for particular age and sex groups.

Overall, the review concluded that evidence about the benefits of THR and hip RS is lacking. We found that certain types of THR appeared to show some benefit over others. However, further research is needed and it should consist of better-quality studies that have longer follow-up of patients after their operations.

Scientific summary

Background

Osteoarthritis (OA) is a leading cause of pain and disability in the UK. The incidence rates of OA of the hip in men and women aged 70–79 years are estimated to be 430 and 600 per 100,000 person-years, respectively. The economic impact of arthritis is vast, because of both direct costs to the health-care system, community and social services and indirect costs to patients, carers and families. Patients who do not respond to non-surgical measures are referred for elective surgical interventions, most commonly total hip replacement (THR) or resurfacing arthroplasty (RS).

Total hip replacement involves replacement of the hip joint with an artificial hip prosthesis consisting of a cup (with or without liner), a femoral stem and head. There are different types of THR including different types of articulation surface (metal, ceramic, polyethylene, ceramicised metal); implant component fixation (cemented, cementless, hybrid, reverse hybrid); and implant component size. RS involves replacement of the femoral head surface with a metal covering. The resurfacing component articulates with a hollow metal cup located in the acetabulum. Revision is undertaken when implants fail because of infection or loosening. Previous National Institute for Health and Care Excellence guidance indicated that the benchmark for selection of prostheses for THR should be a revision rate of $\leq 10\%$ at 10 years.

Decision problem and objectives

The main objective was to undertake a clinical effectiveness and cost-effectiveness analysis of different types of THR and RS for the treatment of pain and disability in people with end-stage arthritis of the hip.

Specific aims were to compare the clinical effectiveness and cost-effectiveness of different types of primary THR:

(a) with RS for people in whom both procedures are suitable
(b) with each other for those not suitable for hip RS.

Systematic reviews

Methods

Searches were undertaken of clinical effectiveness, cost-effectiveness and registry studies in December 2012. For clinical effectiveness, studies were limited to 2008 onwards and a sample size of ≥ 100 participants. Electronic databases searched included MEDLINE, EMBASE, The Cochrane Library and Current Controlled Trials. Reference lists, and manufacturer and professional organisation websites were screened. Full-text English-language reports of randomised controlled trials (RCTs), systematic reviews and meta-analyses were included.

Two independent reviewers screened all records and extracted data, with disagreements resolved through consensus. Methodological quality was assessed using the Cochrane Collaboration's risk of bias and the Assessment of Multiple Systematic Reviews (AMSTAR) tools. Estimates of post-treatment mean difference (MD) for continuous outcomes and risk ratios (RRs) for binary outcomes of individual studies were pooled using a random-effects model. Dichotomous outcomes were pooled as RRs using a Mantel–Haenszel fixed-effects model or as odds ratios (ORs) using the Peto fixed-effects model. Heterogeneity was

determined through Cochran's *Q* and the I^2 statistics. Overall quality of evidence was assessed using GRADE (Grading of Recommendations, Assessment, Development and Evaluation).

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Systematic review results

Clinical effectiveness: resurfacing arthroplasty compared with total hip replacement

A total of 2469 records were screened, of which 37 were included, representing 16 RCTs and eight systematic reviews. Mean age ranged from 45 to 72 years with a maximum follow-up of 20 years. Mean post-THR Harris Hip Score (HHS) and Western Ontario and McMaster University Osteoarthritis Index (WOMAC) and Short Form questionnaire-12 items (SF-12) scores measured at different follow-up periods did not differ between THR groups, including between cross-linked polyethylene and traditional polyethylene cup liners [HHS pooled MD 2.29, 95% confidence interval (CI) –0.88 to 5.45].

There was a reduced risk of implant dislocation with the use of a cemented cup compared with a cementless cup (high-grade evidence; pooled OR 0.34, 95% CI 0.13 to 0.89) or larger femoral head size (36 mm vs. 28 mm). Femoral head penetration rates were reduced for cross-linked compared with conventional polyethylene cup liners (low-grade evidence). Recipients of ceramic-on-ceramic articulations (vs. metal-on-polyethylene) experienced a reduced risk of osteolysis. Evidence from two RCTs indicated an increased infection risk for THR compared with RS (pooled OR 7.94, 95% CI 1.78 to 35.40).

The eight systematic reviews identified (five on different types of THR and three on THR vs. RS) were inconclusive.

Cost-effectiveness

Methods

We drew on systematic review results and the National Joint Registry for England and Wales (NJR). Using a series of cross-tabulations, we identified the top four most commonly used mutually exclusive categories of THR (> 25,000) as reported in the NJR and on clinical advice included a further mutually exclusive fifth category. We built a Markov multistate model to investigate both RS and THR using the observed time to revision. Cycle length was 1 year and we adopted 10-year and lifetime horizons.

Analysis was conducted from the perspectives of the NHS and personal and social services. All costs are in UK pounds at 2011/12 prices. Health outcomes were measured in quality-adjusted life-years (QALYs). Results are expressed as incremental cost per QALY gained. An annual discount rate of 3.5% was applied to both costs and outcomes. We ran the model deterministically and probabilistically with 1000 iterations, calculated cost-effectiveness acceptability curves (CEACs) and undertook sensitivity analyses.

Resurfacing arthroplasty compared with total hip replacement

We propensity matched RS patients drawing age–sex matched pairs from the data set of all identified categories of THR combined, to identify patients who underwent THR but who were also eligible for RS. We used NHS reference costs for RS and THR for follow-up and revision. We drew age- and sex-adjusted utility values from the patient-reported outcome measures (PROMs) data set for both THR and RS.

For the comparison of RS with THR we undertook sensitivity analyses stratified by sex and controlled for age. We assessed estimates of cost-effectiveness for men and women aged 40, 50 and 60 years using lifetime revision rates. We constructed CEACs comparing RS with THR overall, in separate age groups and at different levels of willingness to pay (WTP).

Total hip replacement compared with total hip replacement

We compared the five categories of THR with each other and in sensitivity analyses investigated patients aged > 65 years who are less eligible for RS. In the base case we used costs supplied by the manufacturers for each of the components of THR; we used alternative costs in sensitivity analyses. We used age- and sex-adjusted utility values from the PROMs data set for before and after hip replacement and for revision.

We undertook various sensitivity analyses and analysis of cost drivers. These included different age and sex categories, stratifying by age (> 65 years or < 65 years), different methods of extrapolation of revision rates, varying prosthesis costs (using NHS list prices) and discount rates. We constructed CEACs comparing different types of THR overall and in separate age groups at different levels of WTP.

Cost-effectiveness results

Resurfacing arthroplasty compared with total hip replacement

Using the NJR we found a total of 31,222 people who had undergone RS and 386,556 who had undergone a THR. In total, 3% of those undergoing THR and 11% undergoing RS had a revision by 9 years. The revision rate for all RS was always higher than that for THR. The prostheses cost £2672 and £2571 or RS and THR, respectively.

For all analyses, the mean cost for RS was higher than that for THR and the mean QALYs were lower. The incremental cost-effectiveness ratio for RS was dominated by THR; that is, THR was cheaper and more effective than RS (for a lifetime horizon in the base-case analysis, the incremental cost of RS was £11,490 and the incremental QALYs were –0.0879).

Similar results were obtained for the deterministic and probabilistic analysis of RS compared with THR and when analysed separately in sensitivity analyses for men and women by age group (40, 50 and 60 years). For all age and sex groups RS remained clearly dominated by THR. CEACs showed that, for all patients, THR was almost 100% cost-effective at any WTP level.

Total hip replacement compared with total hip replacement

We identified five categories of commonly used types of THR from the NJR:

- category A: metal head (cemented stem) on cemented polyethylene cup (CeMoP) (125,285 patients)
- category B: metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner) (CeLMoP) (37,874 patients)
- category C: ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner) (CeLCoC) (34,754 patients)
- category D: hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner) (HyMoP) (28,471 patients)
- category E: ceramic head (cemented stem) on cemented polyethylene cup (CeCoP) (12,075 patients).

There were age and sex differences between recipients of different types of THR and variations in revision rates (category A: 2.5%; B: 3.2%; C: 3.5%; D: 2.5%; E: 1.6% at 9 years). For all interventions, revision rates at 9 years were substantially less than the 10% benchmark. Costs of the different prostheses were as follows: category A: £1557.38; B: £3015.60; C: £3868.80; D: £2649.78; and E: £1995.98.

In the base-case analysis, for all age and sex groups combined and using a bathtub model (indicating an increasing likelihood of need for revision with time) and a lifetime horizon, the mean cost for category E (CeCoP) was slightly lower and the mean QALYs for category E were slightly higher than for all other THR categories in both deterministic and probabilistic analyses. Category E dominated the other four categories.

For example, in the deterministic analysis, compared with category E, category A (CeMoP) cost £278 more (£14,801 vs. £14,523) and generated 0.0022 fewer QALYs (14.7887 vs. 14.7909). The probabilistic results were very similar. The CEACs demonstrated that, over a lifetime horizon, category E was 97.2% likely to be cost-effective compared with 2.8% for category A at a WTP of £20,000 per QALY. For patients aged > 65 years, category A was more likely to be cost-effective in all groups (category A: 100% probability of being cost-effective; categories B–E: < 1% probability of being cost-effective). When examining the lifetime horizon for all age groups, category E was more clinically effective (except for men aged 80 years for whom QALYs generated by categories A and E were the same).

Sensitivity analyses using a log-normal model (indicating a decreasing risk of revision over time) for extrapolation beyond the observed data for revision rates resulted in category A being cheaper over a lifetime horizon for all age–sex groups combined. Although category E was more clinically effective than the other four categories, category A was 100% cost-effective at a WTP threshold of £20,000 per QALY. Additional sensitivity analysis using an age- and sex-adjusted log-normal model also demonstrated that over a lifetime horizon category A was 100% cost-effective at a WTP of £20,000 per QALY.

Varying the main inputs by 30% in the base-case analysis for all age–sex groups, and comparing category A with category E, demonstrated that the main drivers of difference were the costs of the components, the discount rate and the modelled revision rates.

Strengths and limitations

We reanalysed comprehensive national audit data to calculate outcomes and used the PROMs data set as a source for utility data coupled with costs from the rigorous literature review, NHS reference costs and manufacturers' costs. We did not find any relevant RCTs comparing RS and THR or different types of THR to allow us to model differences in revision rates relevant to a lifetime horizon. As NJR data are non-randomised and may be subject to selection bias, we worked to reduce confounding by propensity matching and by undertaking extensive analyses by age and sex.

In comparing RS with THR, our clinical advisors suggested that the selection of patients for RS may be based on activity levels (levels of physical fitness, athleticism); however, the only characteristics reliably collected at patient level in the NJR were age and sex. This means that we were unable to identify other characteristics or subpopulations for whom RS might be more beneficial. However, age and sex are likely to act as a proxy for physicality and it is of interest that revision rates for RS were higher in every age and sex group that we examined, including in the youngest category of men.

We identified five categories of the most commonly used combinations of THR components. To our knowledge this is the first time that different types of THR have been investigated in this comparative way. It has the advantage of more precisely reflecting current practice.

Revision rates are one of the main factors affecting the cost-effectiveness of the different categories. We had preselected category E on the recommendation of our clinical advisors before assessing revision rates. We undertook extensive modelling of revision rates to find the best methods for extrapolation beyond the observed data for all THR categories. We found that category E had lower revision rates overall and generally across age–sex groups. This pertained across different methods for extrapolation, suggesting that the relative cost-effectiveness of category E is a robust finding.

Conclusions

Systematic reviews

Total hip replacement is a common operation and is clearly beneficial. Improvements post surgery were reported in the literature for functional/clinical and quality-of-life measures regardless of the type of surgery. Much of the evidence was inconclusive because of poor reporting, missing data, inconsistent results and uncertainty in treatment effect estimates. Evidence on the relative benefits of RS compared with THR or of different types of THR was largely lacking. Certain types of THR appeared to confer some benefit including larger femoral head sizes, use of a cemented cup, use of a cross-linked polyethylene cup liner and a ceramic-on-ceramic articulation as opposed to a metal-on-polyethylene articulation.

Resurfacing arthroplasty compared with total hip replacement

Compared with THR, revision rates for RS were higher, mean costs were higher and mean QALYs were lower; RS was therefore dominated by THR. Very similar results were obtained for deterministic and probabilistic analyses and for all age and sex groups. THR was almost 100% cost-effective at any WTP level.

Total hip replacement compared with total hip replacement

Revision rates for all types of THR were low. Costs of prostheses varied depending on composition. There were small but clear differences between categories in both costs and effectiveness as measured by QALYs, and when age and sex groups were factored in. Category A was more cost-effective for older age groups for whom revision rates are lower. However, across all age–sex groups combined, mean costs were slightly lower and mean QALYs were slightly higher for category E, than for all other THR categories in both deterministic and probabilistic analyses; category E therefore dominated the other four categories.

Recommendations for research

1. Consideration should be given to setting up RCTs with long-term follow-up.
2. We were not able to link PROMs data with NJR data or with costs in our analysis; however, the NJR will embed these utility data from 2013.
3. We would welcome work to validate our new findings on the relative cost-effectiveness of different combinations of prosthesis components for THR.

Study registration

This study is registered as PROSPERO CRD42013003924.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Chapter 1 Background

Description of the health problem

Arthritis is a general term that describes pain and inflammation within a joint. There are many causes, of which the most common is osteoarthritis (OA), a degenerative disease that has become a leading cause of pain and disability both in the UK and worldwide.[1] OA is a chronic syndrome of articular cartilage degeneration with associated synovitis and hypertrophic changes within bone.[2]

Aetiology, pathology and prognosis

Osteoarthritis of the hip

The hip is a weight-bearing ball and socket joint that is commonly affected by OA. OA in the hip manifests itself as loss of articular cartilage, inflammation of synovial tissue and hypertrophy of the associated bone (e.g. osteophytes, bone sclerosis). The loss of cartilage tissue and new bone tissue growth suggests OA may result from disordered repair of cartilage damaged by mechanical and biochemical changes within the joint.[3]

When the repair process is unable to keep up with the rate of tissue damage, the consequence is symptomatic OA characterised by pain, stiffness and progressive disability.[3]

Osteoarthritis of the hip may be classified as primary or secondary. Secondary hip OA can be caused by most intra-articular diseases, including osteonecrosis, trauma, septic arthritis, Paget's disease, hip dysplasia, Perthes' disease and slipped upper femoral epiphysis. Primary hip OA is presumed when no other specific cause has been identified.[3]

Rheumatoid arthritis of the hip

Rheumatoid arthritis (RA) is an autoimmune disease that commonly affects the synovial lining of peripheral joints, including those of the hand, foot and hip. RA is a multisystem disorder with implications for almost every region of the body, including the heart, lungs and eyes.[4] Multiple episodes of synovial inflammation lead to reduced articular cartilage (e.g. causing secondary OA), joint destruction and progressive disability. It has also been associated with reduced quality of life and premature mortality.[5–7]

Rheumatoid arthritis manifests itself by gradual accumulation of structural changes within the joint, which can (particularly in late-stage disease) be detected by radiography or other imaging techniques.[5] In 2010, a joint working group of the American College of Rheumatology and the European League Against Rheumatism[5] developed new criteria for identifying patients with early RA, which place more emphasis on characteristics associated with a high risk of later progression to severe and erosive disease.

Epidemiology of osteoarthritis and rheumatoid arthritis

Osteoarthritis is one of the most commonly encountered musculoskeletal diseases. There are an estimated 2.8 million patients with OA in the UK, based on symptomatic diagnosis in patients aged > 45 years.[8] A further 8.5 million people are estimated to be affected by joint pain that can be attributed to OA.[3]

Current projections estimate that 10% of the world's population aged ≥ 60 years will present with symptoms caused by OA.[9] The prevalence and incidence of OA, including hip OA, increase with age and are higher in women than in men after 50 years of age.[10,11] For example, the incidence rates of hip OA in men and women aged 70–79 years are estimated to be 430 and 600 per 100,000 person-years, respectively.[12]

Estimates of age-standardised incidence rates of hip OA among women and men in Europe are about 53.3 and 38.1 per 100,000, respectively.[13] The prevalence of hip OA among Caucasians is demonstrably higher (range 3–6%) than in Asian, black and East Indian populations (≤ 1%).[14] In light of a longer life expectancy, an ageing population and increasing rates of obesity observed in developed countries, it is expected that both the incidence and the prevalence of OA will rise in future.[1,15,16]

It is difficult to estimate the prevalence and incidence rates of OA accurately because of variable diagnostic criteria (e.g. radiographic, symptomatic or self-reported features).[10,17,18] For example, some patients with radiographic evidence of joint damage indicative of OA may not experience pain or disability whereas some patients with clinical OA may not demonstrate radiographic changes. These discrepancies make it challenging to determine the presence or absence of OA accurately.[10] In general, the prevalence of symptomatic or self-reported OA is higher than that of radiographic OA.[3]

The prevalence of RA is estimated at 400,000 cases in the UK. Estimates of annual incidence suggest that 10,000–20,000 people develop RA in the UK each year. Although the disease may develop in patients at any age, onset classically occurs between the ages of 40 and 60 years. The incidence of RA is approximately two to three times greater in women than in men[4] and approximately 10–40% of cases manifest within the hip.[19]

Risk factors for osteoarthritis

Evidence suggests that contributing factors to OA can be classified broadly as:

(a) biomechanical (e.g. joint injury, reduced muscle strength)
(b) constitutional [e.g. advanced age (≥ 65 years), female sex, obesity and high bone density]
(c) genetic (high heritability estimates for OA).

Biomechanical factors are probably the most important cause and may explain both the relationship between OA and obesity as well as the tendency for OA to affect weight-bearing joints, for example the hips and knees.[2] Malalignment, instability and altered joint loading correlate with OA progression in both clinical and animal studies.[20,21] In the hip, femoroacetabular impingement are related to OA onset; 'cam type' is a bump on the surface of the femoral head typically affecting younger athletic men and 'pincer type' impingements describe an overdeep acetabululm, which restricts the movement of the femoral head – this typically affects middle-aged women. The prevalence of any type of congenital or acquired hip malformation is 4.3% in men and 3.6% in women. Similarly, epidemiological studies have demonstrated associations between certain occupational factors (e.g. long-distance running, farming, heavy physical work load) and hip OA.[22,23]

However, biomechanical factors alone do not explain the onset of OA in non-weight-bearing joints (e.g. the carpometacarpal joints) and metabolic factors may also play a role.[2,24]

Symptoms and diagnosis

Symptoms of hip OA include pain, stiffness and loss of function, that is, limited daily activities such as walking, climbing the stairs and performing household tasks.[1,11,19,25] The diagnosis of primary hip OA is usually based on history and clinical examination with particular assessment of joint pain, deformity and reduced range of movement. Physical examination can also exclude pain resulting from other causes, for example bursitis, tendonitis and muscle spasm. Plain radiographs of the hip are used to identify and stage OA.

Advanced imaging techniques such as magnetic resonance imaging (MRI) and computerised tomography can identify causes of secondary hip OA (e.g. stress fractures, osteonecrosis, Paget's disease, inflammatory arthropathies) as well as evaluating and monitoring the extent of hip damage.[1,18]

Natural history of osteoarthritis

The natural history of OA varies between affected joints but little is known about the natural history of the symptomatic disease. The prognosis of hip OA has been shown to be the least favourable and is the most frequent reason for surgical intervention after 1–5 years of progression.[3] The national clinical guideline (CG) for OA[3] states that hip OA has the worse outcome of all the OA sites discussed in the CG (hip, knee, hand). Occasionally, OA hips can improve without surgical intervention as measured by symptoms and radiographic change.[3] Comorbidity (e.g. diabetes, obesity, cardiovascular disease) may additionally influence the prognosis of OA, as does older age.[3]

Impact of the health problem

Significance for patients in terms of ill health (burden of disease)

Osteoarthritis has a significant impact on an individual patient, resulting in pain, stiffness, limited mobility and reduced function. A UK-based survey assessed the impact of OA on daily living for 1762 people.[26] The majority of the sample consisted of people aged ≥ 50 years, of whom 75% were female. In total, 81% of respondents were found to have experienced constant pain and/or were limited in their ability to perform everyday tasks. Many respondents had visited a general practitioner three or four times before a diagnosis of OA, which was made on average 18 months after the onset of symptoms. Approximately 72% of respondents had comorbid conditions such as heart disease, diabetes and hypertension.

Significance for the NHS

The economic impact of arthritis consists of direct costs to health-care services and indirect costs because of lost productivity and early mortality. The impact of OA on health services and the UK economy has been substantial. The cost of treating OA has been estimated to be approximately £640 per person per year.[19] A report has suggested that, if one-tenth of the 15.2 people per 1000 who experience hip pain severe enough for surgery received medical and/or physical therapy, the cost to the NHS in England and Wales would be of the order of £48M per year in 2002.[19] The costs of both surgical and non-surgical interventions are reviewed in detail later in this chapter.

Because of the ageing of the population, OA is projected to become the fourth leading cause of disability worldwide by 2020.[3] In the present economic climate of tightening health-care spending, the implications of increasing demand for the treatment of arthritis of the hip have led to intense discussions about the cost-effectiveness of new technologies and treatment options.

Measurement of disease

More than 20 tools have been developed and validated for the assessment and monitoring of patient outcomes specific to hip arthritis.[27] One commonly used disease-specific tool is the Western Ontario and McMaster University Osteoarthritis Index (WOMAC).[28] This is a 24-item questionnaire that covers three domains of pain, stiffness and physical function, with a total score ranging from 0 (worst outcome) to 100 (best outcome). Other validated tools designed to measure outcomes specific to hip function and symptoms (e.g. disability, pain, range of motion, limitations in daily living and other activities) have also been used.[27,29]

In the UK the most commonly used tools are the Oxford Hip Score (OHS)[30] and the Harris Hip Score (HHS).[31]

The Oxford Hip Score

The OHS is one of the most commonly used hip-specific measures. It was designed to assess function and pain in relation to daily activities (e.g. walking, dressing, sleeping) for patients undergoing total hip replacement (THR) surgery.[30] The OHS includes 12 multiple choice items and scores range from 0 (worst outcome) to 48 (best outcome).

The Harris Hip Score

The HHS is another frequently used tool. It includes 10 items (maximum score of 100 denoting 'best possible outcome') and consists of four domains: pain (severity, effect on activities, need for pain medication), function (daily activities – stair climbing, sitting, managing shoes/socks; gait – limp, support needed, walking distance), absence of deformity (hip flexion, abduction, internal rotation, extremity length) and range of motion (hip flexion, abduction, internal/external rotation and adduction).[31]

Other commonly used measures include the Hip Disability and Osteoarthritis Outcome Score (HOOS),[29] the Merle d'Aubigné and Postel hip score[32] and the Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH).[33–35]

Current service provision

Management of disease

Treatment and management of arthritis in the UK can be categorised as non-surgical and surgical as detailed below. Patients in the early stages of OA begin treatment with non-surgical options; when non-surgical management has failed, patients are considered for intervention with surgical treatment.

Non-surgical management:

- self-management and patient education
- non-pharmacological (acupuncture, exercise, physical therapy, manual therapy, weight reduction)
- pharmacological [simple analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), topical treatments, intra-articular steroid injections].

Surgical management:

- surgery [e.g. THR or resurfacing arthroplasty (RS), arthrodesis, arthroscopy, osteotomy].

Current service cost

Arthritis has a significant negative impact on the UK economy with an estimated total cost of 1% of gross national product.[36] It is the most common group of conditions for which people receive Disability Living Allowance in England. The benefits provided outweigh those provided for people diagnosed with heart disease, stroke, chest disease and cancer combined.[36] A reported £43M is spent annually on community services and £215M on social services for OA.[36] In 2002 an estimated 36 million workdays were lost because of OA, resulting in £3.2B of lost productivity.[36] Data for the numbers of people who have their symptoms managed by non-surgical interventions (such as pain relief, exercise, physical therapy and manual therapy) within England and Wales are difficult to ascertain.

Chen *et al.*[8] estimated the cost of topical and oral NSAIDs using prescribing data from 2005/6. They reported that an estimated 167,000 people with a diagnosis of OA were found to have been prescribed topical NSAIDs and 1.4 million patients were prescribed oral NSAIDs. The annual costs were £8.5M and £25M, respectively.[8] Adjusting for inflation they found that this would equate to £19.2M and £25.65M, respectively in 2010. Most health economic analyses have reported that surgery for the treatment of arthritis is a cost-effective intervention and maximises cost per quality-adjusted life-year (QALY) gained.[37]

An earlier Health Technology Assessment (HTA) report (reference number 01/21/01)[19] found that the annual cost to the NHS of elective hip replacement surgery for the treatment of OA was £140M and that each trust spent, on average, £257,000 on the purchase of hip prostheses in 1998/9. This study was conducted in 2002.[19] It reported that the cost to the NHS and social services of non-surgical treatment for an individual was approximately £640 per person per year. During the year 2000, £405M was spent on 44,000 hip and 35,000 knee replacements.[36] Since then the costs have increased substantially, as the estimated cost to the NHS of THR surgery alone in 2011 was reported to be £426M.[36]

The cost of one surgical treatment in 2002 was £3891, averaged across all NHS trusts in 1999/2000, with the cost for 50% of trusts falling within the range £3404–4434.23.[19] According to the 8th Annual Report of the National Joint Registry for England and Wales (NJR),[36] the cost of hip replacement surgery varies considerably from trust to trust in the UK, with no set national price for implants. The cost depends considerably on length of hospital stay. For example, the tariff reimbursement paid to a trust in one study in 2005/6 for a primary THR was £6000 whereas, in 2010, the national tariff was set at £5552 for an uncomplicated THR.[36]

When hip replacement surgery fails, revision surgery to replace part or all of the prosthetic hip joint may be required. The number of revision procedures has increased in recent years, with 3012 carried out in 2003/4, rising to 6581 by 2008/9.[36] This accounted for approximately 9.4% of all elective hip replacement procedures performed in England and Wales.[36] Revision surgery is also a key element of the current service expenditure, with unit costs of revision generally higher than those for primary surgery. Briggs *et al.*[38] reported a mean cost for a standard hip revision procedure in 2000/1 as £5294 (£6385 in 2008 prices) compared with £3889 (£4690 in 2008 prices) for a primary procedure. The 2002 HTA report[19] stated that, in 1989/90, one in seven of all procedures (5000 out of a total of 35,000) was a revision of a hip replacement. In 1999/2000 a crude estimate of 6700 revisions was reported.[19]

Randomised controlled trials (RCTs) have compared revision rates across prosthesis types but with insufficient sample sizes or durations of follow-up to produce conclusive results.[39] The largest observational study found that 7-year revision rates were lower for cemented (3.0%) than for hybrid (3.8%) or cementless (4.6%) prostheses.[36] Edlin *et al.*[40] reported that a total of 97% of UK hip replacements are still working (unrevised) at 5 years.

Variation in services and uncertainty about best practice

Outcomes for hip replacement surgery vary by geographical location, surgeon and hospital. The Global Orthopaedic Registry has shown that patient selection criteria vary between practitioners, surgeons and referring doctors and between countries.[41] Nationally, there are reported inconsistencies in the treatment, procedure and prostheses that are offered to patients in the NHS.[42]

In 1998 more than 60 hip prostheses manufactured by 19 companies were available commercially in the UK, with total NHS expenditure of approximately £53M.[43] By 2008 this had risen to 124 brands of acetabular cups and 137 brands of femoral stems at a cost of £67M.[36] This represents a substantial increase in the variety of available prostheses in recent years. Implants are often grouped into cemented, cementless and hybrid prostheses.[44] The reported increasing use of cementless components in the UK has contributed to a doubling of prosthesis costs between 1996 and 2006.[44]

There is variation in the rate of primary hip replacement expenditure in England per 1000 population weighted by age, sex and need. For example, hip RS accounts for 6% of the approximate 70,000 hip arthroplasty operations conducted in England and Wales every year, although the equivalent figure among men aged < 55 years is 33%.[40]

Spend also varies significantly between regions in the UK, with the lowest reported in Tower Hamlets (£560) and the highest in Devon (£8140).[42] When examining data by local authority, the difference in the rate of provision of hip replacements per 1000 people in need was almost 14-fold.[42] National European Quality of Life-5 Dimensions (EQ-5D) data after hip replacement for England and Wales show that variation between the best and worst trusts is large (31–49%) and cost-effectiveness varies considerably between hospitals.[45]

Relevant national guidance

In the UK, the National Collaborating Centre for Chronic Conditions (NCC-CC) of the Royal College of Physicians developed clinical practice guidelines for OA.[3] The National Institute for Health and Care Excellence (NICE) developed clinical guidance on the selection of prostheses for primary THR[46] and metal-on-metal hip RS.[25]

Summary of National Institute for Health and Care Excellence guidance on the selection of prostheses for primary total hip replacement

The 2000 technology appraisal (TA)2[46] stated that the 'benchmark' for the selection of prostheses for THR should be a revision rate of ≤ 10% at 10 years with evidence relating to data from adequately sized, well-conducted observational studies or RCTs. NICE recommended that various patient factors, including age and underlying pathology, should be taken into account when choosing prostheses, for example ease of revision (of particular importance for younger patients).

Specific recommendations on the selection of hip prostheses for primary THR were considered difficult to construct because the evidence base was generally poor and difficult to interpret. However, the available evidence supported the use of a range of cemented prostheses for primary THR. This was further supported by the evidence on immediate and long-term postoperative pain.

There are currently no cost-effectiveness data based on revision rates after ≥ 10 years of follow-up to support the use of the generally more costly cementless and hybrid hip prostheses. Some evidence suggested that these types of prostheses might lead to less bone loss, meaning that they were potentially easier to revise than cemented prostheses. However, no reliable evidence was available to support the proposition that the potential ease of revision of a hip prosthesis would outweigh its poorer revision rate.

Summary of National Institute for Health and Care Excellence guidance on the use of metal-on-metal hip resurfacing arthroplasty

In the June 2002 NICE guidance TA44,[25] metal-on-metal hip RS was recommended as one option for people with advanced hip disease who would otherwise receive, and are likely to outlive, a conventional primary hip replacement. It did note, however, that the current evidence was principally in individuals aged < 65 years and that surgeons should bear this in mind. Furthermore, the guidance stated that all patients receiving this arthroplasty should be made aware of the relative paucity of evidence for medium- to long-term safety and reliability and the likely outcome of revision surgery compared with that for conventional THR.

However, in June 2012 advice about follow-up of patients receiving a metal-on-metal articulation changed. The Medicines and Healthcare products Regulatory Agency (MHRA) issued a medical device alert[47] stating that a small number of patients implanted with these hips might be at risk of developing progressive soft tissue reactions to the wear debris associated with metal-on-metal articulations; this updated the original advice of April 2010. These reactions could also adversely affect the results of later revision surgery. However, it also stated that its evidence pointed to the fact that early revision of such poorly performing metal-on-metal articulations should give a better revision outcome. Therefore, the agency advised that clinicians should perform appropriate follow-up, depending on which group a patient's hip surgery fitted into, as well as whether the patient was symptomatic or asymptomatic. Follow-up, if indicated, should consist of both imaging (MRI or ultrasound) and blood metal ion tests [ion level greater than seven parts per billion indicates the potential for soft tissue reaction]. Revision should be considered if imaging is abnormal and/or blood metal ion levels are rising.

Summary of Medicines and Healthcare products Regulatory Agency alert advice

Metal-on-metal hip RS implants:

- symptomatic: follow-up annually for life of implant
- asymptomatic: follow-up according to local protocols – no need for investigations unless cause for concern about cohort or patients who become symptomatic.

Metal-on-metal THRs with a head diameter < 36 mm:

- symptomatic: follow-up annually for life of implant
- asymptomatic: follow-up according to local protocols – no need for investigations unless cause for concern about implant.

Metal-on-metal THRs with a head diameter ≥ 36 mm:

- annual follow-up for life of implant whether symptomatic or not.

DePuy ASR™ hip replacements (all types) (DePuy, West Chester, PA, USA):

- annual follow-up for life of implant whether symptomatic or not.

National Institute for Health and Care Excellence guidance on the care and management of osteoarthritis in adults

The most recent NICE guidance on OA, issued in February 2008,[3] stresses the importance of a holistic assessment of the patient, including his or her function, quality of life, occupation, mood, relationships and leisure activities. After this assessment, the clinician is advised to formulate and agree a management plan with the patient, which should include 'core treatments' such as education, muscle strengthening and aerobic exercise, and weight-loss programmes for the overweight or obese. It should also include other self-management and 'conservative' strategies such as application of heat/cold packs or transcutaneous electrical nerve stimulation to the site of pain, manipulation and stretching (particularly for hip OA) and assessment for bracing/joint supports/insoles/walking sticks.

Adjuncts to the above 'core' treatment could include pharmacological treatments, in particular paracetamol (regular dosing may be required) and topical NSAIDs or topical capsaicin (although topical treatments are less useful for hips). If these are found to be insufficient for relieving pain, practitioners are advised to consider adding opioid analgesics or oral NSAIDs. Intra-articular corticosteroid injections are recommended for moderate to severe pain. Clinicians are advised to consider a referral for joint surgery if the patient has already been offered the 'core' treatments and is still experiencing joint symptoms that have a substantial impact on quality of life.

The Orthopaedic Data Evaluation Panel

The Orthopaedic Data Evaluation Panel (ODEP) was established to provide an independent assessment of clinical evidence, submitted by suppliers, on the compliance of their implants for THR and hip RS with NICE benchmarks for safety and effectiveness. ODEP produced detailed criteria for this assessment and in 2010 there was an ongoing review of this guidance by all stakeholders.[36] ODEP does have to rely on the honesty of the submitting companies and therefore provides no warranty that the data in its database are accurate, complete or current.

For 10-year benchmark products (those recommended to last for 10 years), ODEP places them in one of four categories according to whether there is evidence that a product meets NICE guidelines:

- level A – strong evidence that product meets NICE guidelines
- level B – reasonable evidence that product meets NICE guidelines
- level C – weak evidence that product meets NICE guidelines
- unacceptable – unacceptable evidence that product meets NICE guidelines.

For products that fail to meet NICE's 10-year benchmark, ODEP looks at evidence at 3, 5 and 7 years. Again, these products are split according to whether there exists acceptable, weak or unacceptable evidence for the product meeting NICE guidelines.

As of March 2011, ODEP ratings had been given to 38% of available brands of femoral stems and 41% of available brands of acetabular cups used in primary procedures. However, 42% of available brands of acetabular cups and 47% of available brands of femoral stems being used in England had not yet submitted data to ODEP. Clearly, for surgeons to make the most informed choices, it is important that all manufacturers submit their product data to ODEP using the pro forma and associated guidelines.

Description of the technology under assessment

Summary of total hip replacement

The predominant surgical intervention for the treatment of arthritis in England and Wales is THR, using a variety of cemented or uncemented stemmed femoral prostheses articulating with a cup that fits into the acetabulum. In 2011, 80,314 hip procedures were carried out in England and Wales; this rose to 88,599 in 2012.[48] THR has been so successful in treating hip OA that it has been described as the operation of the 20th century.[49] The average age of patients undergoing a hip replacement in 2010 was 67.2 years. There was a 3% increase in the percentage of women undergoing a THR in 2010/11 (59%) compared with 2009. On average, female patients were older than male patients at the time of their THR (68.8 years and 66.3 years, respectively).[36]

Modern THR began in the 1970s with widespread use of the Charnley prosthesis (DePuy, West Chester, PA, USA). More than 80,000 procedures are performed every year in England and Wales, with excellent clinical outcomes showing > 95% implant survivorship at 10 years' follow-up and > 80% implant survivorship at 25 years' follow-up.[41]

Rates for primary and revision THR have been increasing, with a 16% increase recorded in the UK between 2005 and 2010.[41] Although rates are 1.5–2 times higher for women than for men, THR is becoming more common for both sexes and for those in younger age groups. The greatest proportion of procedures (65%) is carried out in patients aged ≥ 65 years. However, the proportion of patients undergoing THR who are aged < 65 years is projected to increase to 50% of all arthroplasties by 2030.[41]

The decision to undertake THR is guided by symptoms (pain, functional impairment) and by physical examination and radiographic findings. Patients presenting with hip pain will follow a care pathway similar to the one presented in the following section.

In the early stages, non-surgical treatment options will be provided such as exercise and physical therapy. Non-surgical options are used until the point at which they are deemed to have failed. The patient is then referred to an orthopaedic specialist for secondary assessment and possible surgical intervention. Indications for THR surgery in the UK are:

- OA (93%)
- avascular necrosis (2%)
- fractured neck of femur (2%)
- congenital dislocation (2%)
- inflammatory arthropathy (1%).[48]

The success of surgical intervention can be influenced through patient selection. Assessment of patient and prosthesis outcomes is necessary to identify which designs or surgical techniques provide the best patient benefit. Relative contraindications to THR include severe obesity, advanced age and other medical comorbidities. There is a reported 40% increased risk of complications for every decade above the age of 65 years.[41] THR in younger patients, who are typically more active, is problematic because of the risk of poor prosthesis survivorship over a patient's lifetime. Waiting time for surgery should also be considered as it can be an important factor in patient outcomes following THR. Under the current waiting time targets, people in England should not have to wait longer than 18 weeks for their hip replacement surgery once it has been recommended.

Example patient care pathway for hip arthroplasty

Figure 1 presents a typical care pathway for patients treated for arthritis in the NHS. In general, patients would be treated in primary care services and undergo various non-surgical management options. Once non-surgical management is said to have failed, the patient is classified as having end-stage arthritis and is recommended for surgery in secondary care.

Figure 2 presents the two surgical options THR and hip RS. The care pathways are similar in terms of pre- and postoperative care and follow-up.

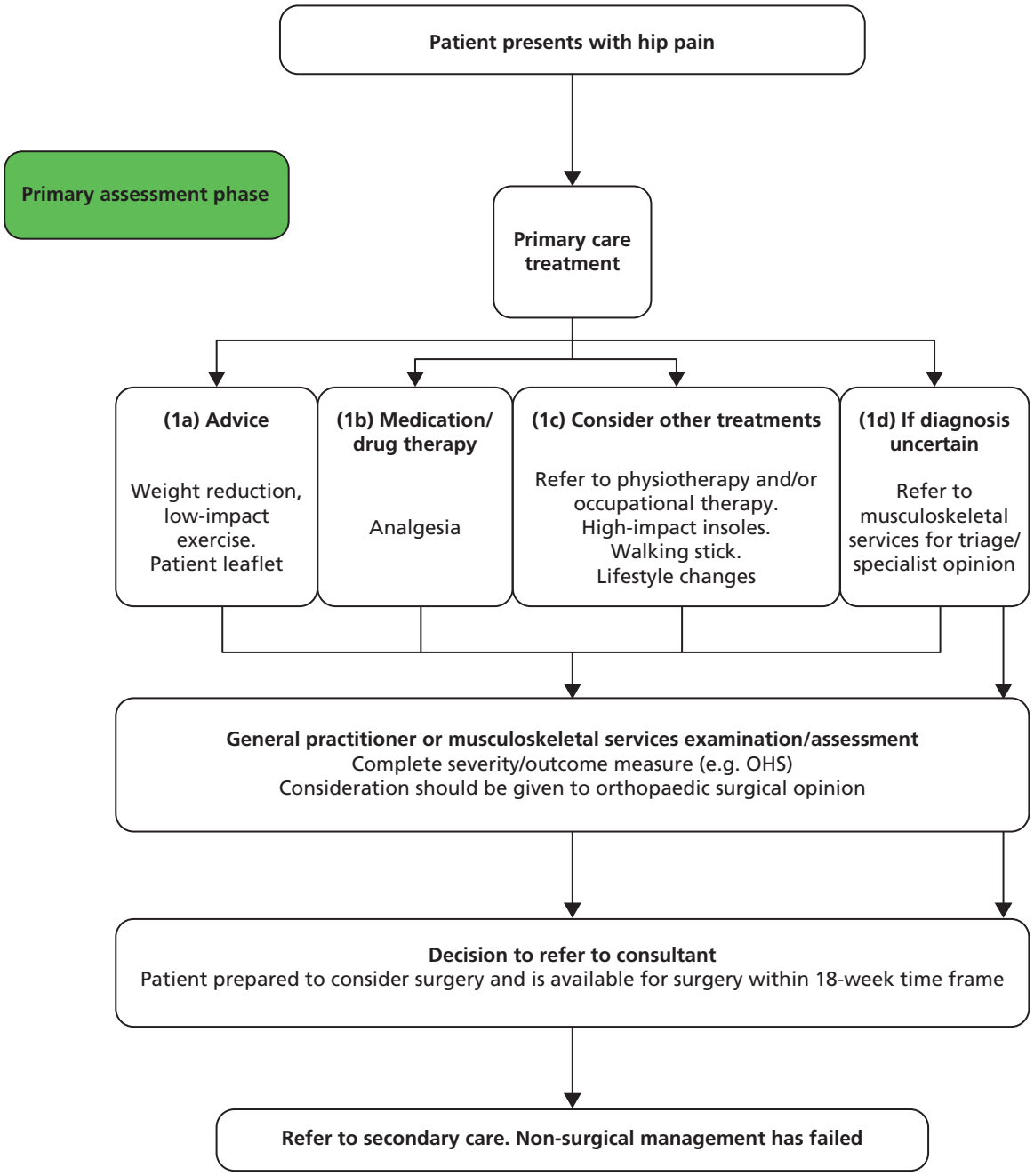


FIGURE 1 Example pathway for patient with arthritis in primary care.

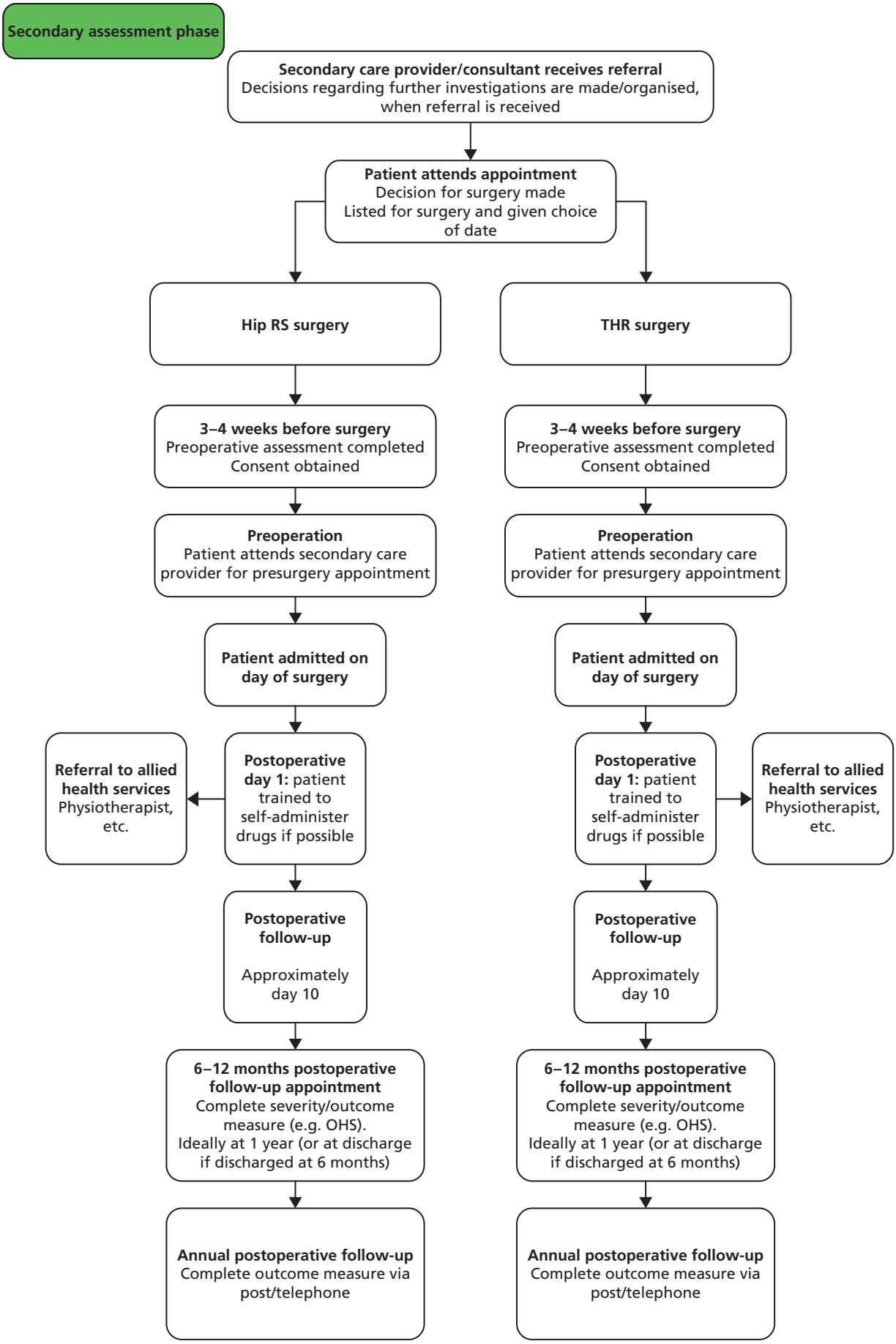


FIGURE 2 Example hip replacement care pathway in secondary care.

Identification of different types of total hip replacement

The different types of THR can be categorised into the following subgroups:

- hip replacement with different fixation methods for implant components (cemented, cementless, hybrid or reverse hybrid prostheses)
- hip replacement with implant components (i.e. femoral stem, femoral head, acetabular cup) made from different materials (metal, ceramic, polyethylene)
- hip replacement with differing femoral head sizes.

Hip replacement with different fixation methods

Hip replacement prostheses can be categorised by their fixation method (*Figure 3*) as (a) cemented, (b) cementless, (c) reverse hybrid with a cemented cup and cementless stem or (d) hybrid with a cemented stem and cementless cup. Cemented prostheses are held in place with bone cement and generally consist of three components: a femoral stem, a femoral head (modular) and an acetabular cup. These components are permanently attached to the pelvis and the femur. According to the NJR, the percentage of cemented procedures did not change between 2009 and 2010. The number of cemented procedures had been in decline since 2005. In 2004 the figure was at 77%, and by 2010 this had reduced to 50%.[36]

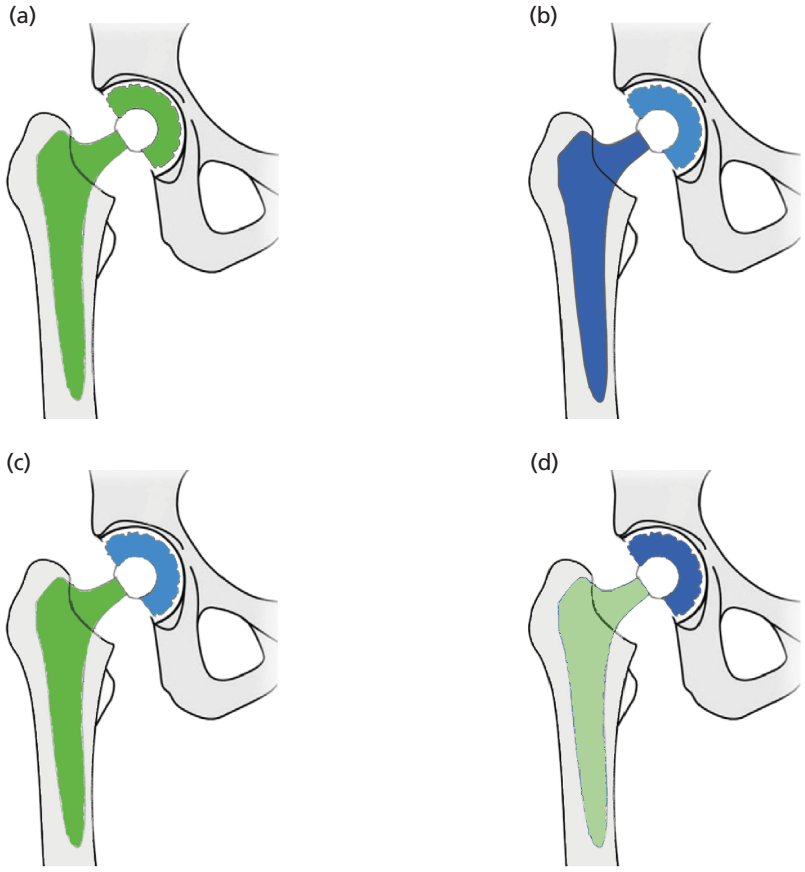


FIGURE 3 Overview of four different fixation options for the femoral stem and acetabular cup in THR. (a) Cemented THR; (b) cementless THR; (c) reverse hybrid THR – cementless stem with a cemented cup; and (d) hybrid THR – cemented stem with a cementless cup.

Cementless prostheses rely on initial press-fit fixation followed by natural bone growth. They typically consist of four components: a femoral stem, a femoral head, an acetabular cup shell and an acetabular liner. The theoretical benefit of the cementless fixation is the possibility of bone–implant interface (human : technology) remodelling. In England and Wales there has been a 4% increase in cementless procedures in recent years.[36]

The cementless prostheses include implant components coated in a porous material (hydroxyapatite) that is compatible with bone growth and which helps to secure the liner in place. Hydroxyapatite is a mineral form of calcium apatite.[50] Hydroxyapatite is also commonly used as a filler to replace amputated bone in addition to a coating to promote bone ingrowth into prosthetic implants.

A hybrid hip replacement consists of a cemented femoral stem and a cementless acetabular cup, whereas the reverse hybrid uses a cementless femoral stem and a cemented acetabular cup. In 2010, 14% of these types of procedure were reverse hybrid (cementless stem, cemented acetabulum) and 86% were standard hybrid (cemented stem, cementless acetabulum).[36]

Hip replacement with components made from different materials

The combinations of prosthetic components that are available are listed in *Table 1*. The different materials used for the implant components (i.e. femoral stem, femoral head, acetabular cup) produce various articulating surfaces or bearing surfaces.

TABLE 1 Total hip replacement and RS articulation and fixation type combinations

Femoral head (press-fit)	Fixation method	Femoral stem	Acetabular cup[a]	Acetabular cup shell	Acetabular liner
THR articulation type					
Metal	Cemented	Metal	Polyethylene	–	–
Metal		Metal	*Metal*	–	–
Ceramic		Metal	*Polyethylene*	–	–
Ceramic		Metal	*Ceramic*	–	–
Ceramic	Cementless	Metal	–	Metal	Ceramic
Metal		Metal	–	Metal	Polyethylene
Metal		Metal	–	Metal	Metal
Ceramic	Hybrid (cemented femoral stem and a cementless acetabular cup)	Metal	–	Metal	Ceramic
Ceramic		Metal	–	Metal	Polyethylene
Metal		Metal	–	Metal	Metal
Metal		Metal	–	Metal	Polyethylene
Metal	Reverse hybrid (cementless femoral stem and a cemented acetabular cup)	Metal	Polyethylene	–	–
Metal		Metal	*Metal*	–	–
Ceramic		Metal	Polyethylene	–	–
Ceramic		Metal	*Ceramic*	–	–
RS articulation type					
–	Cemented	Metal	Metal	–	–
–	Cementless	Metal	Metal	–	–
–	Hybrid	Metal	Metal	–	–

a Components in italics are rarely used in clinical practice.

The NJR report for 2011[36] provided the percentage use of fixation type during 2010 and 2011 (*Table 2*). The cemented fixation type was the most popular fixation method and the polyethylene-on-metal articulation combination was used the most (86.1%) of all the cemented bearing surfaces. The cementless fixation type was the second most common fixation method and the polyethylene-on-metal articulation combination was most popular (35.6%).

Another way of characterising the variation of combination of articulation surface and fixation method is by frequency of use, as reported in the NJR. The most common combinations are listed in *Table 3* along with the associated acronyms that have been used in the remainder of this report.

Polyethylene-on-metal (cup material-on-femoral head material)

A metal ball with polyethylene cup (or polyethylene liner inside a metal cup) (*Figure 4*) is the most common type of articulation combination (both cemented and cementless) and is one of the cheapest. The Charnley low-friction arthroplasty was the first widely accepted polyethylene-on-metal prosthesis to be used. It has a high reported implant survivorship at > 20 years' follow-up (> 80%) and at 35 years' follow-up (78%).[41] It also provides the baseline against which new prosthetic designs are compared. In England and Wales this was the most common articulation type used during 2010 and 2011 (see *Table 2*). Clinical advice suggested that, if a metal cup is used with a polyethylene liner, a cementless cup fixation is most commonly used in England, and the cementing of the metal cup is increasingly rare. Highly cross-linked polyethylene is being used by some surgeons in place of standard polyethylene in THRs because of its lower reported wear rates.[51,52]

TABLE 2 Percentage spread of fixation type and bearing/articulation surface for primary hip replacements reported in the 2011 NJR report (2010/11)[36]

Articulation combination (cup material-on-femoral head material)	Cemented ($n = 132,511$)	Cementless) ($n = 102,688$)	Hybrid[a] ($n = 43,933$)	All ($n = 279,132$)
Other/unknown	2.9	5.7	3.8	4.0
Ceramic-on-ceramic	1.8	25.6	15.1	12.6
Polyethylene-on-ceramic	8.4	14.2	11.7	11.0
Metal-on-metal	0.9	18.9	3.0	7.9
Polyethylene-on-metal	86.1	35.6	66.5	64.4

a The 2011 NJR report[36] did not distinguish between hybrid and reverse hybrid fixation methods.

TABLE 3 Combination of bearing/articulation surface and fixation method by frequency as reported in the 2011 NJR report (2010/11)[36]

Implant characteristics	Acronym used in the report[a]
Metal head (cemented stem) on cemented polyethylene cup	CeMoP
Metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)	CeLMoP
Ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)	CeLCoC
Hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)	HyMoP
Metal head (cementless stem) on cementless non-HA-coated metal cup (polyethylene liner)	CeLMoP (non-HA)
Ceramic head (cemented stem) on cemented polyethylene cup	CeCoP
Hybrid metal head (cemented stem) on cementless non-HA-coated metal cup (polyethylene liner)	HyMoP (non-HA)

HA, hydroxyapatite.
a Acronym = (fixation type), (femoral head material), (cup liner material).

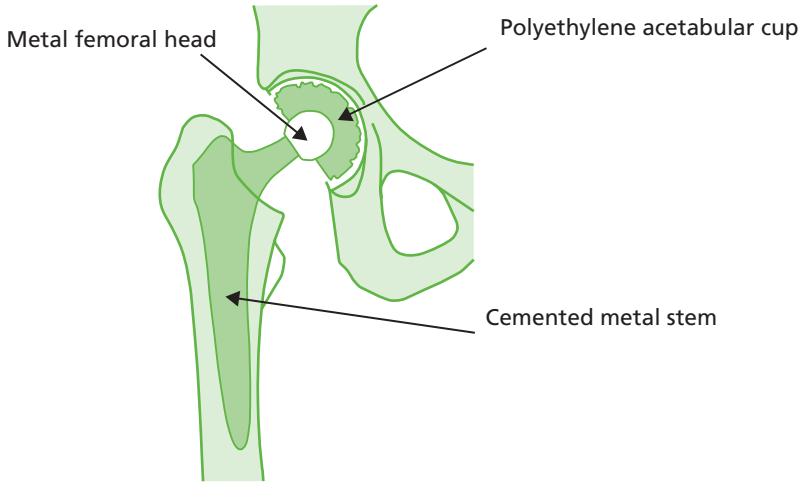


FIGURE 4 Cemented metal stem, metal femoral head and polyethylene acetabular cup.

Polyethylene-on-ceramic

The polyethylene-on-ceramic option combines a polyethylene cup with a hard ceramic femoral head (*Figure 5*). This articulation type is reported to have a lower wear rate than the polyethylene-on-metal combination and is cheaper than the ceramic-on-ceramic option. It is used more often with a cementless fixation (14.2%) than with a cemented fixation (8.4%) (see *Table 2*). The ceramic head is harder than metal and hence reportedly withstands more wear. In the past ceramics were brittle and cracked, leading to failure of the implant, but advances in technology have limited this problem in recent years.

Ceramic-on-ceramic

The ceramic-on-ceramic articulation (*Figure 6*) provides the hardest bearing surface combination and is generally the most expensive combination available.[40] This combination has a lower reported wear rate than other options available to patients in England and Wales. The ceramic-on-ceramic articulation is mostly used without cement, as shown in *Table 2* (25.6% cementless vs. 1.8% cemented). Clinical advice suggests that the cementless ceramic cup is the most common type of ceramic-on-ceramic articulation in England; cementing the ceramic cup is increasingly rare, as demonstrated in the NJR data.[36]

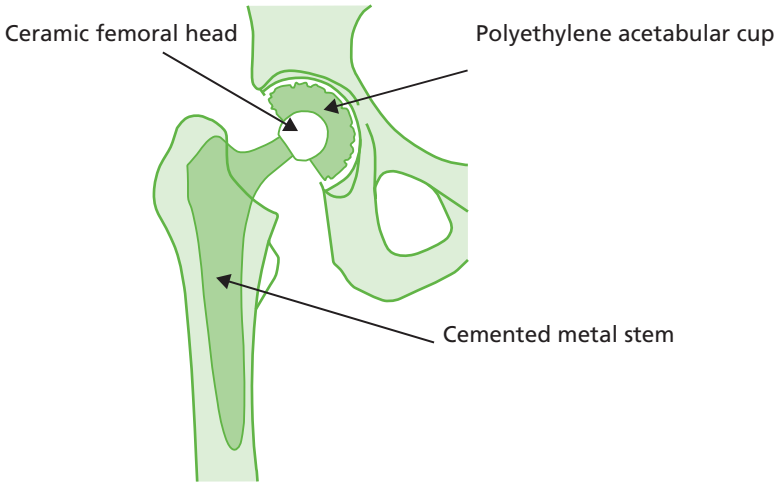


FIGURE 5 Cemented metal stem, ceramic femoral head and polyethylene acetabular cup.

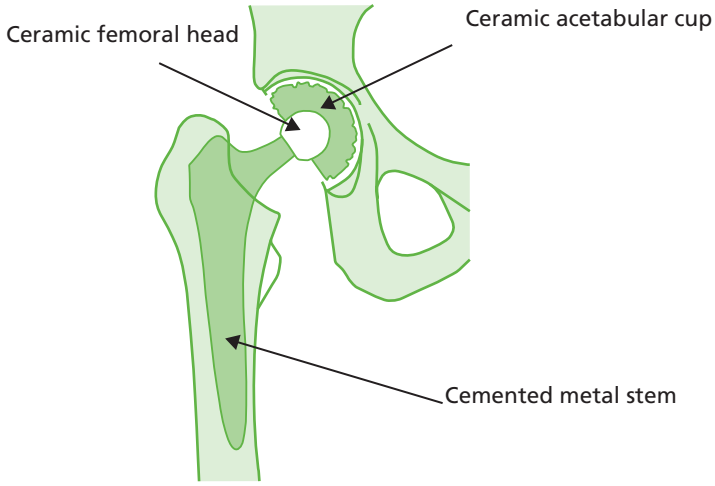


FIGURE 6 Cemented metal stem, ceramic femoral head and ceramic acetabular cup.

Metal-on-metal

Metal-on-metal articulations (*Figure 7*) provide a hard bearing surface; however because of their reportedly high revision rate the MHRA has made recommendations for following up patients implanted with such devices.[47]

The MHRA recommendations apply to four groups of metal-on-metal replacements:

1. metal-on-metal hip RS implants
2. metal-on-metal THRs with a head diameter < 36 mm
3. metal-on-metal THRs with a head diameter ≥ 36 mm
4. DePuy ASR hip replacements comprising:
   - ASR acetabular cups for hip RS or THR
   - ASR surface replacement heads for hip RS
   - ASR XL femoral heads for THR.

Revision is necessary when prostheses fail, more common in younger patients, usually for loosening secondary to wear or dislocation. Interestingly, metal-on-metal bearing surfaces were actually designed by surgeons to reduce the proportion of replacements that require revision. They had been extensively assessed in simulator tests and were noted to be highly resistant to wear, even when used in very large head sizes.[53]

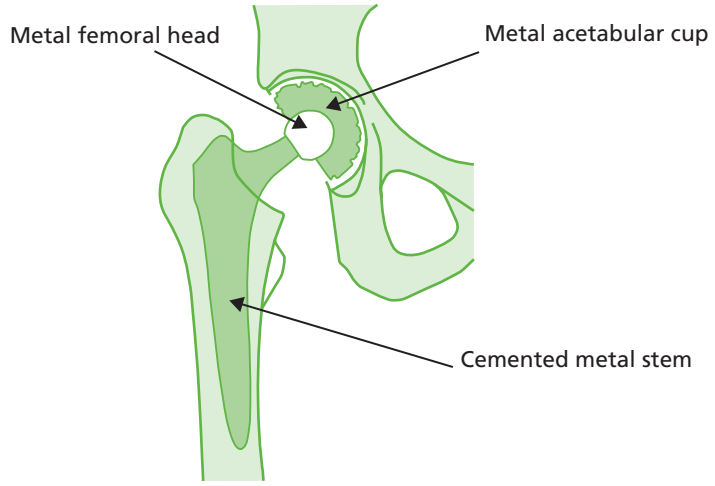


FIGURE 7 Cemented metal stem, metal femoral head and metal acetabular cup.

Head size is important because in simulator tests larger head sizes give lower wear because of the boundary lubrication regime becoming more favourable.[54] Therefore, implantation of large diameter metal-on-metal bearing surfaces on stemmed prostheses became popular on the basis of such evidence, which suggested that they should result in less wear and thus lower failure rates. They seemed to be particularly appropriate for younger, more active patients.

However, several issues have arisen with the practical use of these metal-on-metal prostheses. It soon emerged that one brand of metal-on-metal prosthesis, the DePuy ASR, actually seemed to fail early.[55] Data received from the company[56] showed that 5 years after surgery 12% of patients who received the ASR RS and 13% of patients who received the ASR THR required revision surgery.

This prompted recent analysis of NJR data on 402,051 hip replacements to assess whether metal-on-metal bearing surfaces lead to increased implant survival compared with other bearing surfaces in stemmed THR.[16] These authors additionally challenged the previous evidence that larger head sizes result in improved implant survival.

The results revealed that, in THR, metal-on-metal articulations failed at higher rates than other bearings. For example, 5-year revision rates in younger women were 6.1% [95% confidence interval (CI) 5.2% to 7.2%] for 46-mm metal-on-metal articulations compared with 1.6% (95% CI 1.3% to 2.1%) for 28-mm polyethylene-on-metal articulations. This effect was found even though the ASR data had been removed before analysis (the DePuy ASR articulations had already been removed from the market). Thus, it is a problem with all metal-on-metal prostheses, not an implant-specific characteristic. In addition, their failure was found to be related to head size, with larger heads failing earlier than smaller versions (this effect was the opposite to that found for ceramic-on-ceramic articulations). The authors suggested a number of potential reasons for the finding that larger metal heads fail earlier, such as a failure to achieve optimum lubrication or trunnion (post that inserts into head) wear[55] resulting in metal debris leading to local soft tissue reactions[57] or early loosening because of increased transmitted torque from the larger head. The authors of the paper therefore suggested that metal-on-metal replacements not be performed because of poor implant survival and that patients undergo at least an annual review with both clinical and radiological examination, in line with the MHRA recommendations.[47]

Furthermore, there are the potential dangers of exposure to metals such as chromium and cobalt. Metal alloys used in metal-on-metal bearings degrade through wear, from corrosion or because of a combination of the two.[58] Consequently, they produce a vast number of nanometre- to submicrometre-sized metal particles that cumulatively present a large surface area for corrosion.[59] This is also relevant to the polyethylene-on-metal bearings, which also produce such particles through wear. The consequences of local and systemic exposure to the wear particles and the accompanying biologically active corrosion products have been extensively researched.[60] It is well known that metal debris can induce adverse local soft tissue reactions[41] including the release of inflammatory cytokines from macrophages, histiocytosis, fibrosis and necrosis.[61] Local results include aseptic loosening because of osteolysis induced by some immunological reaction involving hypersensitivity[62] and local pseudotumours (soft tissue masses relating to the joint) that are locally destructive and require revision surgery in the majority of patients.[63]

Furthermore, it seems that metals can disseminate through the body and cause direct damage to end organs such as the kidneys, lungs and brain.[64,65] There is also evidence of genotoxicity and evidence that these metals can signal across biological barriers at concentrations produced after THR.[66] The genotoxic effects of the metal ions are thought to be mediated either by direct action, causing DNA breaks through attacks on free radicals, or through an indirect effect by inhibiting the repair of DNA.[67] There have been concerns that this genotoxicity could cause a long-term increased risk of malignancy, particularly important for the younger, more active patients in whom life expectancy after implantation is long. However, recent studies have failed to find this increase[68] and some have actually found a decrease in the numbers of certain malignancies in metal-on-metal articulation patients.[69]

The US Food and Drug Administration (FDA),[70] the UK MHRA[47] and the British Orthopaedic Association[71] have released statements of concern about metal-on-metal articulations. The MHRA recommendation states that patients with metal-on-metal bearings and a painful hip joint should have yearly measurements of whole blood metal ion concentrations and radiographic assessment to exclude adverse local tissue reactions as the source of pain.[47] These yearly assessments should continue for the lifetime of the hip replacement. The use of metal-on-metal bearing surfaces has consequently declined in England and Wales. In 2010/11 only 7.9% of all procedures used a metal-on-metal implant (see *Table 2*). However, data suggest that they are still being used extensively in other countries. For example, in the USA, 35% of articulations were metal-on-metal in 2009.[72]

Hip replacement with differing femoral head sizes

Research has suggested that differing femoral head sizes leads to variation in the rate of revision. Smith *et al.*[16] reported that the use of larger head sizes (> 36 mm in diameter) improves stability and range of motion compared with the smaller head diameters that are used with other bearing surfaces. Use of large diameter femoral heads increases the distance that the head must travel before dislocation, without decreasing hip range of motion, thus increasing stability.[41]

Summary of hip resurfacing arthroplasty

Hip RS has been developed as a surgical alternative to THR. It is reported to be an option that is predominantly suited to younger, active, male patients.[46] The procedure consists of placing a cobalt–chromium metal cap over the head of the femur while a matching metal cup (similar to that in THR) is placed in the acetabulum. This replaces the articulating surfaces of the hip joint and is bone-conserving compared with THR (*Figure 8*). According to clinical advice, in NHS practice the metal cup is generally cementless and the femoral metal head can be cemented or cementless.

In 2011 patients were on average 54.8 years of age when they underwent RS. Four times as many men underwent this procedure as women.[36] According to the NJR 2011 report,[36] this shows good adherence by the orthopaedic community to guidelines issued by the British Orthopaedic Association during 2009/10 on patient selection criteria for metal-on-metal RS prostheses. As with THR, patient selection is crucial for the outcome of RS.

The FDA has produced patient selection criteria for metal-on-metal hip RS. These include:

1. patient is fit and active
2. patient has normal proximal femoral bone geometry and bone quality
3. patient would otherwise receive a conventional primary THR
4. patient is likely to live longer than current conventional THR prostheses are expected to last.[73]

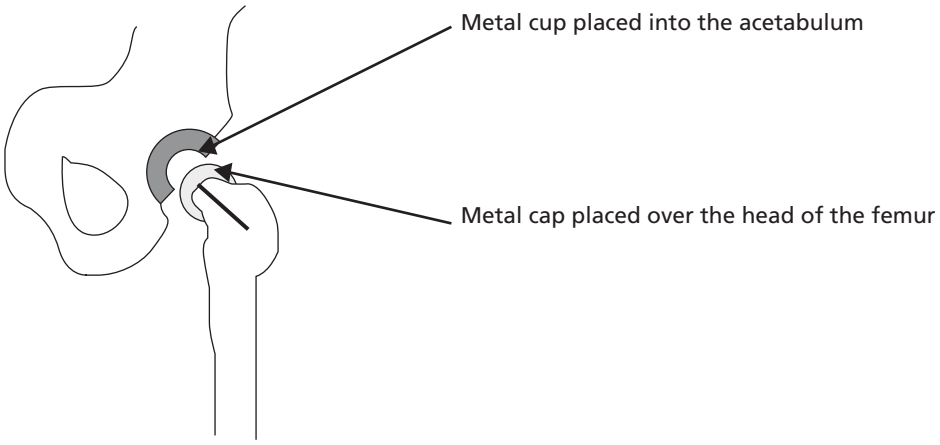


FIGURE 8 Diagrammatic representation of a hip RS.

Johnson *et al.*[74] reported 100% implant survivorship at 5 years' follow-up in 93 patients identified using narrow selection criteria who underwent RS. The selection criteria included avoiding RS in patients with large femoral head or neck cysts, ensuring proper seating of the femoral component band and ensuring an optimal thickness of the cement mantle. The authors of this study suggested that the best results were achieved in male patients aged < 50 years with a primary diagnosis of OA and a native femoral head > 50 mm in diameter.[74] Individual surgeon experience with hip RS is also an important factor and outcomes may differ between operators. Although positioning of the surgical component in RS is comparable in difficulty to that of THR, there is a learning curve that must be negotiated for surgeons inexperienced with the procedure.[41]

Since 2011 there has been a significant decrease in the percentage of RS procedures conducted in England and Wales. There has also been a reduction in the percentage of procedures using a large head implant for RS.[36] This is thought to be because of the withdrawal of the DePuy ASR device from the market following the identification of higher than expected revision rates for this product.

Failure of hip replacement

A hip replacement may fail because of peri- and/or postoperative complications such as implant instability, dislocation, aseptic loosening, osteolysis, implant fracture and infection.

Implant instability and dislocation

Instability and recurrent dislocation are the most common reasons for THR failure and the second most common cause of failure of revision THR. The prevalence of dislocation ranges from 0.3% to 10% for primary THR and is 28% for revision THR.[75–77]

The most common reasons for instability are component malpositioning and abductor (muscle) deficiency such as a loss of abduction power, which can lead to a severe limp. Cup malpositioning can lead to increased wear of particular sections of the prosthesis, for example both 45-degree inclination (tilting) and 20-degree anteversion (forward tilting) have been associated with THR failure.[78,79] However, age, previous fracture, surgical volume, surgical approach, component sizing and polyethylene wear are also contributory factors to revision because of instability and dislocation.[80–83]

Recurrent late dislocation remains a major source of THR failure. There are various treatment options for patients who have recurrent dislocations. These include revision surgery using constrained polyethylene liners (which offers increased stability but at the cost of smaller range of motion), larger diameter femoral heads and dual mobility devices.

Aseptic loosening and osteolysis

Aseptic loosening is a common cause of failure of THR. It arises because of osteoclast-mediated bone reabsorption at the bone–implant interface, which can lead to loosening, implant migration, implant failure and periprosthetic fracture.[84] Osteolysis is one of the most common complications after THR, which may lead to implant failure. It is initiated as a result of an inflammatory process against polyethylene particulate debris. Component malpositioning is a major cause of severe wear and osteolysis, but they are also affected by activity level and material and component design.[85]

Aseptic loosening and osteolysis are diagnosed clinically by patient reports of pain. They are treated with replacement of loose components and correction of component malalignment. Outcomes after revision surgery are generally good, with reported mechanical failure rates < 5% at follow-up.[86]

Periprosthetic fracture

Periprosthetic fracture is a major complication after THR and is associated with increased morbidity and mortality. Risk factors for periprosthetic fracture include previous revision surgery, component malalignment, age, osteoporosis, previous fracture and minor trauma.[87,88]

Treatment for most periprosthetic fractures is usually surgical. Treatment for most periprosthetic fractures is usually surgical and the options depend on the fracture pattern. It can include open reduction and internal fixation with or without cortical strut allografts, longer femoral stems or increases in the setting of acetabular fractures, or tumour prostheses.[89,90]

Infection

Infection of a THR prosthesis is associated with greatly increased morbidity, mortality and use of health-care resources. The infections can by treated with antibiotics; however, deep infections are rarely cured by antibiotics alone and may require revision surgery. As more THRs are performed, the absolute number of deep infections is likely to increase although, because of comprehensive infection control techniques, rates are relatively low. Risk factors for infection include age, obesity, comorbidities and American Society of Anesthesiologists (ASA) score. Longer operative times and reoperation within 90 days have been implicated as risks for infection.[91,92]

Revision of hip arthroplasty

Recent data demonstrated that 7-year revision rates were lower for cemented (3.0%) than for hybrid (3.8%) or cementless (4.6%) prostheses.[36] RCTs have compared revision rates across prosthesis types but with insufficient sample sizes or durations of follow-up to produce conclusive results.[39]

Factors affecting long-term prosthesis survivorship include patient-related factors such as comorbidities and patient activity levels.[41] Once an implant has failed, patients will have implant revision surgery. The rate at which hip replacements are revised is termed the revision burden.

In England and Wales the NJR keeps a record of whether each operation performed is a primary replacement or a secondary revision of a replacement. This allows trends to be followed to estimate how many revision operations are expected in the future, hence the revision burden (*Table 4*).

This shows a rise in the number and proportions of operations that are being conducted for revision of THRs over the last couple of years, which in real terms relates to around 3000 more revisions over the last 5 years. This may be because the recipients of the replacements are living longer and are thus outliving their THR or possibly may be because of more stringent follow-up. At NHS hospitals, revision procedures account for a higher percentage of the total procedures (13%) than at any other type of provider, with 84% of all revision procedures in 2010/11 being performed in the NHS.[36]

Clinical follow-up

Implants should be assessed every year for signs of loosening, migration/measure of prosthesis movement (e.g. femoral head penetration rate) and failure. Although no studies have examined the benefits of specific follow-up frequencies, NICE recommends continued periodic follow-up.

TABLE 4 Revision procedures by type and year as reported in the 2011 NJR report[36]

Procedure	2006/7	2007/8	2008/9	2009/10	2010/11
Hip primary, *n*	58,445	66,556	69,681	70,669	77,800
Hip revision, *n* (%)	6198 (9.6)	6725 (9.2)	7345 (9.5)	8285 (10.5)	9200 (10.6)
Total, *N*	64,643	73,281	77,026	78,954	87,000

Follow-up using radiostereometric analysis allows for precise quantification of any implant movement of the prosthesis; however, visual inspection of the radiograph by the surgeon is commonly used in clinical follow-up.[93] Evidence suggests that early detection of lesions (e.g. aseptic lymphocyte-dominated vasculitis) is more cost-effective than waiting until patients report pain and loss of function and an assessment is conducted.[94]

Disability, function, pain, limitations in daily activities, overall satisfaction and health-related quality of life should be routinely measured and documented at follow-up using validated instruments [e.g. Short Form questionnaire-12 items/Short Form questionnaire-36 items (SF-12/SF-36), EQ-5D].[27]

Current usage in the NHS

The following information was taken from the 8th Annual Report of the NJR.[36]

General statistics

- In total, 179,450 operations (hip, ankle and knee) were reported to the NJR in 2010, a 9.9% increase on the previous year.
- However, 15.8% of these operations were accounted for by operations performed in previous years being added to the register.
- The increase in numbers of hip and knee replacements over the last few years is the result of increases in the number of operations performed in England; Wales has not seen similar growth.

Hip replacement surgery

According to these 2010/11 data, 83,014 hip replacement operations (95%) took place in England and 4024 operations took place in Wales. There are four types of organisation in England carrying out hip replacement surgery (*Table 5*) (note: there are no NHS treatment centres or independent sector treatment centres in Wales).

There have been no major changes in these proportions over the last 5 years although there has been a constant, very slight increase in the proportion of operations carried out by NHS hospitals over this time period and a slight decrease in the proportion carried out by NHS treatment centres. Annual fluctuations between types of provider have been small and the proportion of operations for each type of provider in 2010/11 is within two percentage points of the figures from 2006/7. In total, 93% of patients at independent sector hospitals and independent sector treatment centres were reported to be 'fit and healthy' or with 'mild' disease (ASA grading system) compared with only 80% at NHS centres.

Type of procedure

The operations carried out across the NHS organisations can be categorised by procedure type as displayed in *Table 6*.

TABLE 5 Percentage of procedures by organisation type reported in the NJR[36]

Organisation type	Percentage of procedures in 2010/11
NHS hospitals	67
NHS treatment centres	3
Independent sector hospitals	26
Independent sector treatment centres	5

TABLE 6 Percentage of interventions by fixation method across NHS hospitals and treatment centres reported in the 2011 NJR report[36]

Procedure type	Overall (68,907 treatments)	NHS hospitals (44,054 treatments)	NHS treatment centres (2075 treatments)
Cemented	36	38	25
Cementless	43	42	66
Hybrid	3	17	4
RS	2	3	4

The percentage of primary hip RS undertaken in independent hospitals (5%) is nearly double that carried out at NHS hospitals. Interestingly, at NHS treatment centres, 66% of primary procedures are cementless hip primary procedures, a greater proportion than at any other type of provider.

Background summary

Arthritis is a general term describing pain and inflammation within a joint. It commonly affects the hip, which is a weight-bearing ball and socket joint. The most common causes of the arthritis syndrome are OA and RA.

Osteoarthritis is a degenerative disease in which the degeneration and consequent loss of articular cartilage are associated with synovial inflammation and bone hypertrophy. This leads to symptoms of pain, stiffness and loss of function and mobility. The degeneration can be primary (no specific cause identified) or secondary to a number of intra-articular diseases. Its prevalence is also increased by a number of risk factors including biomechanical, constitutional and genetic factors. OA is by far the most common arthritis of the hip and is diagnosed clinically and by imaging. There are difficulties in estimating the disease burden of OA because of variable diagnostic criteria. However, there are an estimated 2.8 million patients in the UK alone who have the disease and current projections estimate that 10% of the world's population aged ≥ 60 years will be affected at some point. Estimates of the annual incidence of RA suggest that 10,000–20,000 people develop RA in the UK each year. Although the disease may develop in patients at any age, onset is classically between the ages of 40 and 60 years. This is especially important in light of the ageing population as OA and RA mostly affect elderly people with comorbidities. Although the natural history of OA varies between affected joints, the prognosis of hip OA is particularly poor. Approximately 10–40% of cases of RA manifest within the hip joint.

The economic impact of arthritis is vast, both because of direct costs to the health-care system, community and social services and because of indirect costs from lost productivity and early mortality. In the present economic climate in which health-care spending must be carefully justified, the implications of increasing demand for the treatment of arthritis of the hip has led to intense discussion about the cost-effectiveness of new technologies and treatment options. To aid this comparison, different tools such as the OHS and the HHS have been developed and validated for the assessment and monitoring of patient outcomes.

Non-surgical and surgical treatments exist for the management of arthritis to provide symptomatic relief in the short term and to avoid progressive joint damage and improve quality of life in the longer term. Surgical options, including THR, are usually considered for patients with symptoms unmanageable through conservative management. The surgical interventions are believed to be cost-effective interventions that maximise cost per QALY gained. Patient selection criteria, amount spent and outcomes for hip replacement surgery vary across geographical location, hospital and surgeon. The NCC-CC and NICE have developed guidelines to assist clinicians with making clinical decisions about whether or not a patient requires a hip replacement; however, there still exist inconsistencies in surgeries offered at different NHS centres.

Total hip replacement is the predominant surgical intervention for the treatment of arthritis in the UK and is highly successful. Hip replacements can be categorised and compared according to their components, fixation methods, femoral head size and revision rates. For example, there are many different brands of prosthesis for a surgeon to choose from, with fixation types split into cemented, cementless or hybrid, in addition to the option of RS. Failure of the articulations and need for revision surgery are important considerations, especially considering the growing number of primary procedures that are taking place and the overall increasing revision burden. Requirements for revision include instability/dislocation, aseptic loosening and osteolysis, periprosthetic fracture and infection, and NICE recommends periodic follow-up to help identify such issues.

Chapter 2 Definition of the decision problem

Decision problem

This report aims to evaluate the clinical effectiveness and cost-effectiveness of THR and hip RS for the treatment of pain and disability in people with arthritis. More specifically, we aim to investigate, in people with pain and disability resulting from arthritis of the hip for whom non-surgical management has failed:

1. the clinical effectiveness and cost-effectiveness of different types of elective primary THR compared with primary hip RS in those suitable for both procedures
2. the clinical effectiveness and cost-effectiveness of different types of primary THR compared with each other in those not suitable for hip RS.

Overall aims and objectives

1. To undertake a systematic review of the clinical effectiveness and cost-effectiveness of (a) different types of primary THR compared with RS for people in whom both procedures are suitable and (b) different types of primary THR compared with each other for people who are not suitable for hip RS and to investigate factors that influence benefits and costs. If data are sufficient, the influence of patient- and intervention-related factors on the magnitude of treatment effects will be explored through subgroup analysis and meta-regression.
2. To further develop the cost-effectiveness and cost–utility models published in TA44[25] using updated NJR data and model inputs when available.
3. To report on findings and make recommendations for future research.

This report aims to evaluate the clinical effectiveness and cost-effectiveness of THR and RS for the treatment of pain and disability in people with arthritis [*Table 7* provides a summary of the population, intervention, comparator/control and outcome (PICO)].

TABLE 7 Population, intervention, comparator/control and outcome table

PICO	Final scope issued by NICE (17/01/13)[a]	Decision problem addressed in the assessment report
Population	People with pain or disability resulting from arthritis of the hip for whom non-surgical management has failed	People with pain or disability resulting from end-stage arthritis of the hip for whom non-surgical management has failed
Intervention	1. Primary THR 2. Primary hip RS	1. Elective primary THR 2. Primary hip RS
Comparators	Different types of primary THR and hip RS will be compared for people in whom both procedures are suitable Different types of primary THR will be compared for people in whom hip RS is not suitable The different types of hip replacement that will be considered separately are dependent on the available evidence, but may include hip replacements with components made from different materials (metal, ceramic, polyethylene, ceramicised metal); cemented, cementless or hybrid prostheses; prostheses with differing femoral head sizes; prostheses with differing revision rates	Different types of primary THR and hip RS will be compared for people in whom both procedures are suitable Different types of primary THR will be compared for people in whom hip RS is not suitable
Outcomes	The outcome measures to be considered include functional result, pain, bone conservation, revision rates, radiosteriometric analysis to assess prosthesis movement, dislocation rates, adverse effects of treatment (peri- and postprocedural) including degradation products when appropriate, health-related quality of life and mortality	Outcome measures considered include function, pain, bone conservation, revision rates (device failure/revision rates/time to revision), radiosteriometric analysis (to assess prosthesis movement), radiological results, dislocation rates, health-related quality of life and mortality Adverse events include peri- and postprocedural complications (e.g. infection, nerve palsy, dislocation rates, femoral neck fracture, metallosis, muscle weakness) and metal and other degradation products
Economic analysis	The reference case stipulates that the cost-effectiveness of treatments should be expressed in terms of incremental cost per QALY. The reference case stipulates that the time horizon for estimating clinical effectiveness and cost-effectiveness should be sufficiently long to reflect any differences in costs or outcomes between the technologies being compared. Costs will be considered from NHS and Personal Social Services perspectives	Cost-effectiveness outcomes include mean difference in costs and clinical effectiveness measures or utility measures, ICERs, uncertainty measures, ceiling WTP ratios and probabilities from CEAC
Different types of THR to be considered	If the evidence allows, subgroups based on activity levels will be compared. Guidance will be issued in accordance with CE marking only. If the recommendations remain based on long-term performance (revision rates, for example ODEP ratings), the collection and monitoring of performance data and arrangements for the effective implementation of such recommendations should be considered	With components made from different materials (metal, ceramic, polyethylene, ceramicised metal); cemented, cementless or hybrid prostheses; prostheses with differing femoral head sizes

CE, Conformité Européene; CEAC, cost-effectiveness acceptibility curve; ICER, incremental cost-effectiveness ratio; WTP, willingness to pay.
a See http://guidance.nice.org.uk/TA304 (accessed August 2014).

Chapter 3 Joint registries

Description of the three largest international registries

National joint registries have improved the recording of interventions, patient outcomes, implant survival and different surgical techniques for joint replacement. They aim to collect data on large samples, that is, countrywide to improve the outcome of replacement surgery for patients. Interest in national registries has continued to grow and annual reporting from the registries is important for decision-makers, academia and the various industry professionals. Registries available worldwide include those from the UK, Canada, Australia, New Zealand, Sweden, Italy, Norway and Denmark (among others). We conducted a review of the recent annual reports published from these databases. A summary of the three longest-established joint registries is provided for information (*Table 8* and following sections).

Australian Orthopaedic Association National Joint Replacement Registry

The Australian Orthopaedic Association established the National Joint Replacement Registry (AOANJRR) in 1999. At that time, outcomes of surgery in Australia were unknown. The registry began data collection in South Australia on 1 September 1999 followed by the inclusion of each of the Australian states until 2002.[95] The register was expanded to include other joint replacements in November 2007, with all hospitals undertaking joint replacement in Australia approving participation in the collection of additional data. The number of hip replacements has been steadily increasing since 1999, with > 37,000 hip replacements undertaken in Australia in 2012.[95]

The most recent report from the AOANJRR discussed the large increase in revision hip procedures in Australia.[95] In 2010, revision procedures represented 11.3% of all hip replacements, but by 2011 this had

TABLE 8 The three longest-established joint (hip) replacement registries available worldwide[a]

Name	Country	Year established	Lifetime reporting	Most recent report	Data collected
NJR	England and Wales	2002	10 years	2011, surgical data to 31 December 2010	Reports a large number of process and outcome variables across England and Wales, including operation totals, provider sector and type; patient characteristics and procedure details; implant and operation details; implant survival (88.6%); compliance (85.2%)
Swedish Hip Arthroplasty Register	Sweden	1979	33 years	2010	Reports a large number of outcome variables at unit and aggregate county council levels, including reported health gains (EQ-5D index gain after 1 year); patient satisfaction after 1 year; short-term complications after 2 years; 10-year implant survival (95%); compliance (98.5%)
Australian Orthopaedic Association National Joint Replacement Registry	Australia	1999	13	2012	Reports outcome variables across all states: 10-year implant survival (95%); RS reported to be 1.6% of procedures; compliance (93.9%)

a More than 1000 entries.

increased to 12.5%. The authors associated this increase with the DePuy ASR hip (discontinued metal-on-metal hip replacement) and its reported problems. The use of primary RS had declined by 39.7% between 2010 and 2011, accounting for only 1.6% of all hip procedures. In 2012 a reduction in the use of new hip prostheses and prosthetic combinations was reported. In 2010 there were 330 combinations being used in Australia; this had reduced to 97 in 2011.

The Swedish Hip Arthroplasty Register

The Swedish Hip Arthroplasty Register (SHAR) is entering its 33rd year of activity.[96] National coverage for 2010 was 98.5% and 15,935 primary THRs were performed. The registry collects data on all implant types, surgical techniques and reoperation frequency. Individual patient data (IPD) such as age, sex, diagnosis, surgical technique and type of implant used are recorded and, since 2002, patient-reported outcome measures (PROMs) such as pain relief, satisfaction and health-related quality of life have been included. The response rate for PROMs at the 1-year follow-up is just over 90%.

All units in Sweden (78 hospitals) that carry out total hip arthroplasty, both public and private, are included in the registry. The registry's aim is to identify predictors for both good and poor outcomes.[96] In international comparisons, Sweden has the world's highest reported 10-year implant survival rate for total hip arthroplasties. At county council level there are no large and significant differences that are detectable at unit level. The 10-year survival rate of the most common implants was > 95% in 2010.[96] The 2010 report stated that the potential for improvement lies chiefly among certain patient groups. Sweden reports the lowest frequency of revision worldwide. However, it states that problem areas still exist and that these can be overcome with systematic local analyses and subsequent improvement work.

National Joint Registry for England and Wales

The NJR aims to improve patient safety and clinical outcomes by providing information to patients and to all those involved in the management and delivery of joint replacement surgery. This is achieved by collecting data to monitor the effectiveness of hip, knee and ankle replacement surgery and prosthetic implants.[36]

The NJR was established in October 2002 and began collecting data on hip and knee replacement operations on 1 April 2003. The most recent report[36] was from the period 1 April 2010–31 March 2011 and also included statistics on joint replacement activity and a survivorship analysis of hip replacement surgery using data from 1 April 2003 to 31 December 2010.[36] The NJR is one of the largest registries with over one million recorded procedures and a compliance rate of 85.2% (from 1 April 2003 to 31 March 2010). Compliance has shown a steady upwards trend since 2003.[36]

Quality assessment of the NJR[36] is undertaken as a part of the annual reporting of the NJR process using robust statistical techniques. The following factors are considered: random variation, differences in surgical case mix and factors related to the practice of care. The quality assessment results from 2011 reported:

- data from 1.2 million procedures
- a sophisticated method of classifying implant components
- a patient consent rate of 90.4%
- activity and outcomes data at trust, health board and unit level.

Since 1 April 2009, providers of hip replacement surgery have been required to collect and report PROMs under the terms of the Standard NHS Contract for Acute Services.[36] This means that all providers of NHS-funded surgery are expected to invite patients undergoing this procedure to complete a preoperative PROMs questionnaire in accordance with the relevant guidance. Postoperative questionnaires are then sent to patients following their operation after a specified time period. Data collected in the NJR can be linked to the PROMs data collected by the Health and Social Care Information Centre.[97] The NJR is currently working to extend its own study of the follow-up of PROMs to 12 months. This will allow for investigation of population-level quality-of-life reporting after hip replacement.[36]

Summary of national registries

Joint registries, such as those in the UK and Australia, are 'government' organisations. Some are funded by fees levied on orthopaedic implant manufacturers, with fund disbursement conducted under the discretion of the registry steering committee. Although the costs associated with the development and maintenance of national joint registries vary, registries are considered a beneficial medical development because of their ability to detect poorly performing implants at a national level.

The three national registries summarised here report long-term data and have compliance rates of 83.2% (NJR), 98.5% (SHAR) and 93.9% (AOANJRR). Implant survival rates are reported as 88.6%, 95% and 95% at 9, 10 and 10 years, respectively. In England and Wales the incorporation of new PROMs data is planned, which will allow for linkage between activity and patient outcomes.

Chapter 4 Assessment of evidence

Methods for the review of clinical effectiveness

A protocol was developed and approved by NICE (www.nice.org.uk/nicemedia/live/13690/62831/62831.pdf). General principles were applied as recommended by the NHS Centre for Reviews and Dissemination (CRD).[98]

This report contains reference to confidential information provided as part of the NICE appraisal process. This information has been removed from the report and the results, discussions and conclusions of the report do not include the confidential information. These sections are clearly marked in the report.

Identification of studies

Initial scoping searches were undertaken in MEDLINE in October 2012 to assess the volume and type of literature relating to the assessment question. The scoping searches also informed development of the final search strategies (see *Appendix 1*). An iterative procedure was used to develop these strategies with input from clinical advisors and previous HTA reports (e.g. Vale *et al.*,[19] de Verteuil *et al.*[11]). The strategies have been designed to capture generic terms for arthritis, THR and RS.

Search strategies

Final searches were undertaken in November and December 2012 (see *Appendix 1*) and were date limited from 2002 (the date of the most recent NICE guidance in this area[25]). Searches of the clinical effectiveness literature were restricted to RCTs and systematic reviews; additional searches were undertaken to capture literature relating to costs, resource use, utilities, cost-effectiveness, cost-effectiveness models and registries to inform the survival and cost-effectiveness analysis.

The following main sources were searched to identify relevant published and unpublished studies and studies in progress:

- electronic bibliographic databases
- contact with experts in the field
- references of included studies
- screening of relevant websites.

The following databases of published studies were searched: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, Science Citation Index and Conference Proceedings Citation Index – Science, The Cochrane Library [specifically the Cochrane Database of Systematic Reviews (CDSR), Cochrane Central Register of Controlled Trials (CENTRAL), Database of Abstracts of Reviews of Effects (DARE), NHS Economic Evaluation Database (NHS EED), HTA database], Current Controlled Trials, ClinicalTrials.gov and UK Clinical Research Network (UKCRN) Portfolio Database. The search strategies were initially developed for MEDLINE and were adapted as appropriate for other databases.

The reference lists of included studies and relevant review articles were checked and the following websites of hip implant manufacturers were screened for relevant publications:

- Amplitude
- Biomet
- B Braun/Aesculap
- Comis Orthopaedics
- Corin
- DePuy

- Exactech
- Finsbury
- JRI Orthopaedics
- Implantcast
- Implants International
- Lima WG Healthcare
- Mathys Orthopaedics
- Medacta UK
- Orthodynamics
- Peter Brehm
- SERF Dedienne santé
- Smith & Nephew
- Stanmore Implants Worldwide
- Stryker
- Symbios SA
- Waldemar Link
- Wright Medical UK
- Zimmer, Inc.

Grey literature searches were undertaken using Google (Google Inc., Mountain view, CA, USA) and the online resources of the following regulatory bodies, health services, research agencies and professional societies:

- British Hip Society
- British Orthopaedic Association
- Orthopaedic Research UK
- ODEP
- NJR
- Arthritis Research UK
- Cochrane Musculoskeletal Group
- Arthritis Care
- MHRA
- American Association of Hip and Knee Surgeons
- American Academy of Orthopedic Surgeons (AAOS)
- The Hip Society
- Royal College of Surgeons
- Royal College of Surgeons of Edinburgh.

All bibliographic records identified through the electronic searches were collected in a managed reference database.

Inclusion criteria

Study design

- RCTs.
- Systematic reviews.
- Meta-analyses.

Given the wide scope and large amount of identified evidence, we limited studies to those published since 2008 with a sample size of $\geq$ 100 participants.

Population

- People with pain or disability resulting from end-stage arthritis of the hip for whom non-surgical management has failed.

Intervention

- Elective primary THR.
- Primary hip RS.

Comparator

- Different types of primary THR compared with RS for people in whom both procedures are suitable.
- Different types of primary THR compared with each other for people who are not suitable for hip RS.

Outcomes

Clinical effectiveness outcome measures were mortality, validated functional/pain and health-related quality of life total scores, revision rate, implant survival rate and femoral head penetration rate (measure of prosthesis movement). Adverse events included incidence of peri-/postprocedural complications (i.e. implant dislocation, infection, osteolysis, aseptic loosening, femoral fracture and deep-vein thrombosis).

Exclusion criteria

The exclusion criteria were as follows:

- indications for hip replacement other than end-stage arthritis of the hip
- revision surgery as the primary procedure of interest
- abstract/conference proceedings, letters and commentaries
- non-English language publications.

Study selection process

All retrieved records were collected in a specialised database. All duplicate records were identified and removed. Two reviewers pilot tested an a priori screening form based on the predefined study eligibility criteria. Afterwards, two independent reviewers applied the same inclusion/exclusion criteria and screened all identified bibliographic records for title/abstract (level I) and then for full text (level II). Disagreements over eligibility were resolved through consensus or by a third party reviewer. Reasons for exclusion of full-text papers were documented. The study flow was documented using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram.[99]

Quality assessment strategy

Two reviewers independently assessed the risk of bias of individual studies using validated tools I (see *Appendix 2*).[100,101] Any disagreements between the two reviewers were resolved by a third reviewer through discussion.

Randomised controlled trials were assessed using the Cochrane Collaboration risk of bias tool,[100] which covers the following domains of threat to internal validity: selection bias (randomisation sequence generation, treatment allocation concealment), performance bias (blinding of participants/personnel), detection bias (blinding of outcome assessors), attrition bias (incomplete outcome data), reporting bias (selective outcome/analysis reporting) and other prespecified bias [e.g. funding source, adequacy of statistical methods used, type of analysis (intention to treat/per protocol), imbalance in the distribution of baseline prognostic factors between the compared treatment groups]. The risk of bias assessment results fall into three distinct categories of high, low and unclear risk of bias. For each RCT, the risk of bias for the performance, detection and attrition bias domains was assessed for a priori defined groups of subjective (e.g. patient-administered clinical and functional scores) and objective (e.g. mortality, revision,

survival, radiography result, complications) outcomes separately. Afterwards, the within-study summary risk-of-bias rating across all of the domains was derived for subjective and objective outcomes separately. The decision for determining the within-study summary risk of bias was based on the ratings prevailing for the selection, performance and detection bias domains. At data synthesis stage, the across-study average summary risk of bias was determined and assigned to each outcome of interest.

The methodological quality of included systematic reviews was assessed with the Assessment of Multiple Systematic Reviews (AMSTAR) tool,[101] which covers the following domains: (1) research question, (2) inclusion/exclusion criteria, (3) search strategy (at least two major electronic databases), (4) data extraction by independent reviewers, (5) assessment of risk of bias by independent reviewers, (6) consideration of risk of bias in the analysis, (7) exploration of heterogeneity and (8) publication bias. For convenience of presentation, the methodological quality of each systematic review was graded according to the number of items satisfied as follows: high (range 9–11), medium (range 5–8) and low (range 0–4).

Grading the overall quality of clinical effectiveness evidence

The overall quality of evidence for each preselected (i.e. gradable) outcome across studies was assessed using the systematic approach developed by the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) Working Group (see www.gradeworkinggroup.org).

The GRADE approach[102] indicates levels of confidence in the observed treatment effect estimate(s), which are categorised as high, moderate, low or very low. The grading of overall quality of evidence for each gradable outcome is based on assessments across five domains: (1) summary risk of bias across studies per gradable outcome (internal validity across studies, study limitations), (2) consistency of results (heterogeneity), (3) directness of the evidence (applicability of the results, indirect treatment comparisons), (4) precision of the results (the width of the 95% CI around the estimate) and (5) publication/reporting bias (detection of asymmetry in the funnel plot, selective outcome reporting). The definitions and explanations of the grading levels and the grading process across the five domains are presented later in this chapter (see *Tables 35* and *43*).

The gradable outcomes, selected according to their meaningfulness and importance for decision-making, were the following: HHS, WOMAC score, revision, mortality, femoral head penetration rate and implant dislocation.

Data extraction strategy

The relevant data were extracted from included studies independently by one reviewer using a data extraction form informed by the CRD.[103] The extracted data were cross-checked by a second reviewer. Uncertainty and/or any disagreements with the second researcher were resolved by discussion. The extracted data were entered into summary and full extraction tables (see *Appendices 3* and *4*, respectively). The extracted information included the following:

- Study characteristics (i.e. authors, country, design, study setting, sample size, funding source, duration of follow-up and information relevant to risk-of-bias assessment such as generation of randomisation, allocation concealment, blinding, completeness of outcome ascertainment, patient withdrawals/attrition for randomised trials; for observational studies and non-randomised trials, and information on potential confounding was additionally ascertained).
- Patient baseline characteristics [i.e. inclusion/exclusion criteria, number of enrolled/analysed participants, age, race, sex, body mass index (BMI), underlying conditions, concomitant conditions, co-interventions, disability, activity levels, function, pain intensity and quality of life and disease-specific measures such as the OHS[30] and HHS[31]].
- Experimental treatment characteristics (e.g. type – THR, RS; training/experience of the operator and postoperative rehabilitation staff; method of fixation – cemented, cementless, hybrid; bearing surface material – metal-on-metal, ceramic-on-ceramic, polyethylene-on-metal; femoral head size; name/brand and country of manufacturer; postoperative rehabilitation).

- Outcome characteristics [e.g. definition; timing of measurement; scale of measurement – dichotomous, continuous; measures of association – mean difference (MD), risk ratio (RR), odds ratio (OR), hazard ratio (HR)]. Statistical test results and measures of variability were also extracted [standard deviation (SD), 95% CI, standard error (SE), *p*-value).

Any additional relevant information found in multiple publications of included studies was also extracted. For studies of clinical effectiveness in which summary measures and 95% CIs for the association between the treatments were not reported, MDs with 95% CIs were calculated if data allowed (*t*-tests for independent samples and continuous outcomes and RRs for dichotomous outcomes). No RRs and 95% CIs were estimated for individual studies that observed zero events in one or both treatment arms. The 95% CIs and SEs were used to derive SDs or vice versa. All calculated parameters were entered into the data extraction sheets.

Data management

Study, treatment, population and outcome characteristics were summarised in text, evidence and summary tables. The study results were compared qualitatively and quantitatively in text and summary tables. For each outcome of interest, the effectiveness of treatments reported in individual studies was compared as follows:

- different types of primary THR compared with each other for people who are not suitable for hip RS
- different types of primary THR compared with RS for people in whom both procedures are suitable.

Meta-analysis

The decision to pool individual study results was based on a degree of similarity with respect to methodological and clinical characteristics of studies under consideration (e.g. design, population, comparator treatment and outcome). Estimates of post-treatment MDs for continuous outcomes and RRs for binary outcomes (except for rare events) of individual studies were pooled using a DerSimonian and Laird random-effects model.[104] The choice of this model was based on the assumption that some residual clinical and methodological diversity will exist across pooled studies. Dichotomous outcomes with low event rates (5.0–10.0%) were pooled as RRs using a Mantel–Haenszel fixed-effects model. Dichotomous outcomes for studies with very low event rates ($\leq$ 5.0%) or zero events in one of the treatment arms were pooled as ORs using a Peto fixed-effects model.[105]

Trials were not pooled if the mean and/or SD for the continuous outcome of interest could not be ascertained.

The degree of statistical heterogeneity across pooled studies was determined through inspection of the forest plots, Cochran's *Q* and the I^2 statistic. The presence of heterogeneity was judged according to predetermined levels of statistical significance (chi-squared $p < 0.10$ and/or $I^2 > 50\%$). Statistical pooling was performed using The Cochrane Collaboration software package Review Manager version 5.2 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark).

Publication bias

It was planned to examine the extent of publication bias, given a sufficient number of data points, by visual inspection of funnel plots with respect to plot asymmetry as well as using linear regression tests.[106]

Analysis to explore heterogeneity

If data allowed, exploration of study-level clinical and methodological sources of statistical heterogeneity of effect estimates across studies was planned through a priori-defined subgroup analysis (i.e. age, sex, function), sensitivity analysis (risk of bias item-specific ratings, intention-to-treat vs. per-protocol analysis) and meta-regression.

Data synthesis and interpretation

For both RCTs and systematic reviews, the comparison and synthesis of results for each outcome of interest was summarised and categorised as conclusive evidence (either there is a 'difference' or there is 'no difference') or inconclusive evidence (indeterminate results because of statistical uncertainty, statistical heterogeneity/inconsistency in treatment effects and/or incomplete information). This conclusion was based on several factors determined separately or in combination such as statistical significance of the observed difference (*p*-value), magnitude of the effect estimate, width of the 95% CIs, a minimal clinically important difference (MCID) for a given outcome, if known, and consistency in terms of effect direction and statistical significance. We ascertained the MCIDs for clinical/functional measures such as HHS (MCID range 7–10), OHS (MCID range 5–7), WOMAC score (MCID 8) and EQ-5D score (MCID 0.074) from previous empirical research evidence.[107–109]

Evidence was considered conclusive in showing a 'difference' if a treatment effect estimate was statistically significant and the 95% CI included the MCID for any given outcome. Evidence was considered conclusive in showing 'no difference' if a treatment effect estimate was not statistically significant and the 95% CI around it was narrow enough to exclude the MCID for any given outcome. Alternatively, evidence was considered conclusive in showing 'no difference' if a treatment effect estimate was statistically significant but the 95% CI around it did not include the MCID for an outcome.

Evidence was considered inconclusive if a treatment effect estimate was not statistically significant and had 95% CIs that were sufficiently wide to include the MCID or any large effect size values. (Because for such studies the possibility of type II error cannot be ruled out, the observed non-significant results should not be interpreted as if there is no difference between the treatment effects. The lack of precision around the effect estimates may be a result of an insufficient sample size, a short follow-up period and/or low event counts, leading to inadequate study power and an increased chance of a type II error.)

The results were also considered inconclusive if there were partially missing data for continuous outcomes (e.g. reporting treatment arm-specific means without SDs; reporting only *p*-values for the between-treatment difference) or zero events for binary outcomes in both treatment arms. Evidence from studies showing inconsistent results, that is, significant effects but in opposing directions, was also classified as inconclusive.

Evidence from systematic reviews not reporting pooled results of RCTs (i.e. reporting only narrative syntheses), those reporting inappropriate pooling methods (e.g. indirect naive comparison of single group cohorts; pooling of studies of different design) or those reporting inconsistent summary findings was also considered inconclusive.

Industry submissions regarding effectiveness of treatments

The included clinical effectiveness evidence was compared with the evidence submitted by industry. These industry submissions will be discussed in *Appendix 5*.

Results of the review of clinical effectiveness

Search results

A total of 2469 records were identified through our searches of different sources. The removal of duplicates left 1522 records to be screened. Of these, 1281 records were excluded as irrelevant at title and abstract screening, leaving 241 potentially relevant records. Of these 241 full-text records screened, 146 were excluded, leaving 95 potentially relevant full-text records, of which 58 were additionally excluded based on publication date (published before 2008 unless a companion paper to an included study) and sample size (< 100 participants). The remaining 37 records were included in the review.[107,110–145]

The flow chart outlining the process of identifying relevant literature can be found in *Figure 9*.

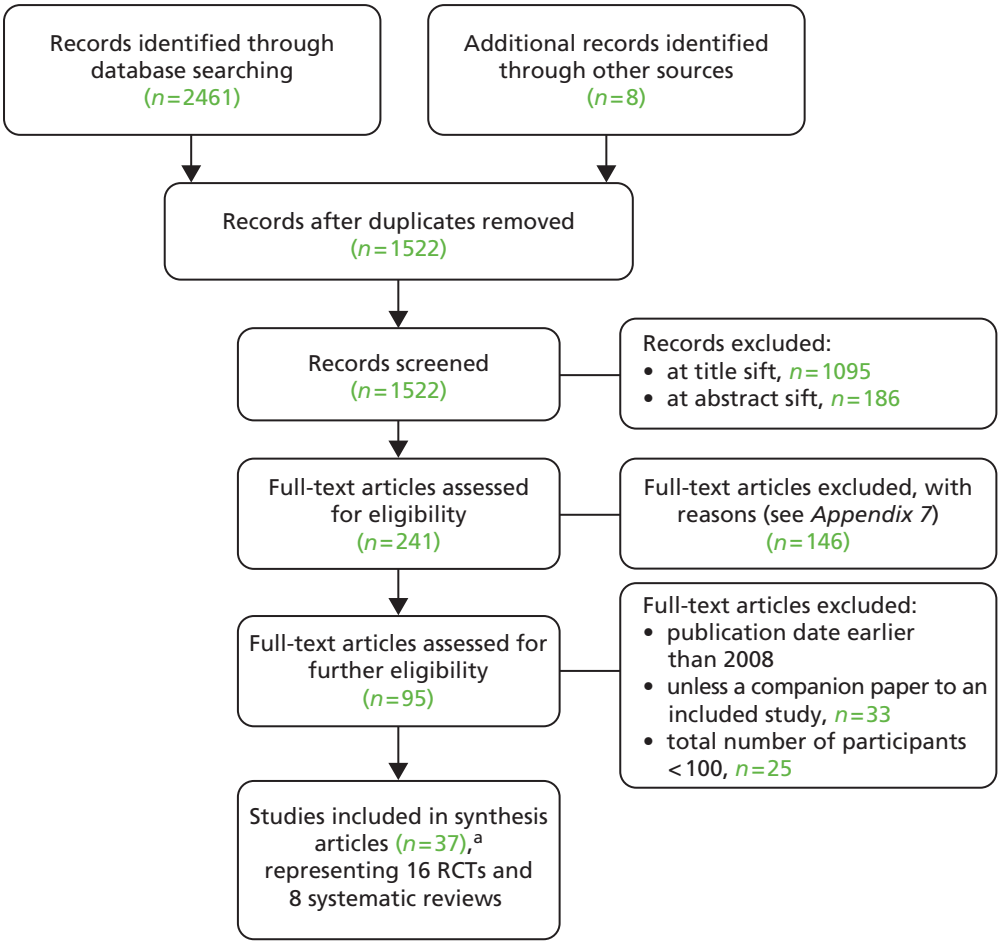


FIGURE 9 Flow diagram of study identification for the clinical effectiveness review. a, A further 20 ongoing clinical trials were identified.

A list of records excluded at full-text screening with reasons for exclusion is provided in *Appendix 6*. The main reasons for exclusion were the comparison of different surgical/operative approaches ($n = 42$[11,146–186]), study published before 2008 (unless a companion paper to an included study) ($n = 33$[19,39,187–217]) and study includes < 100 participants ($n = 25$[83,218–241]).

A separate search in December 2012 of Clinical Trials.gov, Current Controlled Trials, the UKCRN Portfolio Database and the National Library of Medicine (NLM) Gateway Health Services Research Projects in Progress (HSRProj) database retrieved 511 potential trials or health services research projects. After screening titles and full records (if available), 20 clinical trials and one health services research project were identified, one of which[130] had already been identified from the original database search (see *Appendix 7*). The identified clinical trials were considered potentially relevant based on the available information. The trials were ongoing or completed since 2009 or their status was unknown.

The included 37 records represent 16 RCTs[107,110–136,145] and eight systematic reviews.[137–144]

Six of the 16 RCTs were represented by multiple publications:

1. Bjørgul *et al.*[110,111]
2. Engh *et al.*[113,114]
3. Capello *et al.*,[115] D'Antonio *et al.*[116,117] and Mesko *et al.*[118]
4. Corten *et al.*,[119,122] Laupacis 2002[120] and Bourne and Corten[121]
5. Costa *et al.*[130] and Achten *et al.*[107]
6. Vendittoli *et al.*,[132,133,136] Girard *et al.*[134] and Rama *et al.*[135]

These six RCTs are cited as follows: Bjørgul *et al.*[110] Engh *et al.*,[113] Capello *et al.*,[115] Corten *et al.*,[119] Costa *et al.*[130] and Vendittoli *et al.*[132] Thirteen RCTs[110,112,113,115,119,123–129,145] and five systematic reviews [137–141] comparing different types of primary THR and three RCTs[130–132] and three systematic reviews[142–144] comparing primary THR with RS were finally included in the current review.

In the following sections we will begin by reporting the findings for the comparison of different types of THR and will then report the findings for the comparison between THR and RS.

Comparison of different types of total hip replacement

Study and participant characteristics

Randomised controlled trials

The study and participant characteristics of the 13 included RCTs[110,112,113,115,119,123–129,145] are summarised in *Table 9*. More details can be found in *Appendices 3* and *4*. Briefly, four RCTs were conducted in the USA,[113,115,125,127] one in the UK,[112] one in Australia,[123] two in Norway[110,126] two in the Republic of Korea[128,129] and three in Canada.[111,119,124] A total of 3175 participants were randomised across the 13 RCTs, with the number of participants in each study ranging from 100[124,128,145] to 557.[123] The mean age of participants across the RCTs ranged from 45[129] to 72[123,145] years. The proportion of women across the studies ranged from 24%[129] to 73%.[110] The length of follow-up of the studies ranged from 3 months[119] to 20 years.[119,129] The proportion of participants diagnosed with primary OA was reported for nine studies[110,112,113,115,123,124,127–129] and ranged from 14%[129] to 96%.[123]

Comparison of THR interventions in the included RCTs was based on differences in hip replacement implant components (e.g. acetabular cup/shell, femoral stem and femoral head) according to their composition,[127] design,[115,128] bearing surface,[113,115,124–126,145] fixation method[110,112,119,129] and component size.[123] *Table 10* shows the distribution of RCTs across the THR comparison categories.

Reported outcomes across the 13 RCTs varied. Most RCTs reported HHS[110,112,113,115,119,124–129,145] and risk of revision.[112,113,115,119,123–125,127–129] The follow-up of outcome assessments ranged from 3 months[119] to 20 years.[119,129] Outcomes reported in the included studies can be found in *Appendix 8*. A summary of the functional/clinical and quality of life measures/tools used is provided in *Appendix 9*.

Systematic reviews

The five included systematic reviews[137–141] evaluated RCTs and non-RCTs of the clinical effectiveness of THR (see *Appendix 3*). The primary focus of these systematic reviews was the comparison of the effects of different cup fixation methods (cemented vs. cementless)[137–139] and materials used for implant articulations[140,141] on postoperative clinical/functional scores (HHS, OHS)[137,138,140] and risk of revision rate.[138,139] Searches in these systematic reviews were undertaken between July 2007[141] and June 2011.[139] Further details on specific outcomes reported in the included systematic reviews can be found in *Appendix 8*.

TABLE 9 Overall study characteristics across the 13 RCTs comparing different types of THR

Study characteristic	Metric
Geographical region	UK ($n = 1$); Australia ($n = 1$); Norway ($n = 2$); the Republic of Korea ($n = 2$); Canada ($n = 3$); USA (n=4)
Total number of randomised participants	3175 (range 100–557)
Mean age (years)	Range 45–72
Female participants (%)	Range 24–73
Length of follow-up	Range 3 months–20 years
Diagnosis of primary OA (%)	Range 14–96

TABLE 10 Distribution of 13 RCTs according to basis of THR comparison

Basis of comparison	Study
1. Cup fixation	Bjørgul 2010[110]
	Angadi 2012[112]
2. Cup liner bearing surface	McCalden 2009[145]
	Engh 2012[113]
3. Cup shell design	Capello 2008[115]
4. Cup/stem fixation	Corten 2011[119]
5. Femoral head size	Howie 2012[123]
6. Femoral head bearing	Lewis 2008[124]
7. Femoral head-on-cup liner bearing surface	Amanatullah 2011[125]
	Capello 2008[115]
	Kadar 2011[126]
8. Stem composition	Healy 2009[127]
9. Stem design	Kim 2011[128]
10. Stem fixation	Kim 2011[129]

Risk of bias and methodological quality

Risk of bias in the randomised controlled trials

The risk-of-bias assessments for the 13 included RCTs comparing different types of THR are presented in risk-of-bias tables (see *Appendix 2*), the summary table (*Table 11*) and the risk-of-bias graph (*Figure 10*). Overall, four[112,119,123,128] of the 13 RCTs reported an adequate method for random sequence generation and eight[110,112,119,123–126,129] reported adequate treatment allocation concealment (low risk of bias). A greater proportion of the RCTs were rated as having a low risk of performance and detection bias for objective (e.g. mortality, dislocation) than for subjective (e.g. patient-administered functional scores) outcomes (92–100% vs. 15–23%, respectively). For at least eight of the RCTs, it was unclear whether or not awareness of THR type would influence the ascertainment of clinical/functional scores by patients/study personnel (performance bias)[110,112,113,115,124,125,127–129,145] or outcome assessors (detection bias).[112,113,115,124–126,128,129] Most RCTs failed to report the blinding status of the patients, study personnel and/or outcome assessors. Eight RCTs were judged as having a low risk of attrition bias. Five RCTs[115,124,125,127,128] were judged as being at high risk for selective outcome and/or analysis bias. The risk of other biases (e.g. funding source, baseline imbalance in important characteristics, inappropriate analysis) for about one-third of the RCTs was judged to be high.

Methodological quality of the systematic reviews

The assessment of methodological quality of the five included systematic reviews comparing different types of THR is presented in *Table 12* and the quality assessment sheets (see *Appendix 2*). Briefly, based on the number of methodological items that were satisfied, two systematic reviews[137,140] were judged to be of high quality (falling into the score range of 9–11) and two systematic reviews[138,141] were of medium quality (falling into the score range of 5–8). The one remaining systematic review[139] was judged to be of low quality (falling into the score range of 0–4). The specific unmet methodological items related to inappropriate analysis, absence of duplicate study selection, limited literature search, failure to address issues of publication bias and no information on conflicts of interest.

TABLE 11 Risk of bias summary for RCTs: review authors' judgements about each risk of bias item – THR vs. THR

Study	Selection bias: random sequence generation	Selection bias: allocation concealment	Performance bias: subjective (e.g. patient reported)	Performance bias: objective (e.g. mortality, radiography, dislocation)	Detection bias: subjective (e.g. patient reported)	Detection bias: objective (e.g. mortality, radiography, dislocation)	Attrition bias: subjective (e.g. patient reported)	Attrition bias: objective (e.g. mortality, radiography, dislocation)	Reporting bias: selective reporting of the outcome, subgroups or analysis	Other bias [funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics]
Amanatullah 2011[125]	?	+	?	+	?	+	?	–	–	–
Angadi 2012[112]	+	+	?	+	?	+	+	+	+	?
Bjørgul 2010[110]	?	+	?	+	+	+	–	–	+	–
Capello 2008[115]	?	?	?	+	?	+	+	+	–	–
Corten 2011[119]	+	+	+	+	+	+	+	+	+	+
Engh 2012[113]	?	?	?	?	?	+	–	–	+	?
Healy 2009[127]	–	–	?	+	–	+	+	+	–	+
Howie 2012[123]	+	+	NA	+	NA	+	NA	+	+	+
Kadar 2011[126]	?	+	+	+	+	+	+	+	+	+
Kim 2011[128]	+	?	?	+	?	+	+	+	–	+
Kim 2011[129]	?	+	?	+	?	+	+	+	+	?
Lewis 2008[124]	?	+	?	+	?	+	?	?	–	?
McCalden 2009[145]	?	?	?	+	+	+	+	+	+	–

–, high risk of bias; +, low risk of bias; ?, unclear risk of bias; ITT, intention to treat; NA, not applicable; PP, per protocol.

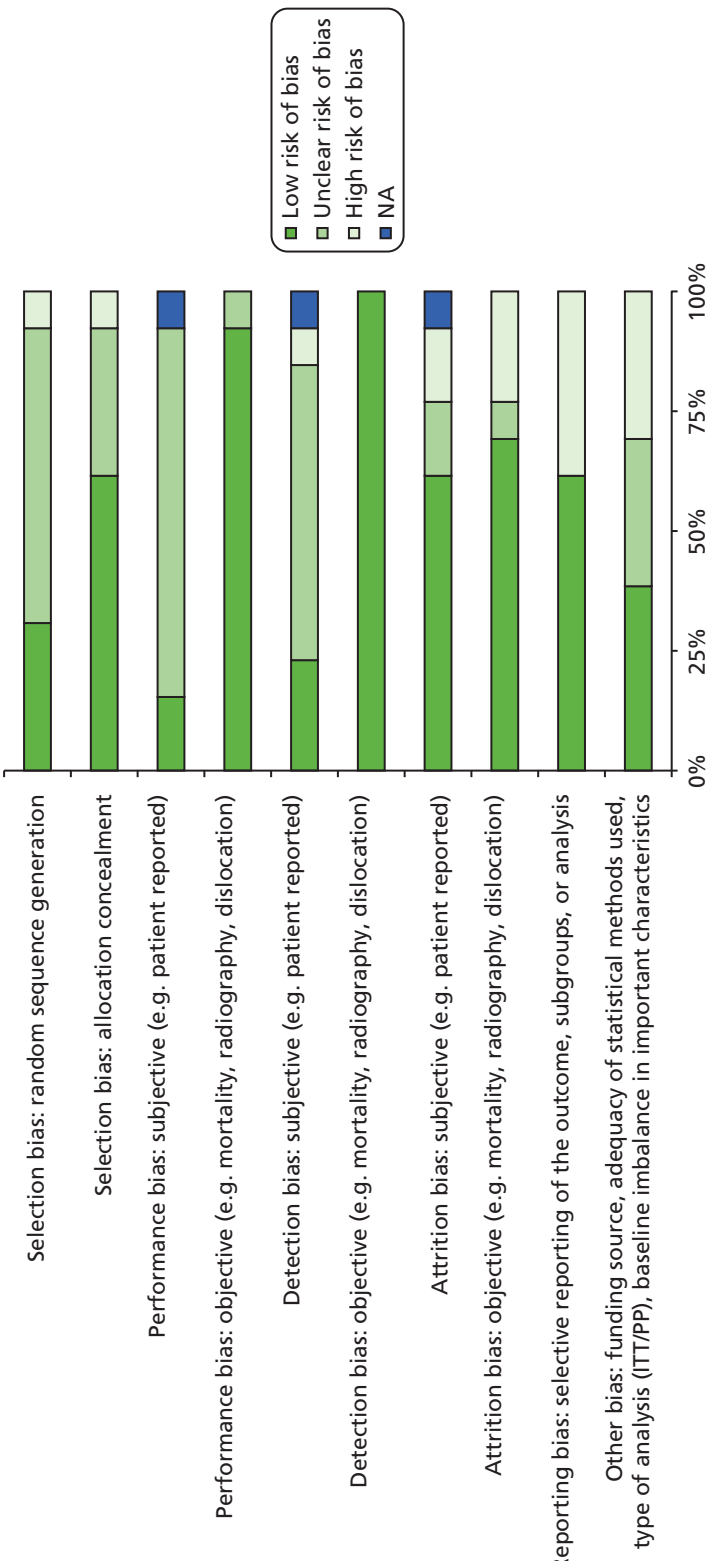


FIGURE 10 Risk of bias graph for RCTs: review authors' judgements about each risk of bias item – THR vs. THR. ITT, intention to treat; NA, not applicable; PP, per protocol.

TABLE 12 Methodological quality assessment summary for systematic reviews: THR vs. THR

Study	Was an a priori design provided?	Was there duplicate study selection and data extraction?	Was a comprehensive literature search performed?	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	Was a list of studies (included and excluded) provided?	Were the characteristics of the included studies provided?	Was the scientific quality of the included studies assessed and documented?	Was the scientific quality of the included studies used appropriately in formulating conclusions?	Were the methods used to combine the findings of studies appropriate?	Was the likelihood of publication bias assessed?	Was the conflict of interest stated?	Overall
Clement 2012[139]	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	Low quality
Pakvis 2011[138]	Yes	No	Yes	Yes	No	Yes	Yes	No	No	No	No	Medium quality
Sedrakyan 2011[140]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	High quality
Voigt 2012[137]	Yes	Yes	Yes	CA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	High quality
Yoshitomi 2009[141]	Yes	Yes	Yes	CA	Yes	Yes	Yes	NA	No	Yes	No	Medium quality

CA, cannot answer; NA, not applicable.

Clinical effectiveness findings for the comparison of different types of total hip replacement

This section summarises the evidence from the 13 RCTs[110,112,113,115,119,123–129,145] and five systematic reviews.[137–141]

The reported outcomes for this section were HHS (12 RCTs;[110,112,113,115,119,124–129,145] three systematic reviews[137,138,140]), WOMAC score (four RCTs[119,124,129,145]), McMaster Toronto Arthritis Patient Preference Questionnaire (MACTAR) score (one RCT[119]), Merle d'Aubigné and Postel hip score (one RCT[119]), University of California Los Angeles (UCLA) activity score (one RCT[129]), OHS (one systematic review[137]), SF-12 score (three RCTs;[124,125,145] one systematic review[140]), risk of revision (10 RCTs;[112,113,115,119,123–125,127–129] five systematic reviews[137–141]), mortality (six RCTs[110,113,119,123,128,145]), femoral head penetration rate (three RCTs[113,126,145]), implant dislocation (seven RCTs;[110,112,115,123–125,127] two systematic reviews[139,140]), osteolysis (seven RCTs;[112,113,115,125,127,129,145] two systematic reviews[138,139]), aseptic loosening (five RCTs;[112,113,119,124,127] one systematic review[139]), femoral fracture (three RCTs[113,115,127]), infection (four RCTs[112,124,125,127]) and deep-vein thrombosis (one RCT[125]).

Neither the RCTs nor the systematic reviews reported any evidence for the following clinical effectiveness outcomes:

- HOOS
- LISOH
- AAOS Hip and Knee Questionnaire
- Arthritis Impact Measurement Scale (AIMS)
- Nottingham Health Profile (NHP) questionnaire
- EQ-5D
- SF-36
- time to revision
- pain score [visual analogue scale (VAS)].

Summary results for the following outcomes are presented separately for RCTs and systematic reviews in the following sections. The outcomes of interest are as follows:

- mortality
- validated functional/pain (total scores): HHS, OHS, pain score (VAS), Merle d'Aubigné and Postel score, UCLA activity score, WOMAC, MACTAR, HOOS, LISOH, AAOS Hip and Knee Questionnaire, AIMS
- health-related quality of life (total scores): EQ-5D, SF-36/SF-12, NHP
- revision rate (risk of revision, mean time to revision)
- femoral head penetration rate (measure of prosthesis movement)
- adverse events (peri-/postprocedural complications): implant dislocation, infection, osteolysis, aseptic loosening, femoral fracture, deep-vein thrombosis, muscle weakness, nerve palsy and pulmonary embolism.

Functional/clinical measures

Twelve of the 13 included RCTs comparing different types of THR reported at least some results for the following functional scores measured at different postprocedure follow-up times: HHS (12 studies[110,112,113,115,119,124–129,145]) WOMAC score (four studies[119,124,129,145]), MACTAR score (one study[119]), Merle d'Aubigné and Postel score (one study[119]) and UCLA activity score (one study[129]). None of these 12 studies reported measurements of the OHS.

Three of the five included systematic reviews comparing different types of THR reported at least some evidence on HHS[137,138,140] and OHS.[137] None of the three reviews reported any summary evidence for WOMAC, MACTAR, Merle d'Aubigné and Postel, and UCLA scores.

Harris Hip Score

Randomised controlled trials ($n = 12$) Mean HHS at follow-up (range 6 months–10 years) did not differ between the following interventions: cup fixation (two studies;[110,112] cemented vs. cementless), cup liner bearing surface (two studies;[113,145] cross-linked polyethylene vs. non-cross-linked polyethylene), cup and femoral stem fixation (one study;[119] cemented vs. cementless) and femoral head-on-cup liner bearing surfaces (one study;[126] cobalt–chromium/oxinium-on-polyethylene vs. cobalt–chromium/oxinium-on-cross-linked polyethylene) (*Table 13*). The pooled MD for HHS in our meta-analysis of two studies[113,145] comparing cup liners made with cross-linked polyethylene compared with non-cross-linked polyethylene was 2.29 (95% CI –0.88 to 5.45), suggesting a non-significant benefit of cross-linked polyethylene cup liners (*Figure 11*).

TABLE 13 Harris Hip Score (range 0–100): RCTs

Follow-up	Arm-specific estimate, mean (SD or 95% CI)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
6 months	90.2 (87.9 to 92.6) vs. 89.1 (86.9 to 91.3)[110]	$p > 0.05$ (NS)	2 (unclear)	No difference
2 years	92.7 (89.6 to 95.8) vs. 94.0 (92.4 to 95.7)[110]	$p > 0.05$ (NS)		
5 years	93.9 (91.6 to 96.2) vs. 91.4 (89.3 to 93.5)[110]	$p > 0.05$ (NS)		
10 years	89.8 (87.0 to 92.6) vs. 87.3 (84.1 to 90.6)[110]	$p > 0.05$ (NS)		
10 years	74.5 (NR) vs. 78.0 (NR)[112]	$p > 0.05$ (NS)		
Cup liner bearing surface: XLPE vs. non-XLPE				
1 year	85.0 (10.3) vs. 83.4 (13.1)[145]	MD 1.60 (–3.07 to 6.27)[c]	2 (unclear)	No difference
5 years	86.0 (13.1) vs. 83.1 (15.4)[145]	MD 2.90 (–2.77 to 8.57)[c]		
10 years	88.0 (14.0) vs. 86.0 (15.0)[113]	MD 2.00 (–1.85 to 5.85)[c]		
		Pooled estimate of MD[c] 2.29 (–0.88 to 5.45)[113,145]		
Cup shell design: porous-coated shell vs. arc-deposited HA-coated shell				
5 years	97.0 (NR) vs. 96.4 (NR)[115]	$p > 0.05$ (NS)	1 (unclear)	Inconclusive
10 years	96.0 (NR) vs. 96.7 (NR)[115]	$p > 0.05$ (NS)		
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
3 months	41 (12.0) vs. 41 (11.0)[119]	MD 0.0 (–3.00 to 3.00)[c]	1 (low)	No difference
6 months	47 (12) vs. 50 (13)[119]	MD –3.0 (–6.32 to 0.32)[c]		
1 year	52 (10.0) vs. 53 (11.0)[119]	MD –1.0 (–3.86 to 1.86)[c]		
3 years	50 (14.0) vs. 52 (11.0)[119]	MD –2.0 (–5.62 to 1.62)[c]		
5 years	47 (14.0) vs. 48 (13.0)[119]	MD –1.0 (–4.88 to 2.87)[c]		
7 years	44 (15) vs. 46 (14)[119]	MD –2.0 (–7.07 to 3.05)[c]		
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	92 (NR) vs. 92.5 (NR)[124]	$p > 0.159$ (NS)	1 (unclear)	Inconclusive

TABLE 13 Harris Hip Score (range 0–100): RCTs (*continued*)

Follow-up	Arm-specific estimate, mean (SD or 95% CI)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. metal-on-PE				
5 years	96.4 (NR) vs. 97.0 (NR)[115]	$p > 0.05$ (NS)	1 (unclear)	Inconclusive
10 years	96.7 (NR) vs. 96.4 (NR)[115]	$p > 0.05$ (NS)		
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	NR[125]	$p > 0.05$ (NS)	1 (unclear)	Inconclusive
Femoral head-on-cup liner bearing surface: steel-on-PE vs. CoCr-on-PE vs. oxinium-on-PE vs. CoCr-on-XLPE vs. oxinium-on-XLPE				
2 years	91 (10.8) vs. 91 (8.5) vs. 91 (11.1) vs. 93 (11.3) vs. 88 (9.5)[126]	$p = 0.7$ (NS); ANOVA-based $p = 0.5$ (NS)[c]	1 (low)	No difference
Femoral stem composition: CoCr vs. titanium				
5 years	83 (NR) vs. 87 (NR)[127]	$p = 0.029$ (SS)	1 (high)	Inconclusive
Femoral stem design: short metaphyseal-fitting stem vs. conventional metaphyseal- and diaphyseal-fitting stem				
3 years	97.0 (NR) vs. 96.0 (NR)[128]	$p = 0.79$ (NS)	1 (unclear)	Inconclusive
Femoral stem fixation: cemented vs. cementless				
18 years	91 (NR) vs. 90 (NR)[129]	$p = 0.71$ (NS)	1 (unclear)	Inconclusive

ANOVA, analysis of variance; CoCr, cobalt–chromium; HA, hydroxyapatite; NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias; SS, statistically significant; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported; one RCT did not report any evidence for this outcome.

The evidence for the other comparisons based on cup shell design (porous coated vs. arc-deposited hydroxyapatite coated),[115] femoral head bearing surface (oxinium vs. cobalt–chromium),[124] femoral head-on-cup liner bearing surfaces (ceramic-on-ceramic vs. metal-on-polyethylene or ceramic-on-polyethylene),[115,125] femoral stem composition (cobalt–chromium vs. titanium),[127] femoral stem design (short metaphyseal fitting vs. conventional diaphyseal fitting)[128] and femoral stem fixation (cemented vs. cementless)[129] was judged to be inconclusive.

Systematic reviews (n = 3) One systematic review[140] reported the pooled MD for the HHS (*Table 14*). Pooled estimates for the comparison between metal-on-metal and metal-on-polyethylene bearing surfaces at two different follow-up times were not consistent: at 2 years metal-on-metal bearing surfaces gave a significantly higher HHS than metal-on-polyethylene, but at > 2 years there was no significant difference between the two types of THR. The remaining two systematic reviews presented only narrative summaries.[137,138] In summary, for the HHS the systematic review-based evidence was judged to be inconclusive.

Western Ontario and McMaster University Osteoarthritis Index score

RCTs (n = 4) Results from all four RCTs reporting postprocedural mean WOMAC scores indicated statistically non-significant differences between the THR groups compared with respect to cup liner bearing surface (cross-linked polyethylene vs. non-cross-linked polyethylene),[145] cup and femoral stem fixation (cemented vs. cementless),[119] femoral head bearing surface (oxinium vs. cobalt–chromium)[124] and femoral stem fixation (cemented vs. cementless)[129] (*Table 15*). The MD in WOMAC score of –0.12 (95% CI –7.58 to 7.34) observed for one RCT[145] suggested no difference between cross-linked

Study or subroup	XLPE cup liner Mean	SD	Total	Non-XLPE cup liner Mean	SD	Total	Weight	Mean difference IV, random, 95% CI
Engh *et al.* 2012[113]	88	14	111	86	15	109	68.1%	2.00 (–1.84 to 5.84)
McCalden *et al.* 2009[145]	86	13.1	50	83.1	15.4	50	31.9%	2.90 (–2.70 to 8.50)
Total (95% CI)			**161**			**159**	**100.0%**	**2.29 (–0.88 to 5.45)**

Heterogeneity: $\tau^2=0.00$; $\chi^2=0.07$, df=1 ($p=0.80$); $I^2=0\%$
Test for overall effect: $z=1.42$ ($p=0.16$)

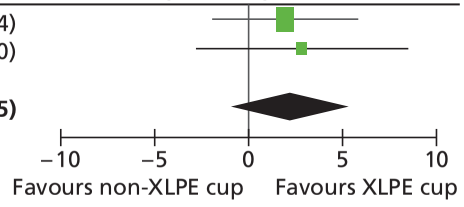


FIGURE 11 Harris Hip Score. XLPE, cross-linked polyethylene.

TABLE 14 Harris Hip Score (range 0–100): systematic reviews

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
3 years	NR[137]	2[137]	High quality[137]	Inconclusive
2–5 years	NR[138]	3[138]	Low quality[138]	
Femoral head-on-cup liner surface: metal-on-metal vs. metal-on-PE				
2 years	MD –2.40 (–4.47 to –0.33) (SS)[140]	4[140]	High quality[140]	Inconclusive
> 2 years	MD 1.21 (–2.41 to 4.83) (NS)[140]	2[140]		
Femoral head-on-cup liner surface: ceramic-on-ceramic vs. ceramic-on-PE				
NR	NR[140]	5[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: ceramic-on-PE vs. metal-on-PE				
NR	NR[140]	2[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: metal-on-metal vs. ceramic-on-ceramic				
NR	NR[140]	1[140]	High quality[140]	Inconclusive

MA, meta-analysis; NR, not reported; NS, not significant; PE, polyethylene; SS, statistically significant.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported. Two systematic reviews did not report this outcome.

TABLE 15 Western Ontario and McMaster University Osteoarthritis Index (range 0–100): RCTs

Follow-up	Arm-specific estimate, mean (SD or 95% CI)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup liner bearing surface: XLPE vs. non-XLPE				
1 year	83.0 (17.2) vs. 81.6 (17.6)[145]	MD 1.43 (–5.48 to 8.34)[c]	1 (unclear)	No difference
5 years	78.0 (19.4) vs. 78.1 (18.2)[145]	MD –0.12 (–7.58 to 7.34)[c]		
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
NA	Mean domain subscores only[119]	–	1 (low)	NA
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	84.9 (NR) vs. 87.0 (NR)[124]	$p > 0.159$ (NS)	1 (unclear)	Inconclusive
Femoral stem fixation: cemented vs. cementless				
16 years	11 (NR) vs. 13 (NR)[129]	$p = 0.927$ (NS)	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; NA, not applicable; NR, not reported; NS, not significant; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported. Nine RCTs[110–118,123,125–128] did not report any evidence on this outcome.

polyethylene and non-cross-linked polyethylene cup liners. Results for WOMAC score in the remaining three RCTs[119,124,129] were judged to be inconclusive because of incompletely reported data.

Systematic reviews (n = 0) No evidence was identified.

Other functional/clinical scores

Randomised controlled trials (n = 2) In one RCT[119] there was no difference in mean MACTAR scores (at 7 years: mean change difference 0.20, 95% CI –0.74 to 1.14) and Merle d'Aubigné and Postel scores (at 7 years: mean change difference –0.40, 95% CI –1.34 to 0.54) between patients who received a THR with cemented components and those who received a THR with cementless components (*Tables 16* and *17*). Results from one RCT[129] comparing femoral stem fixation (cemented vs. cementless) by the postoperative UCLA activity score were inconclusive because of incomplete data reporting (*Table 18*).

Systematic reviews (n = 1) The OHS was reported in one systematic review[137] comparing cup fixation methods (cemented vs. cementless), but the results were inconclusive (*Table 19*). This evidence was based on one RCT showing a statistically non-significant result.

TABLE 16 McMaster Toronto Arthritis Patient Preference Questionnaire scores (range 0–30): RCTs

Follow-up	Arm-specific estimate, mean (SD)	Difference (95% CI)	No. of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
	Mean change (postoperative):	Mean change difference:	1 (low)	No difference
3 months	–5.3 (2.5) vs. –5.2 (2.2)[119]	MD 0.10 (–0.51 to 0.71)[c]		
6 months	–6.6 (1.9) vs. –6.4 (2.1)[119]	MD 0.20 (–0.33 to 0.73)[c]		
1 year	–7.0 (1.8) vs. –6.9 (2.0)[119]	MD 0.10 (–0.41 to 0.61)[c]		
3 years	–6.6 (2.3) vs. –6.4 (2.3)[119]	MD 0.20 (–0.46 to 0.86)[c]		
5 years	–6.0 (2.8) vs. –6.2 (2.4)[119]	MD –0.20 (–0.45 to 0.55)[c]		
7 years	–6.2 (2.8) vs. –6.0 (2.6)[119]	MD 0.20 (–0.74 to 1.14)[c]		

SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported. Only the study by Corten *et al.*[119] reported any evidence for this outcome.

TABLE 17 Merle d'Aubigné and Postel scores (range 0–18): RCTs

Follow-up	Arm-specific estimate, mean (SD)	Difference (95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
	Mean change (postoperative):	Mean change difference:	1 (low)	No difference
3 months	5.8 (1.9) vs. 5.6 (2.2)[119]	MD 0.20 (–0.34 to 0.74)[c]		
6 months	6.7 (2.1) vs. 7.0 (2.2)[119]	MD –0.30 (–0.87 to 0.27)[c]		
1 year	7.5 (1.8) vs. 7.4 (2.1)[119]	MD 0.10 (–0.43 to 0.63)[c]		
3 years	7.1 (2.2) vs. 6.9 (2.1)[119]	MD 0.20 (–0.41 to 0.81)[c]		
5 years	6.5 (2.3) vs. 6.6 (2.4)[119]	MD –0.10 (–0.77 to 0.57)[c]		
7 years	6.1 (2.6) vs. 6.5 (2.8)[119]	MD –0.40 (–1.34 to 0.54)[c]		

SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported. Only the study by Corten *et al.*[119] reported any evidence for this outcome.

TABLE 18 University of California Los Angeles activity scores (range 1–10): RCTs

Follow-up	Arm-specific estimate, mean (SD or 95% CI)	Difference (*p*-value)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Femoral stem fixation: cemented vs. cementless				
16 years	7.6 (NR) vs. 7.8 (NR)[129]	$p = 0.814$ (NS)	1 (unclear)	Inconclusive

NR, not reported; NS, not significant; SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported. Only the study by Kim *et al.*[129] reported any evidence for this outcome.

TABLE 19 Oxford Hip Score (range 0–48): systematic review

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
3 years	NR[137]	1[137]	High quality[137]	Inconclusive

MA, meta-analysis; NR, not reported.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported. Only the study by Voigt *et al.*[137] reported this outcome.

Health-related quality of life

Only three RCTs[124,125,145] and one systematic review[140] reported any comparative evidence for measures of health-related quality of life.

Randomised controlled trials ($n = 3$) In one RCT,[145] at follow-up times of 1 and 5 years, there was no difference in quality of life (on the mental and physical subscales of SF-12) between two groups of patients receiving cross-linked and non-cross-linked polyethylene cup liner bearings (*Table 20*).

In two other RCTs[124,125] there were no statistically significant differences in mean SF-12 mental and physical subscale scores between THR groups with different femoral head bearings (oxinium vs. cobalt–chromium)[124] and femoral head-on-cup liner articulations (ceramic-on-ceramic vs. ceramic-on-polyethylene).[125] This evidence was judged to be inconclusive (see *Table 20*).

TABLE 20 Short Form questionnaire-12 items (range 0–100): RCTs

Follow-up	Arm-specific estimate, mean (SD)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup liner bearing surface: XLPE vs. non-XLPE				
1 year	MCS: 55.79 (7.38) vs. 56.01 (8.55);[145] PCS: 42.20 (11.37) vs. 40.86 (11.11)[145]	MCS: MD –0.22 (–3.38 to 2.94);[c] PCS: MD 1.34 (–3.12 to 5.80)[c]	1 (unclear)	No difference
5 years	MCS: 55.24 (8.01) vs. 53.36 (10.13);[145] PCS: 37.24 (12.16) vs. 40.00 (11.78)[145]	MCS: MD 1.88 (–1.74 to 5.50);[c] PCS: MD –2.76 (–7.51 to 1.99)[c]		
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	MCS: 53.80 (NR) vs. 52.57 (NR);[124] PCS: 45.20 (NR) vs. 49.20 (NR)[124]	MCS: $p > 0.05$ (NS); PCS: $p > 0.05$ (NS)	1 (unclear)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	NR[125]	$p > 0.05$ (NS)	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; MCS, mental component summary score; NR, not reported; NS, not significant; PCS, physical component summary score; PE, polyethylene; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported.

Systematic reviews ($n = 1$) One systematic review[140] reported evidence from two studies that compared SF-12 scores across different articulations (metal-on-metal vs. metal-on-polyethylene) (*Table 21*). The review did not provide any formal narrative or quantitative synthesis of the data. The evidence was considered to be inconclusive.

Revision

Evidence on revision was reported for 10 RCTs[112,113,115,119,123–125,127–129] and five systematic reviews.[137–141]

Randomised controlled trials ($n = 10$) One RCT[113] demonstrated a reduced risk of revision in patients who received cross-linked polyethylene compared with non-cross-linked polyethylene cup liners (RR 0.18, 95% CI 0.04 to 0.78) (*Table 22*). The evidence reported in the remaining nine RCTs showed statistically non-significant differences in the risk of revision between the different types of THR with wide CIs compatible with large size effects in both directions (i.e. favouring one or other of the treatment group). This evidence was deemed to be inconclusive (see *Table 22*).

Systematic reviews ($n = 5$) Of the five systematic reviews reporting on revisions, two[137,141] provided pooled estimates for risk of revision (*Table 23*). According to one review,[141] at 9 years post surgery the recipients of zirconium femoral heads were at similar risk for revision as the recipients of non-zirconium femoral heads (three pooled RCTs; risk difference 0.02, 95% CI –0.01 to 0.06). This evidence was considered conclusive in detecting no difference in revision rates between these two types of femoral head.

In another review[137] the risk of revision at 10 years after surgery did not significantly differ between cemented and cementless cup fixation THR groups (pooled RR 0.15, 95% CI 0.02 to 1.18). This result was considered inconclusive given the uninformative 95% CIs. Evidence from the remaining three reviews[138–140] was of a narrative nature, which precluded us drawing conclusions (see *Table 23*).

Mortality

Evidence on mortality was reported for six RCTs.[110,113,119,123,128,145] None of the five systematic reviews reported on mortality.

TABLE 21 Short Form questionnaire-12 items (range 0–100): systematic review

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Femoral head-on-cup liner surface: metal-on-metal vs. metal-on-PE				
2–3 years	NR[140]	2[140]	High quality[140]	Inconclusive

MA, meta-analysis; NR, not reported; PE, polyethylene.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 22 Revision rate: RCTs

Follow-up	Arm-specific counts, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	17/183 vs. 11/104[112]	$p > 0.05$ (NS); RR 0.87 (95% CI 0.42 to 1.80)[c]	1 (low)	Inconclusive
Cup liner bearing surface: XLPE vs. non-XLPE				
10 years	2/111 vs. 11/109[113]	$p < 0.05$ (SS); RR 0.18 (95% CI 0.04 to 0.78)[c]	1 (unclear)	In favour of XLPE cup liner
Cup shell design: porous-coated shell vs. arc-deposited HA-coated shell				
5 years	2/113 vs. 4/109[115]	$p > 0.05$ (NS); RR 0.48 (95% CI 0.09 to 2.57)[c]	1 (low)	Inconclusive
5–10 years	2/113 vs. 2/109[115]	$p > 0.05$ (NS); RR 0.96 (95% CI 0.13 to 6.72)[c]		
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
7 years	13/124 vs. 6/126[119]	$p = 0.11$ (NS); RR 2.20 (95% CI 0.86 to 5.60)[c]	1 (low)	Inconclusive[d]
Femoral head size: 36 mm vs. 28 mm				
1 year	4/273 vs. 6/284[123]	$p =$ NR; RR 0.69 (95% CI 0.19 to 2.43)[c]	1 (low)	Inconclusive
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	1/50 vs. 1/50[124]	$p =$ NR; RR 1.00 (95% CI 0.06 to 15.50)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. metal-on-PE				
5 years	6/222 vs. 8/106[115]	$p = 0.045$ (SS); RR 0.35 (95% CI 0.12 to 1.00)[c]	1 (low)	Inconclusive
5–10 years	4/222 vs. 5/106[115]	$p = 0.08$ (NS); RR 0.38 (95% CI 0.10 to 1.39)[c]		
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	11/196 vs. 3/161[125]	$p = 0.06$ (NS); RR 3.01 (95% CI 0.85 to 10.61)[c]	1 (low)	Inconclusive
Femoral stem composition: CoCr vs. titanium				
5 years	2/199 vs. 0/191[127]	$p = 0.16$ (NS); RR and 95% CI not estimated	1 (unclear)	Inconclusive
Femoral stem design: short metaphyseal-fitting stem vs. conventional metaphyseal- and diaphyseal-fitting stem				
3 years	0/50 vs. 0/50[128]	$p =$ NR; RR and 95% CI not estimated	1 (low)	Inconclusive
Femoral stem fixation: cemented vs. cementless				
20 years	Acetabular: 14/109 vs. 18/110[129]	$p = 0.673$ (NS); RR 0.78 (95% CI 0.41 to 1.49)[c]	1 (low)	Inconclusive
	Femoral: 3/109 vs. 4/110[129]	$p = 0.912$ (NS); RR 0.75 (95% CI 0.17 to 3.30)[c]		

CoCr, cobalt–chromium; HA, hydroxyapatite; NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias; SS, statistically significant; XLPE, cross-linked polyethylene.

a Decision was consensus based.

b Favours THR-1 (or THR-2), no difference or inconclusive.

c Calculated.

d The use of cementless implants (cup and femoral stem) was associated with a better implant survival rate than the use of cemented implants at 10 years (83% vs. 94%, $p = 0.007$), 15 years (66% vs. 80%, $p = 0.007$) and 20 years (48% vs. 69%, $p = 0.007$) post procedure[119] and cementless implants were therefore rated as favourable.

Only those comparisons listed for which any evidence for the given outcome was reported. Three RCTs[110,111,126,145] did not report any evidence for this outcome.

TABLE 23 Revision rate: systematic reviews

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
4–8 years	RR 0.15 (0.02 to 1.18) (NS)[137]	2[137]	High quality,[137] low quality,[138] low quality[139]	Inconclusive
10 years	RR 1.36 (0.81 to 1.29) (NS)[137]	2[137]		
< 10 years	NR[138]	6[138]		
5–15 years	NR[139]	NR[139]		
Femoral head-on-cup liner surface: metal-on-metal vs. metal-on-PE				
2–5 years	NR[140]	2[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: ceramic-on-ceramic vs. metal-on-PE				
6–8 years	NR[140]	1[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: ceramic-on-ceramic vs. ceramic-on-PE				
2–8 years	NR[140]	5[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: ceramic-on-PE vs. metal-on-PE				
8 years	NR[140]	1[140]	High quality[140]	Inconclusive
Femoral head-on-cup liner surface: zirconia-on-PE vs. non-zirconia-on-PE				
9 years	RD 0.02, 95% CI –0.01 to 0.06 (NS)[141]	3[141]	Medium quality[141]	No difference

MA, meta-analysis; NR, not reported; NS, not significant; PE, polyethylene; RD, risk difference.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported. All systematic reviews reported this outcome.

Randomised controlled trials ($n = 6$) Evidence from the six RCTs[110,113,119,123,128,145] that reported mortality was inconclusive because of non-significant RR estimates and wide 95% CIs (*Table 24*). For example, based on a pooled RR estimate of 1.39 (95% CI 0.78 to 2.49),[113,145] 5- to 10-year post-surgery mortality rates in the group receiving cross-linked polyethylene cup liners were not significantly different from those in the group receiving non-cross-linked polyethylene cup liners (*Figure 12*). Similarly, the rest of the studies showed non-significant results for mortality between THR groups defined by femoral stem and/or cup fixation (cemented vs. cementless)[110,119] and femoral head size (36 mm vs. 28 mm).[123] One RCT[128] reported no deaths for both treatment groups receiving femoral stems of different design.

Systematic reviews ($n = 0$) No evidence was identified.

Femoral head penetration rate (measure of prosthesis movement)

Evidence on femoral head penetration rate was reported by three RCTs.[113,126,145] None of the five systematic reviews reported this end point.

RCTs ($n = 3$) Two RCTs[113,145] demonstrated reduced femoral head penetration in favour of cross-linked polyethylene cup liners compared with non-cross-linked (conventional) polyethylene cup liners at 5–10 years of follow-up (*Table 25*). Similarly, in another RCT,[126] cross-linked polyethylene cup liners with either metal or oxinium femoral heads outperformed conventional polyethylene cup liners in reducing femoral head penetration during 2 years of follow-up.

TABLE 24 Mortality rate: RCTs

Follow-up	Arm-specific count (*n*/*N*)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	12/107 vs. 14/108[110]	p = NR; RR 0.86 (95% CI 0.41 to 1.78)[c]	1 (low)	Inconclusive
Cup liner bearing surface: XLPE vs. non-XLPE				
5 years	7/50 vs. 2/50[145]	$p > 0.05$ (NS); RR 3.50 (95% CI 0.76 to 16.03)[c]	2 (unclear)	Inconclusive
10 years	17/111 vs. 15/109[113]	$p > 0.05$ (NS); RR 1.11 (95% CI 0.58 to 2.11)[c]		
		Pooled estimate of MH-RR: RR 1.39 (95% CI 0.78 to 2.49)[113,145]		
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
7 years	18/124 vs. 17/126[119]	p = NR; RR 1.07 (95% CI 0.58 to 1.98)[c]	1 (low)	Inconclusive
Femoral head size: 36 mm vs. 28 mm				
1 year	5/273 vs. 2/284[123]	p = NR; RR 2.58 (95% CI 0.53 to 13.20)[c]	1 (low)	Inconclusive
Femoral stem design: short metaphyseal-fitting stem vs. conventional metaphyseal- and diaphyseal-fitting stem				
3 years	0/50 vs. 0/50[128]	p = NR; RR and 95% CI not estimated	1 (low)	Inconclusive

MH-RR, Mantel–Haenszel rate ratio; NR, not reported; PE, polyethylene; NS, not significant; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported. Seven RCTs[112,115–118,124–127,129] did not report any evidence for this outcome.

FIGURE 12 Mortality. XLPE, cross-linked polyethylene.

TABLE 25 Femoral head penetration rate: RCTs

Follow-up	Arm-specific estimate (mm/year), mean (SD or 95% CI)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Cup liner bearing surface: XLPE vs. non-XLPE				
5 years	0.003 (–0.024 to 0.030) vs. 0.051 (0.029 to 0.073)[145]	$p = 0.006$ (SS)	2 (unclear)	In favour of XLPE
5 years	0.24 (0.42) vs. 1.26 (0.62)[113]	$p < 0.001$ (SS)		
10 years	0.06 (0.05) vs. 0.22 (0.11)[113]	$p < 0.001$ (SS)		
Femoral head-on-cup liner bearing surface: steel-on-PE vs. CoCr-on-PE vs. oxinium-on-PE vs. CoCr-on-XLPE vs. oxinium-on-XLPE				
2 years	0.19 (0.16 to 0.23) vs. 0.40 (0.33 to 0.46) vs. 0.44 (0.37 to 0.51) vs. 0.19 (0.15 to 0.23) vs. 0.18 (0.13 to 0.22)[126]	$p < 0.001$ (SS; steel-on-PE, CoCr-on-XLPE and oxinium-on-XLPE vs. CoCr-on-PE and oxinium-on-PE)	1 (low)	In favour of CoCr-on-XLPE, oxinium-on-XLPE and steel-on-PE

CoCr, cobalt–chromium, PE, polyethylene; SROB, summary risk of bias; SS, statistically significant; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported. Three RCTs[113,126,145] reported this outcome.

Systematic reviews ($n = 0$) No evidence was identified.

Complications

Evidence on the occurrence/absence of complications was reported by nine RCTs[112,113,115,123–125,127,129,145] and three systematic reviews.[138–140] In most studies[112,113,115,123–125,129,145] the reported complications were classified as postoperative. In one RCT[127] some of the complications were classified as perioperative.

Implant dislocation

Randomised controlled trials ($n = 7$) Evidence on the occurrence/absence of implant dislocation was reported by seven RCTs[110,112,115,123–125,127] (*Table 26*). Our pooled estimate of two studies[110,112] (*Figure 13*) indicated a reduced risk of implant dislocation at 10 years' follow-up in recipients of cemented compared with cementless cups (pooled OR 0.34, 95% CI 0.13 to 0.89). Moreover, in one RCT[123] after 1 year of follow-up, the THR recipients with a larger size of femoral head experienced a lower risk of implant dislocation than those with a smaller size of femoral head (36 mm vs. 28 mm: RR 0.17, 95% CI 0.04 to 0.78). Evidence on implant dislocation for the remaining four RCTs[115,124,125,127] was inconclusive because of incomplete data and non-significant results.

Systematic reviews ($n = 2$) Overall, no conclusions on implant dislocation could be drawn from the two systematic reviews, given the narrative evidence summary[140] and the mixed study designs[139] (*Table 27*). The pooled data from one review[139] was based on nine studies, most of which were not randomised and which indicated a lower risk of dislocation in the groups receiving cemented compared with cementless cups.

TABLE 26 Implant dislocation rate: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	4/107 vs. 10/108[110]	$p > 0.05$ (NS); RR 0.40 (95% 0.13 to 1.24)[c]	2 (low)	In favour of cemented cup
	1/183 vs. 3/104[112]	$p =$ NR; RR 0.18 (95% 0.02 to 1.79)[c]		
		Pooled estimate of Peto OR: 0.34 (95% 0.13 to 0.89)[110,112c]		
Cup shell design: porous-coated shell vs. arc-deposited HA-coated shell				
10 years	2/113 vs. 3/109[115]	$p =$ NR; RR 0.64 (95% 0.10 to 3.77)[c]	1 (low)	Inconclusive
Femoral head size: 36 mm vs. 28 mm				
1 year	2/258 vs. 12/275[123]	$p =$ NR; RR 0.17 (95% 0.04 to 0.78)[c]	1 (low)	In favour of 36-mm head size
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	2/50 vs. 1/50[124]	$p =$ NR; RR 2.00 (95% 0.18 to 21.35)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	10/166 vs. 9/146[125]	$p = 0.672$ (NS); RR 0.97 (95% CI 0.40 to 2.33)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. metal-on-PE				
10 years	5/222 vs. 5/106[115]	$p = 0.25$ (NS); RR 0.47 (95% CI 0.14 to 1.61)[c]	1 (low)	Inconclusive
Femoral stem composition: CoCr vs. titanium				
5 years	3/199 vs. 0/191[127]	$p = 0.678$ (NS); RR and 95% CI not estimated	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; HA, hydroxyapatite; NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported. Six RCTs[113,119,126,128,129,145] did not report any evidence for this outcome.

Study or subgroup	Cemented cup Events	Total	Cementless cup Events	Total	Weight	Peto's OR Peto, fixed, 95% CI
Angadi *et al.* 2012[112]	1	183	3	104	21.8%	0.18 (0.02 to 1.42)
Bjørgul *et al.* 2010[110]	4	107	10	108	78.2%	0.41 (0.14 to 1.20)
Total (95% CI)		**290**		**212**	**100.0%**	**0.34 (0.13 to 0.89)**
Total events	5		13			
Heterogeneity: χ^2=0.45, df=1 (p=0.50); I^2=0%						
Test for overall effect: z=2.20 (p=0.03)						

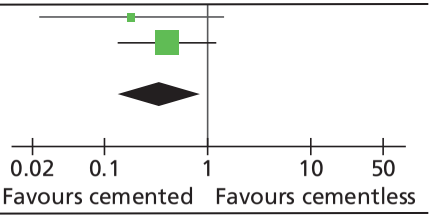


FIGURE 13 Implant dislocation.

TABLE 27 Implant dislocation rate: systematic reviews

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
5–15 years	12/914 (1.3%) vs. 28/696 (4.1%) ($p = 0.001$).[139] Pooled data from nine comparative studies (most non-RCTs) suggested that cemented cups had lower dislocation rates than cementless cups	NR[139]	Low quality[139]	Inconclusive
Femoral head-on-cup liner surface: metal-on-metal vs. metal-on-PE				
2–5 years	NR.[140] No significant difference based on results from three RCTs	3[140]	High quality[140]	Inconclusive

MA, meta-analysis; NR, not reported; PE, polyethylene.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported.

Osteolysis

Randomised controlled trials ($n = 7$) Evidence on osteolysis was reported by seven RCTs[112,113,115,125,127,129,145] (*Table 28*). In one RCT[115] comparing different femoral head-on-cup liner bearing surfaces, recipients of ceramic-on-ceramic articulations had a reduced risk of osteolysis compared with recipients of metal-on-polyethylene articulations at 10 years post operation (RR 0.10, 95% CI 0.02 to 0.32).

For seven RCTs, the evidence for osteolysis was inconclusive across the comparisons based on different methods of cup fixation (cemented vs. cementless),[112] cup liner bearing surface (cross-linked polyethylene vs. non-cross-linked polyethylene),[113,145] cup shell design (porous coated vs. arc-deposited hydroxyapatite coated),[115] femoral head-on-cup liner bearing surface (ceramic-on-ceramic vs. ceramic-on-polyethylene),[125] femoral stem composition (cobalt–chromium vs. titanium)[127] and femoral stem fixation (cemented vs. cementless).[129]

Systematic reviews ($n = 2$) Overall, no conclusions could be drawn on the incidence of osteolysis from two low-quality systematic reviews[138,139] comparing cemented and cementless methods of cup fixation, given the narrative evidence summaries, mixed study designs and inconsistent findings (*Table 29*).

Other complications

Randomised controlled trials ($n = 7$) Seven RCTs reported other complications such as aseptic loosening (*Table 30*),[112,113,119,124,127] femoral fracture (*Table 31*),[113,115,127] infection (*Table 32*),[112,124,125,127] and deep-vein thrombosis (*Table 33*).[125] This evidence was judged to be inconclusive because of low event or zero event counts and CIs indicating great uncertainty.

Systematic reviews ($n = 1$) Of other complications, only aseptic loosening was reported in one low-quality systematic review[139] (*Table 34*). Pooled data from 11 studies, most of which were not randomised, pointed towards a greater risk of aseptic loosening with cemented compared with cementless cups; however, the evidence is inconclusive given the lack of numerical data and the evidence synthesis being based on mixed study designs.

TABLE 28 Osteolysis: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	0/183 vs. 1/104[112]	p = NR; RR and 95% CI not estimated	1 (low)	Inconclusive
Cup liner bearing service: XLPE vs. non-XLPE				
5 years	0/50 vs. 0/50[145]	p = NR; RR and 95% CI not estimated	2 (unclear)	Inconclusive
10 years	0/111 vs. 15/109[113]	$p < 0.001$; RR and 95% CI not estimated		
Cup shell design: porous-coated shell vs. arc-deposited HA-coated shell				
10 years	1/113 vs. 2/109[115]	p = NR; RR 0.48 (95% CI 0.04 to 5.24)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	1/166 vs. 1/146[125]	p = 0.797 (NS); RR 0.87 (95% CI 0.05 to 13.93)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. metal-on-PE				
10 years	3/222 vs. 15/106[115]	$p < 0.001$ (SS); RR 0.10 (95% CI 0.02 to 0.32)[c]	1 (low)	In favour of ceramic-on-ceramic bearing surface
Femoral stem composition: CoCr vs. titanium				
5 years	0/199 vs. 0/191[127]	p = NR; RR and 95% CI not estimated	1 (unclear)	Inconclusive
Femoral stem fixation: cemented vs. cementless				
20 years	Acetabular: 35/109 vs. 40/110[129]	p = 0.168 (NS); RR 0.88 (95% CI 0.61 to 1.27)[c]	1 (low)	Inconclusive
	Femoral: 31/109 vs. 35/110[129]	p = 0.159 (NS); RR 0.89 (95% CI 0.59 to 1.33)[c]		

CoCr, cobalt–chromium; HA, hydroxyapatite; NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 29 Osteolysis: systematic reviews

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
2–6 years	NR.[138] The analysis and narrative synthesis of RCT data showed no statistically significant difference in the occurrence of osteolysis between cemented and cementless cups	3[138]	Low quality[138]	Inconclusive
5–15 years	NR.[139] The narrative synthesis of nine comparative studies (most non-RCTs) indicated lower rates of osteolysis with cemented cups	NR[139]	Low quality[139]	Inconclusive

MA, meta-analysis; NR, not reported.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 30 Aseptic loosening: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	11/183 vs. 2/104[112]	*p* = NR; RR 3.12 (95% CI 0.70 to 13.83)[c]	1 (low)	Inconclusive
Cup liner bearing surface: XLPE vs. non-XLPE				
10 years	0/111 vs. 0/109[113]	NA; RR and 95% CI not estimated	1 (unclear)	Inconclusive
Cup and femoral stem fixation: cemented cup/femoral stem vs. cementless cup/femoral stem				
20 years	9/124 vs. 4/126[119]	*p* = NR; RR 2.28 (95% CI 0.72 to 7.23)[c]	1 (low)	Inconclusive
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	0/50 vs. 1/50[124]	*p* = NR; RR and 95% CI not estimated	1 (low)	Inconclusive
Femoral stem composition: CoCr vs. titanium				
5 years	1/199 vs. 0/191[127]	*p* = 0.324 (NS); RR and 95% CI not estimated	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; NA, not applicable; NR, not reported; NS, not significant; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 31 Femoral fracture: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup liner bearing service: XLPE vs. non-XLPE				
10 years	2/111 vs. 0/109[113]	*p* = NR; RR and 95% CI not estimated	1 (unclear)	Inconclusive
Cup shell design: porous-coated shell vs. arc-deposited HA-coated shell				
10 years	0/113 vs. 0/109[115]	NA; RR and 95% CI not estimated	1 (low)	Inconclusive
Femoral stem composition: CoCr vs. titanium				
5 years	0/199 vs. 1/191[127]	*p* = 0.309 (NS); RR and 95% CI not estimated	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; HA, hydroxyapatite; NA, not applicable; NR, not reported; NS, not significant; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 32 Infection: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Cup fixation: cemented vs. cementless				
10 years	0/183 vs. 2/104[112]	*p* = NR; RR and 95% CI not estimated	1 (low)	Inconclusive
Femoral head bearing surface: oxinium femoral heads vs. CoCr femoral heads				
2 years	1/50 vs. 1/50[124]	*p* = NR; RR 1.00 (95% CI 0.06 to 15.55)[c]	1 (low)	Inconclusive
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	Superficial: 6/166 vs. 3/146[125]	*p* = 0.357 (NS); RR 1.75 (95% CI 0.44 to 6.90)[c]	1 (low)	Inconclusive
	Deep: 1/166 vs. 2/146[125]	*p* = 0.909 (NS); RR 0.43 (95% CI 0.04 to 4.79)[c]		
Femoral stem composition: CoCr vs. titanium				
5 years	1/199 vs. 0/191[127]	*p* = 0.324 (NS); RR and 95% CI not estimated	1 (unclear)	Inconclusive

CoCr, cobalt–chromium; NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 33 Deep-vein thrombosis: RCTs

Follow-up	Arm-specific count, *n*/*N*	Difference (*p*-value or 95% CI)	Number of RCTs (SROB)[a]	Treatment effect conclusion[b]
Femoral head-on-cup liner bearing surface: ceramic-on-ceramic vs. ceramic-on-PE				
5 years	3/166 vs. 2/146[125]	*p* = 0.909 (NS); RR 1.31 (95% CI 0.22 to 7.78)[c]	1 (low)	Inconclusive

NR, not reported; NS, not significant; PE, polyethylene; SROB, summary risk of bias.
a Decision was consensus based.
b Favours THR-1 (or THR-2), no difference or inconclusive.
c Calculated.
Only those comparisons listed for which any evidence for the given outcome was reported.

TABLE 34 Aseptic loosening: systematic review

Follow-up	Pooled effect estimate (95% CI)	Number of RCTs in MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
Cup fixation: cemented vs. cementless				
5–15 years	NR.[139] Pooled data from 11 comparative studies (most non-RCTs) presented only graphically suggested higher rates of aseptic loosening with cemented vs. cementless cups	NR[139]	Low quality[139]	Inconclusive

MA, meta-analysis; NR, not reported.
a Favours THR-1 (or THR-2), no difference or inconclusive.
Only those comparisons listed for which any evidence for the given outcome was reported.

Grading the overall quality of the evidence

The results for graded outcomes are presented in the evidence profile (*Table 35*). For a meaningful grading process and for consistency, only the THR comparison categories that included at least two studies (cup fixation – cemented vs. cementless and cup liner bearing surface: cross-linked polyethylene vs. non-cross-linked polyethylene) were selected. The overall quality for gradable outcomes across the THR comparison categories (cup fixation and cup liner bearing surface) was as follows: HHS – moderate grade; WOMAC score – not graded and very low grade, respectively; revision – very low grade; mortality – very low grade and low grade, respectively; femoral head penetration – not graded and moderate grade, respectively; and implant dislocation – high grade and not graded, respectively.

Summary conclusions for the comparison between different types of total hip replacement

Randomised controlled trials

The majority of the evidence comparing THRs was rated as inconclusive by us (*Table 36*). In three RCTs there was evidence of a reduced risk of implant dislocation with the use of a cemented cup (vs. a cementless cup)[110,112] or a larger femoral head size (36 mm vs. 28 mm)[123] (high-grade evidence for the cup fixation comparison). In three other RCTs, patients who received a THR with a cross-linked polyethylene cup liner experienced a reduced (i.e. improved) femoral head penetration rate (moderate-grade evidence) [113,126,145] and risk for revision (very low-grade evidence)[113] compared with recipients of conventional polyethylene cup liners. In one RCT[119] the use of cementless fixation of the cup and femoral stem (vs. cemented fixation) was associated with a better implant survival rate. Moreover, the recipients of ceramic-on-ceramic articulations (vs. metal-on-polyethylene) experienced a reduced risk of osteolysis.[115] For half of the studies,[110,112,113,119,126,145] the mean post-THR clinical and functional scores (i.e. HHS, WOMAC score, SF-12 score, MACTAR score, Merle d'Aubigné and Postel score) measured at different follow-up times were similar between the different THR treatment groups (moderate-grade evidence for no difference in HHS across the comparisons for cup fixation and cup liner surface types).

Evidence from studies reporting the UCLA activity score,[129] mortality (very low-grade evidence),[110,113,119,123,128,145] aseptic loosening,[112,113,119,124,127] femoral fracture,[113,115,127] infection[112,124,125,127] and deep-vein thrombosis[125] was all inconclusive. Also, the evidence reported in four studies was considered inconclusive for all outcomes (very low-grade evidence).[124,125,127,128] Results were considered inconclusive by us because of partial reporting (missing data for effect estimates, CIs, SEs, SDs, *p*-values), great uncertainty (wide CIs), zero event counts and/or inconsistency in estimates.

Systematic reviews

The majority of evidence from the five systematic reviews comparing different types of THR[137–141] was considered inconclusive. This is because of unreported pooled results across RCTs (i.e. reporting only narrative syntheses), the reporting of inappropriate pooling methods (e.g. indirect naive comparison of single-group cohorts; pooling of studies of different design)[138,139,141] or the reporting of inconsistent summary findings[140] (*Table 37*). The evidence from one review[141] indicated no difference in the risk for revision between two different articulations of zirconium-on-polyethylene and non zirconium-on-polyethylene.

TABLE 35 Grading of Recommendations, Assessment, Development and Evaluation evidence profile for gradable outcomes reported in RCTs of THR[a]

Outcome (follow-up timing)	Number of studies reporting outcome (participants)	Pooled effect estimate (95% CI) and conclusion	SROB across studies	Consistency	Directness	Precision	Outcome reporting bias	Quality of the evidence (GRADE)[b]
Cup fixation (cemented vs. cementless) – two RCTs[110,112]								
HHS (6 months-10 years)	2 (502)	None; no difference	Unclear	Consistent	Direct	Precise	Unlikely	Moderate
WOMAC score (NA)	0	NA	NA	NA	NA	NA	NA	NA (no evidence)
Revision (10 years)	1 (287)	None; inconclusive	Low	NA	Direct	Imprecise	Likely	Very low
Mortality (10 years)	1 (215)	None; inconclusive	Low	NA	Direct	Imprecise	Likely	Very low
Femoral head penetration (NA)	0	NA	NA	NA	NA	NA	NA	NA (no evidence)
Implant dislocation (10 years)	2 (502)	OR 0.34 (0.13 to 0.89); in favour of cemented cup	Low	Consistent	Direct	Precise	Unlikely	High
Cup liner bearing surface (XLPE vs. non-XLPE) – two RCTs[113,145]								
HHS (1–10 years)	2 (320)	MD 2.29 (–0.88 to 5.45); no difference	Unclear	Consistent	Direct	Precise	Unlikely	Moderate
WOMAC score (1–5 years)	1 (100)	None; no difference	Unclear	NA	Direct	Precise	Likely	Very low
Revision (10 years)	1 (220)	None; in favour of XLPE cup liner	Unclear	NA	Direct	Precise	Likely	Very low
Mortality (5–10 years)	2 (320)	RR 1.39 (0.78 to 2.49); inconclusive	Unclear	Consistent	Direct	Imprecise	Unlikely	Low
Femoral head penetration (5–10 years)	2 (320)	None; in favour of XLPE cup liner	Unclear	Consistent	Direct	Precise	Unlikely	Moderate
Implant dislocation (NA)	0	NA	NA	NA	NA	NA	NA	NA (no evidence)

NA, not applicable; SROB, summary risk of bias; XLPE, cross-linked polyethylene.
a Adapted from Guyatt *et al.*[102]
b GRADE categories: high, moderate, low, very low, NA (no evidence).

TABLE 36 Summary of evidence regarding the differences between the different types of THR for each reported outcome: RCTs

Conclusive evidence suggesting a difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Cup fixation: cemented vs. cementless[110,112]		
Implant dislocation (high-grade evidence)[110,112] (in favour of cemented)	HHS (moderate-grade evidence)[110,112]	Mortality (very low-grade evidence),[110] revision (very low-grade evidence),[112] osteolysis,[112] aseptic loosening,[112] infection[112]
Cup liner bearing surface: XLPE vs. non-XLPE[113,145]		
Femoral head penetration (moderate-grade evidence)[113,145] revision rate (very low-grade evidence)[113] (in favour of XLPE)	HHS (moderate grade evidence),[113,145] WOMAC score (very low grade evidence)[145] SF-12 score (mental/physical)[145]	Mortality (low-grade evidence),[113,145] aseptic loosening,[113] femoral fracture[113]
Cup shell design: porous coated vs. arc-deposited HA coated[115]		
None	None	HHS, revision, implant dislocation, osteolysis, femoral fracture
Cup and femoral stem fixation: cemented vs. cementless[119]		
None[a]	HHS, Merle d'Aubigné and Postel score, MACTAR score	WOMAC score, mortality, revision, aseptic loosening
Femoral head size: 36 mm vs. 28 mm[123]		
Implant dislocation (in favour of 36 mm)	None	Mortality, revision
Femoral head bearing surface: oxinium vs. CoCr[124]		
None	None	HHS, SF-12 score, WOMAC score, revision, implant dislocation, aseptic loosening, infection
Femoral head-on-cup liner bearing: ceramic-on-ceramic vs. metal-on-PE[115]		
Osteolysis (in favour of ceramic-on-ceramic)	None	HHS, revision, implant dislocation
Femoral head-on-cup liner bearing: ceramic-on-ceramic vs. ceramic-on-PE[125]		
None	None	HHS, SF-12 score, revision, implant dislocation, osteolysis, infection, deep-vein thrombosis
Femoral head-on-cup liner bearing: steel-on-PE vs. CoCr/oxinium-on-XLPE vs. CoCr/oxinium-on-PE[126]		
Femoral head penetration (in favour of steel-on-PE or CoCr/oxinium-on-XLPE)	HHS	None
Femoral stem composition: CoCr vs. titanium[127]		
None	None	HHS, revision, implant dislocation, osteolysis, aseptic loosening, femoral fracture, infection
Femoral stem design: short metaphyseal fitting vs. conventional metaphyseal and diaphyseal fitting[128]		
None	None	HHS, mortality, revision
Femoral stem fixation: cemented vs. cementless[129]		
None	None	HHS, UCLA activity score, WOMAC score, revision, osteolysis

CoCr, cobalt–chromium; HA, hydroxyapatite; PE, polyethylene; XLPE, cross-linked polyethylene.
a Implant survival rate was in favour of cementless.[119]

TABLE 37 Summary of evidence regarding the differences between the different types of THR for each reported outcome: systematic reviews

Conclusive evidence suggesting a difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Cup fixation: cemented vs. cementless[137–139]		
None	None	HHS,[137,138] OHS,[137] revision,[137–139] aseptic loosening[139]
[a]***Femoral head-on-cup liner bearing: different comparisons***[140,141]		
None	Revision[141]	HHS,[140] SF-12 score,[140] revision,[140] implant dislocation[140]

PE, polyethylene.
a Metal-on-metal vs. metal-on-PE,[140] ceramic-on-ceramic vs. ceramic-on-PE,[140] ceramic-on-PE vs. metal-on-PE,[140] metal-on-metal vs. ceramic-on-ceramic,[140] zirconia-on-PE vs. non-zirconia-on-PE.[141]

Other analysis

Publication bias

The extent to which publication bias could have influenced the pooled treatment effect estimates (i.e. degree of funnel plot asymmetry) could not be explored because of an insufficient number of data points in the forest/funnel plots.

Heterogeneity, subgroup effects and sensitivity analysis

The data reviewed from RCTs were too sparse and heterogeneous (in terms of different types of THR) to allow exploration of whether or not the relative effect of any given THR differed by study-level methodological characteristics (i.e. risk of bias, type of data analysis) or patient-related characteristics (i.e. age, sex or functional status). None of the included RCTs reported within-study subgroup effects of the different THRs compared.

Comparison between total hip replacement and resurfacing arthroplasty

Study and participant characteristics

Randomised controlled trials

Study and participant characteristics of the three included RCTs[130–132] are summarised in *Table 38*. More details can be found in *Appendices 3* and *4*. Two RCTs[131,132] were conducted in Canada and one[130] was conducted in the UK. A total of 422 participants were randomised across the three RCTs, ranging from 104[131] to 192[132] participants. The mean age of participants ranged from 50[132] to 56[130] years and the proportion of female participants across the studies ranged from 10.5%[131] to 41%.[130] The total length of follow-up the studies ranged from 1 year[130] to 6 years.[132] The proportion of participants diagnosed with primary OA was reported for two studies[130,132] and was 33%[132] and 95%.[130]

TABLE 38 Overall study characteristics across the three RCTs comparing THR and RS

Study characteristic	Metric
Geographical region	UK ($n = 1$), Canada ($n = 2$)
Total number of randomised participants	422 (range 104–192)
Mean age (years)	Range 50–56
Female participants (%)	Range 10.5–41
Length of follow-up (years)	Range 1–6
Diagnosis of primary OA (%)	Range 33–95

The three RCTs reported on clinical/functional scores (e.g. HHS, OHS, UCLA activity score, WOMAC score), health-related quality of life and risk of revision. Follow-up of outcome assessments ranged from 3 weeks[130] to 5 years.[132] Outcomes reported in the included studies can be found in *Appendix 8*.

Systematic reviews

Three systematic reviews[142–144] were included that evaluated the clinical effectiveness of THR compared with RS with respect to postoperative clinical/function (HHS, WOMAC score), risk of revision, mortality and complications. Searches for these systematic reviews were undertaken between March 2008[144] and January 2010.[143] Evidence was synthesised from both RCTs and non-RCTs (see *Appendices 3* and *4*). Further details on specific outcomes reported (or not reported) in the included systematic reviews can be found in *Appendix 8*.

Risk of bias and methodological quality

Risk of bias in randomised controlled trials

The risk of bias assessment for the three included RCTs[130–132] comparing THR with RS is presented in risk of bias tables (see *Appendix 2*), the summary table (*Table 39*) and the risk of bias graph (*Figure 14*). Overall, two studies[130,132] reported an adequate method for random sequence generation and all three studies[130–132] reported treatment allocation concealment (low risk of bias). Two of the three studies[130,132] were rated as having a low risk of performance and detection bias for objective outcomes (e.g. revision, dislocation). The same two studies had a high risk of performance bias for subjective outcomes (e.g. patient-administered functional scores). Patients and study personnel were blinded in only one study.[131] For two studies[130,132] the influence of attrition bias on objective outcomes was judged to be of low risk. All three studies were judged as being at low risk for selective outcome and/or analysis bias. Risk of other biases (e.g. funding source, balance/imbalance in important characteristics, inappropriate analysis) for one of the three studies was judged to be high.[131]

Methodological quality of systematic reviews comparing total hip replacement with resurfacing arthroplasty

The assessment of methodological quality of the three included systematic reviews[142–144] is presented in *Table 40* and the data extraction sheets (see *Appendices 3* and *4*). Given the number of methodological items that were satisfied, one of the three reviews was judged as being of high quality (falling into the score range 9–11),[143] one was judged as being of medium quality (falling into the score range 5–8)[142] and one was judged as being of low quality (falling into the score range 0–4).[144] The specific unmet methodological items related to inappropriate analysis, failure to address issues of publication bias and no information on conflicts of interest.

Clinical effectiveness findings for the comparison between total hip replacement and resurfacing arthroplasty

This section summarises the findings from the three RCTs[130–132] and three systematic reviews.[142–144]

The reported outcomes for this section were the HHS (one RCT;[130] two systematic reviews[142,143]), WOMAC score (two RCTs;[131,132] two systematic reviews[142,143]), Merle d'Aubigné and Postel score (one RCT;[132] one systematic review[142]), UCLA activity score (two RCTs;[131,132] one systematic review[42]), OHS (one RCT[130]), health-related quality of life scales (SF-36 and EQ-5D; two RCTs[130,131]), risk of revision (one RCT;[132] two systematic reviews[142,143]), mortality (two systematic reviews[142,143]), infection (two RCTs;[130,132] one systematic review[142]), aseptic loosening (one RCT;[132] two systematic reviews[142,143]), implant dislocation (two RCTs;[130,132] one systematic review[142]) and deep-vein thrombosis (two RCTs[130,132]).

TABLE 39 Risk of bias summary for RCTs: review authors' judgements about each risk of bias item – THR vs. RS

Study	Selection bias: random sequence generation	Selection bias: allocation concealment	Performance bias: subjective (e.g. patient reported)	Performance bias: objective (e.g. mortality, radiography, dislocation)	Detection bias: subjective (e.g. patient reported)	Detection bias: objective (e.g. mortality, radiography, dislocation)	Attrition bias: subjective (e.g. patient reported)	Attrition bias: objective (e.g. mortality, radiography, dislocation)	Reporting bias: selective reporting of the outcome, subgroups or analysis	Other bias [funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics]
Costa 2012[130]	+	+	–	+	+	+	+	+	+	+
Garbuz 2010[131]	?	+	+	NA	?	NA	+	NA	+	–
Vendittoli 2010[132]	+	+	–	+	?	+	–	+	+	+

–, high risk of bias; +, low risk of bias; ?, unclear risk of bias; ITT, intention to treat; NA, not applicable; PP, per protocol.

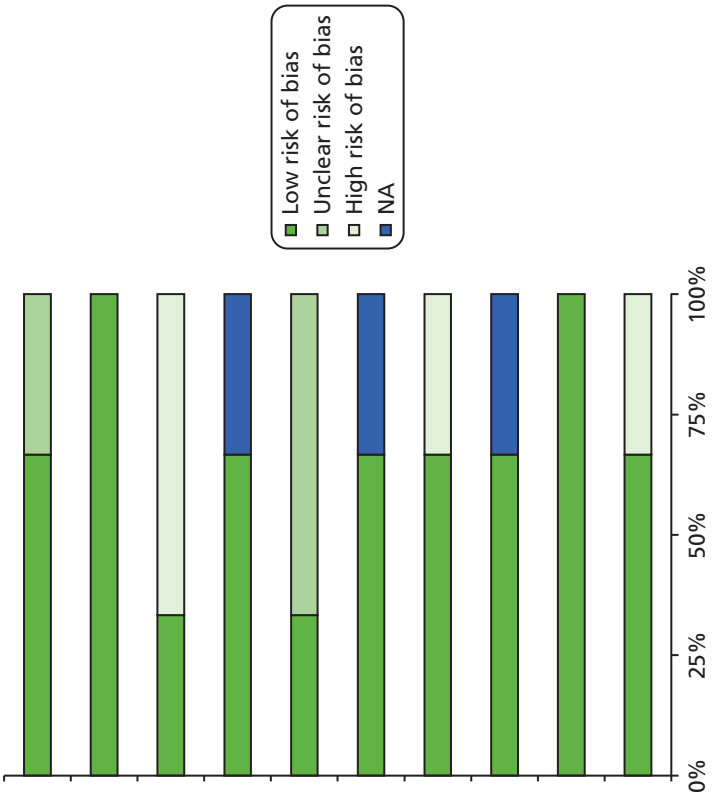


FIGURE 14 Risk of bias graph for RCTs: review authors' judgements about each risk of bias item – THR vs. RS. ITT, intention to treat; NA, not applicable; PP, per protocol.

TABLE 40 Methodological quality assessment summary for systematic reviews: THR vs. RS

Study	Was an a priori design provided?	Was there duplicate study selection and data extraction?	Was a comprehensive literature search performed?	Was the status of publication (i.e. grey literature) used as an inclusion criterion?	Was a list of studies (included and excluded) provided?	Were the characteristics of the included studies provided?	Was the scientific quality of the included studies assessed and documented?	Was the scientific quality of the included studies used appropriately in formulating conclusions?	Were the methods used to combine the findings of studies appropriate?	Was the likelihood of publication bias assessed?	Was the conflict of interest stated?	Overall
Jiang 2011[142]	Yes	Yes	Yes	No	Yes	No	Yes	CA	No	No	No	Medium quality
Smith 2010[143]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	CA	No	Yes	Yes	High quality
Springer 2009[144]	Yes	Yes	No	No	Yes	Yes	CA	No	No	No	No	Low quality

CA, cannot answer.

Neither the RCTs nor the systematic reviews reported any evidence for the following clinical effectiveness outcomes:

- HOOS
- LISOH
- AAOS Hip and Knee Questionnaire
- AIMS
- MACTAR
- NHP questionnaire
- SF-12
- time to revision
- pain score (VAS)
- femoral head penetration.

Summary results for the included outcomes are presented separately for RCTs and systematic reviews.

Evidence from randomised controlled trials

Functional/clinical measures

All three included RCTs comparing THR and RS reported some evidence for the following functional scores measured at 12–24 months after the procedure: HHS,[130] OHS,[130] WOMAC score,[131,132] UCLA activity score[131,132] and Merle d'Aubigné and Postel score.[132]

In two RCTs there was no difference between the THR group and the RS group in mean postoperative OHS (12 months; MD –2.23, 95% CI –5.98 to 1.52),[130] Merle d'Aubigné and Postel score (24 months; MD 0.0, 95% CI –1.06 to 1.06)[132] or WOMAC score (12 months; MD 2.20, 95% CI –1.57 to 5.97).[132] One of these RCTs showed a significantly improved mean WOMAC score for the RS group compared with the THR group at 24 months of follow-up; however, this difference was not deemed to be clinically important (MD 3.30, 95% CI 0.01 to 6.58).[132]

All three included RCTs comparing THR with RS reported some evidence for the following functional scores measured at 12–24 months after the procedure: HHS,[130] OHS,[130] WOMAC score,[131,132] UCLA score[131,132] and Merle d'Aubigné and Postel score.[132]

Health-related quality of life

Two RCTs reporting quality of life measures showed statistically non-significant differences between the THR group and the RS group for both the SF-36 ($p = 0.55$ and $p = 0.97$ for mental and physical components, respectively)[131] and the EQ-5D (MD –0.08, 95% CI –0.18 to 0.03).[130] These results were deemed to be inconclusive given the wide CI[130] and incomplete data reporting.[131]

Revision

The occurrence of implant revision was reported for only one RCT.[132] There was no statistically significant difference between the THR group and the RS group for risk of revision at 6 months (RR 1.01, 95% CI 0.06 to 15.92), 24 months (RR 0.50, 95% CI 0.04 to 5.48) or 56 months (RR 0.54, 95% CI 0.10 to 2.91) post surgery. The 95% CIs around the effect estimates embraced the value 1.00 and therefore did not allow definitive conclusions to be made regarding the effectiveness of THR compared with RS.

Mortality rate
No evidence on mortality rates was identified from the RCTs.

Complications
Evidence on complications was reported for two RCTs.[130,132] Meta-analysis of the data on risk of infection from the two RCTs indicated that, at 12–56 months post operation, THR recipients were at an increased risk of infection compared with RS recipients (pooled OR 7.94, 95% CI 1.78 to 35.40) (*Figure 15*).
In addition, evidence on the differences between groups for the risk of deep-vein thrombosis (*Figure 16*; pooled OR 0.60, 95% CI 0.15 to 2.42),[130,132] implant dislocation (*Figure 17*; pooled OR 3.97, 95% CI 0.79 to 19.90),[130,132] wound complications (RR 4.01, 95% CI 0.92 to 18.18)[130] and aseptic loosening (RR not estimable)[132] was judged to be inconclusive by us.

A summary of the results for the difference outcomes is presented in *Table 41*.

Evidence from systematic reviews

Functional/clinical measures
Two of the three included systematic reviews comparing THR with RS reported evidence on HHS,[142,143] WOMAC score,[142,143] Merle d'Aubigné and Postel score[142] and UCLA activity score[142] (*Table 42*).
The evidence was inconclusive because of the lack of pooled MD estimates for all four scores as well as the inconsistent results for the mean HHS and WOMAC score.

Health-related quality of life
No evidence was identified.

Revision
Both systematic reviews[142,143] found a higher risk of revision in patients receiving RS than in those receiving THR. One review meta-analysed data from four RCTs that compared risk of revision in RS and THR recipients, reporting a pooled RR estimate of 2.60 (95% CI 1.31 to 5.15) (see *Table 42*).[142]

Mortality
Overall, evidence on mortality reported by both systematic reviews[142,143] was inconclusive because of great uncertainty in the effect estimates and the variability around them. For example, the pooled RR for mortality in one review[143] for the comparison between RS and THR was 1.10 (95% CI 0.10 to 17.8) (see *Table 42*).

Failure rate
One systematic review[144] reported an indirect naive comparison analysis (i.e. analysis without a common comparator) based on data from 15 studies of RS and 19 studies of THR (see *Table 42*). The analysis suggested a reduced risk of failure in the RS recipients compared with the THR recipients (3.70% vs. 11.60%). Given the well-recognised problems with validity of such methodology, this evidence was judged to be inconclusive.

Study or subgroup	THR Events	THR Total	Hip RS Events	Hip RS Total	Weight	Peto's OR Peto, fixed, 95% CI
Costa *et al.* 2012[130]	2	66	0	60	28.8%	6.85 (0.42 to 111.12)
Vendittoli *et al.* 2010[132]	5	100	0	109	71.2%	8.42 (1.43 to 49.54)
Total (95% CI)		**166**		**169**	**100.0%**	**7.94 (1.78 to 35.40)**
Total events	7		0			
Heterogeneity: $\chi^2=0.02$, df=1 ($p=0.90$); $I^2=0\%$						
Test for overall effect: $z=2.72$ ($p=0.007$)						

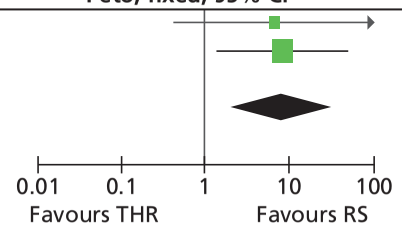


FIGURE 15 Risk of infection.

Study or subgroup	THR Events	THR Total	Hip RS Events	Hip RS Total	Weight	Peto's OR Peto, fixed, 95% CI
Costa *et al.* 2012[130]	0	66	4	60	49.7%	0.12 (0.02 to 0.85)
Vendittoli *et al.* 2010[132]	3	100	1	109	50.3%	3.02 (0.42 to 21.76)
Total (95% CI)		**166**		**169**	**100.0%**	**0.60 (0.15 to 2.42)**
Total events	3		5			
Heterogeneity: $\chi^2=5.19$, df=1 ($p=0.02$); $I^2=81\%$						
Test for overall effect: $z=0.72$ ($p=0.47$)						

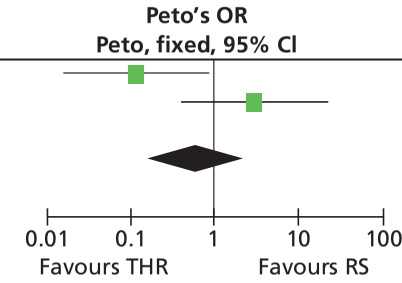


FIGURE 16 Risk of deep-vein thrombosis.

Study or subgroup	THR Events	THR Total	Hip RS Events	Hip RS Total	Weight	Peto's OR Peto, fixed, 95% CI
Costa *et al.* 2012[130]	1	66	1	60	33.5%	0.91 (0.06 to 14.73)
Vendittoli *et al.* 2010[132]	4	100	0	109	66.5%	8.34 (1.16 to 60.14)
Total (95% CI)		**166**		**169**	**100.0%**	**3.97 (0.79 to 19.90)**
Total events	5		1			
Heterogeneity: $\chi^2=1.62$, df=1 ($p=0.20$); $I^2=38\%$						
Test for overall effect: $z=1.68$ ($p=0.09$)						

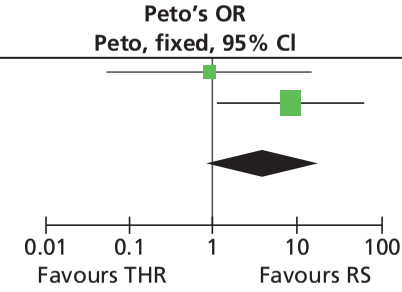


FIGURE 17 Risk of implant dislocation.

TABLE 41 Summary of the results for THR vs. RS: RCTs

Follow-up (months)	Arm-specific estimates, *n*/*N* or mean (SD or 95% CI) (THR vs. RS)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
HHS (range 0–100)				
12	82.3 (77.2 to 87.5) vs. 88.4 (84.4 to 92.4)[130]	MD –6.04 (12.58 to 0.51)	1 (low)	Inconclusive
OHS (range 0–48)				
12	38.2 (35.3 to 41.0) vs. 40.4 (37.9 to 42.9)[130]	MD –2.23 (–5.98 to 1.52)	1 (low)	No difference
WOMAC score (range 0–100)				
3	[c]19.2 (NR) vs. 19.9 (NR)[132]	$p = 0.76$ (NS)[132,136]	2 (unclear)	No difference
6	[c]11.3 (NR) vs. 13.9 (NR)[132]	$p = 0.20$ (NS)[132,136]		
12	[c]10.2 (10.7) vs. 8.0 (13.2)[132]	[d]MD 2.20 (–1.57 to 5.97)[132,136]		
12	[e]90.18 (NR) vs. 90.40 (NR)[131]	$p = 0.95$ (NS)[131]		
24	[c]9.0 (11.9) vs. 5.7 (8.6)[132]	[d]MD 3.30 (0.01 to 6.58)[132,136]		
Merle d'Aubigné and Postel score (range 0–18)				
3	15.8 (NR) vs. 16.2 (NR)[132]	$p = 0.59$ (NS)	1 (unclear)	No difference
6	17.1 (NR) vs. 17.2 (NR)[132]	$p = 0.72$ (NS)		
12	16.6 (NR) vs. 16.7 (NR)[132]	$p = 0.94$ (NS)		
24	17.5 (1.3) vs. 17.5 (1.3)[132]	$p = 0.94$ (NS); MD 0.0 (–1.06 to 1.06)[c]		
UCLA activity score (range 1–10)				
12	6.3 (NR) vs. 6.8 (NR)[131]	$p = 0.24$ (NS)[131]	2 (unclear)	Inconclusive
12	6.3 (NR) vs. 7.1 (NR)[132]	$p = 0.03$ (SS)[132,136]		
24	NR (NR) vs. NR (NR)[132]	$p = 0.09$ (NS)[132,136]		
SF-36 score (range 0–100)				
12	MCS: 55.13 (NR) vs. 53.87 (NR)[131]	$p = 0.55$ (NS)	1 (unclear)	Inconclusive
12	PCS: 51.28 (NR) vs. 51.22 (NR)[131]	$p = 0.97$ (NS)		
EQ-5D score (range 0–1)				
12	0.71 (0.63 to 0.80) vs. 0.79 (0.72 to 0.87)[130]	MD –0.077 (–0.188 to 0.034)	1 (low)	Inconclusive
Revision rate				
3	1/102 vs. 0/103[132]	$p =$ NR; RR and 95% CI not estimated	1 (low)	Inconclusive
6	1/102 vs. 1/103[132]	$p =$ NR; RR 1.01 (0.06 to 15.92)[d]		
12	1/102 vs. 2/103[132]	$p =$ NR; RR 0.50 (0.04 to 5.48)[d]		
24	1/102 vs. 2/103[132]	$p =$ NR; RR 0.50 (0.04 to 5.48)[d]		
56	2/100 vs. 4/109[132]	$p = 0.47$ (NS); RR 0.54 (0.10 to 2.91)[d]		

TABLE 41 Summary of the results for THR vs. RS: RCTs (*continued*)

Follow-up (months)	Arm-specific estimates, *n*/*N* or mean (SD or 95% CI) (THR vs. RS)	Difference (*p*-value or 95% CI)	Number of RCTs (SROB across studies)[a]	Treatment effect conclusion[b]
Complications				
Infection				
12	2/66 vs. 0/60[130]	$p = 0.49$ (NS); RR and 95% CI not estimated	2 (low)	In favour of RS
56	5/100 vs. 0/109[132]	$p = 0.02$ (SS); RR and 95% CI not estimated		
		[d]Pooled estimate of Peto OR: 7.94 (1.78 to 35.40)[130,132]		
Deep-vein thrombosis				
12	0/66 vs. 4/60[130]	$p = 0.05$ (NS); RR and 95% CI not estimated	2 (low)	Inconclusive
56	3/100 vs. 1/109[132]	$p =$ NR (NS); RR 3.27 (95% CI 0.30 to 30.90)[d]		
		[d]Pooled estimate of Peto OR: 0.60 (95% CI 0.15 to 2.42)[130,132]		
Implant dislocation				
12	1/66 vs. 1/60[130]	$p = 1.00$ (NS); RR 0.90, 95% CI 0.05 to 14.21[d]	2 (low)	Inconclusive
56	4/100 vs. 0/109[132]	$p = 0.038$ (SS); RR and 95% CI not estimated		
		[d]Pooled estimate of Peto OR: 3.97 (95% CI 0.79 to 19.90)[130,132]		
Superficial wound complication				
12	9/66 vs. 2/60[130]	$p = 0.06$ (NS); RR 4.01 (95% CI 0.92 to 18.18)[d]	1 (low)	Inconclusive
Aseptic loosening				
56	0/100 vs. 6/109[132]	$p = 0.017$ (SS); RR and 95% CI not estimated	1 (low)	Inconclusive

MCS, mental component summary score; NR, not reported; NS, not significant; PCS, physical component summary score; SROB, summary risk of bias; SS, statistically significant.
a Decision was consensus-based.
b Favours THR (or RS), no difference or inconclusive.
c Mean change in WOMAC score from baseline.
d Calculated.
e End-point mean WOMAC score.
Only those outcomes listed for which any evidence was reported; studies not reporting a given outcome are not listed.

TABLE 42 Summary results for RS vs. THR: systematic reviews

Follow-up (years)	Pooled effect estimate (95% CI) (RS vs. THR)	Number of RCTs in the MA or narrative synthesis	AMSTAR rating	Treatment effect conclusion[a]
HHS (range 0–100)				
1–2	NR;[142] no significant difference	3[142]	Medium quality[142]	Inconclusive
2	MD 2.51 (1.24 to 3.77) (SS);[143] better for RS than for THR	NR[143]	High quality[143]	
WOMAC score (range 0–100)				
1–2	NR;[142] no significant difference	3[142]	Medium quality[142]	Inconclusive
2	MD –2.41 (–3.88 to –0.94) (SS);[143] better for RS than for THR	NR[143]	High quality[143]	
Merle d'Aubigné and Postel score (range 0–18)				
1–2	NR;[142] no significant difference	3[142]	Medium quality[142]	Inconclusive
UCLA activity score (range 1–10)				
1–2	NR;[142] mean UCLA activity score significantly higher in RS group than in THR group	2[142]	Medium quality[142]	Inconclusive
Revision rate				
1–10	RR 2.60 (1.31 to 5.15) (SS)[142]	4[142]	Medium quality[142]	In favour of THR
NR	RR 1.72 (1.20 to 2.45) (SS);[143] higher in RS group than in THR group (19 pooled RCTs and non-RCTs)	NR[143]	High quality[143]	
Mortality rate				
3	NR;[142] one study showed no significant difference between RS and THR RR 1.05 (0.24 to 4.66)	1[142]	Medium quality[142]	Inconclusive
NR	RR 1.10 (0.10 to 17.8) (NS)[143]	NR[143]	High quality[143]	
Failure rate				
NR	3.70% (2.0% to 6.5%) vs. 11.60% (7.50% to 17.40%);[144] indirect naive comparison of 15 studies of RS and 19 studies of THR	NA[144]	Low quality[144]	Inconclusive
Dislocation rate				
1–2	RR 0.25 (0.05 to 1.21) (NS)[142]	3[142]	Medium quality[142]	In favour of RS
NR	RR 0.20 (0.10 to 0.50) (SS);[143] lower in RS group than in THR group (no. of pooled studies NR)	NR[143]	High quality[143]	
Component loosening				
1–10	RR 4.96 (1.82 to 13.50) (SS);[142] higher in RS group than in THR group	4[142]	Medium quality[142]	In favour of THR
NR	RR 3.00 (1.11 to 8.50) (SS);[143] higher in RS group than in THR group (10 pooled RCTs and non-RCTs)	NR[143]	High quality[143]	
Infection				
1–3	RR 2.25 (0.61 to 8.31) (NS)[142]	3[142]	Medium quality[142]	Inconclusive

MA, meta-analysis; NR, not reported; NS, not significant; SS, statistically significant.
a Favours THR (or RS), no difference or inconclusive.
Only those reviews listed for which any evidence for the given outcome was reported.

Complications

Evidence on complications was reported by both systematic reviews[142,143] (i.e. implant dislocation, infection and component loosening) (see *Table 42*). The evidence consistently showed an increased risk for component loosening[142,143] but a reduced risk for implant dislocation[142] among RS recipients compared with THR recipients. One review,[142] which provided the risk of infection pooled across three studies, was not informative enough to draw any conclusions (RR 2.25, 95% CI 0.61 to 8.31).

Grading the overall quality of the evidence

The results for graded outcomes are presented in the evidence profile (*Table 43*). The overall quality for gradable outcomes across the reviewed evidence comparing THR with RS was as follows: HHS – very low grade; WOMAC score – low grade; revision – very low grade; mortality – not graded because of absence of evidence; and implant dislocation – very low grade.

Summary conclusions for the comparison between total hip replacement and resurfacing arthroplasty

The majority of the evidence from three RCTs[130–132] (*Table 44*) and three systematic reviews[142–144] (*Table 45*) comparing THR and RS was rated as inconclusive (RCTs – very low-grade evidence). Nevertheless, the evidence from two RCTs and two systematic reviews indicated a reduced risk of infection[130,132] and implant dislocation[142,143] among RS patients compared with THR patients. However, the evidence from the same reviews also indicated that recipients of RS were at higher risk of revision and component loosening than patients who received a THR. In three RCTs[130–132] the mean postoperative OHS, WOMAC score (low-grade evidence) and Merle d'Aubigné and Postel score were not different between patients who received THR and those who received RS.

TABLE 43 Grading of Recommendations, Assessment, Development and Evaluation evidence profile for gradable outcomes reported in RCTs of THR vs. RS[a]

Outcome (follow-up timing)	Number of studies reporting outcome (participants)	Pooled effect estimate (95% CI) and conclusion	SROB across studies	Consistency	Directness	Precision	Outcome reporting bias	Quality of the evidence (GRADE)[b]
HHS (12 months)	1 (126)[130]	None; inconclusive	Low	NA	Direct	Imprecise	Likely	Very low
WOMAC score (3–24 months)	2 (313)[131,132]	None; no difference	Unclear	Consistent	Direct	Precise	Likely	Low
Revision (3–56 months)	1 (209)[132]	None; inconclusive	Low	NA	Direct	Imprecise	Likely	Very low
Mortality (NA)	0	NA	NA	NA	NA	NA	NA	NA (no evidence)
Implant dislocation (12–56 months)	2 (335)[130,132]	OR 3.97 (0.79 to 19.90); inconclusive	Low	Inconsistent	Direct	Imprecise	Likely	Very low

NA, not applicable; SROB, summary risk of bias.
a Adapted from Guyatt *et al.*[102]
b GRADE categories: high, moderate, low, very low, NA (no evidence).

TABLE 44 Summary of evidence regarding the differences between THR and RS for each reported outcome in the RCTs[130–132]

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Infection[130,132] (in favour of RS)	OHS,[130] WOMAC score (low-grade evidence),[131,132] Merle d'Aubigné and Postel score[132]	HHS (very low-grade evidence)[130] UCLA activity score,[131,132] SF-36,[131] EQ-5D,[130] revision (very low-grade evidence),[132] mortality (no evidence; not graded), deep-vein thrombosis,[130,132] implant dislocation (very low-grade evidence),[130,132] superficial wound complications,[130] aseptic loosening[132]

TABLE 45 Summary of evidence regarding the differences between THR and RS for each reported outcome in the systematic reviews[142–144]

Conclusive evidence suggesting difference	Conclusive evidence suggesting no difference	Inconclusive evidence
Revision[142,143] (in favour of THR), implant dislocation[142,143] (in favour of RS), component loosening[142,143] (in favour of THR)	None	HHS,[142,143] WOMAC score,[142,143] Merle d'Aubigné and Postel score,[142] UCLA activity score,[142] mortality,[142,143] failure,[144] infection[142]

There was inconclusive evidence on mortality (three RCTs[130–132] and two systematic reviews[142,143]), HHS (one RCT[130] and two systematic reviews[142,143]), UCLA activity score (two RCTs[131,132] and one systematic review[142]) and selected complications (i.e. infection, wound complication, deep-vein thrombosis; two RCTs[130,132] and one systematic review[142]).

Results from individual RCTs were considered inconclusive because of the partial reporting (missing data for effect estimates, CIs, SEs, SDs, *p*-values) and great uncertainty in the estimates (wide CIs). The findings from the systematic reviews were inconclusive because of great uncertainty in the pooled estimates (wide CIs), lack of reporting of pooled results across RCTs (i.e. only narrative synthesis reported) or inconsistent summary findings.

Other analysis

Publication bias

The extent to which publication bias could have influenced the pooled treatment effect estimates (i.e. degree of funnel plot asymmetry) could not be explored because of insufficient numbers of data points in the forest/funnel plots.

Heterogeneity, subgroup effects and sensitivity analysis

The reviewed data from RCTs were too sparse (only three RCTs) to allow an exploration of whether or not the effect of any given THR relative to RS differed by study-level methodological (i.e. risk of bias, type of data analysis) or patient-related (i.e. age, sex or functional status) characteristics. None of the included RCTs reported within-study subgroup effects of the THR relative to RS (or vice versa).

Overall summary of the clinical effectiveness findings

A large proportion of evidence appraised and summarised in this review has been judged to be inconclusive (very low to low grade) because of poor reporting, missing data, inconsistent results and/or great uncertainty in the treatment effect estimates. Nevertheless, results from most studies suggested significantly improved post-surgery scores for functional/clinical measures (HHS, OHS, WOMAC score, MACTAR score, Merle d'Aubigné and Postel score and SF-12 score), regardless of the type of THR or RS received. Some moderate- or lower-grade evidence indicated no difference for these measures between different types of THR (or between THR and RS) at different follow-up times. There was a reduced risk of implant dislocation for participants receiving a THR with a larger femoral head size (vs. a smaller head size) or with a cemented cup (vs. cementless; high-grade evidence). Moreover, the evidence suggested a reduced femoral head penetration rate (moderate grade evidence) and risk of implant revision (very low-grade evidence) for participants who received cross-linked polyethylene compared with conventional polyethylene cup liner bearings. Participants with ceramic-on-ceramic articulations (vs. metal-on-polyethylene articulations) experienced a reduced risk of osteolysis. Recipients of RS had a lower risk of infection than recipients of a THR. The evidence on mortality and other complications (e.g. loosening, femoral fracture and deep-vein thrombosis) was inconclusive (very low grade).

Limitations of the reviewed evidence and pitfalls in interpretation

The review findings warrant cautious interpretation given the limitations of the available evidence. Specifically, great uncertainty in the treatment effect estimates (i.e. wide 95% CIs) because of limited sample sizes and/or small numbers of events (especially for deaths, revisions and complications), as well as incomplete or poor reporting (e.g. missing effect measures, SDs/SEs, 95% CIs, *p*-values), rendered some of the reviewed evidence inconclusive. Moreover, reported evidence on complications was scarce. It is unclear whether this is because of the absence or rarity of these events or because of under-reporting. In light of poor reporting, it was not possible to explore contextual factors that might have influenced the study results. For example, lack of blinding of participants and study personnel may have led to systematic differences in caregiving or co-interventions across implant groups, which would independently influence outcome measures. Furthermore, none of the studies reported the between-group distribution of experience and skills of study personnel, including surgeons, physicians, physiotherapists and occupational therapists. Any imbalance between the study treatment groups in the above-mentioned factors would influence the participants' prognosis apart from treatment.

The paucity of data did not allow the exploration of any variation in the treatment effect across the predefined subgroups of patients or methodological features of studies; likewise, the extent of publication bias could not be examined using funnel plots because of the small numbers of studies in the meta-analyses.

Scenario analysis around revision rates

We did not feel that it would be appropriate to use data from other clinical trials/registries to check our findings from the economic modelling because the clinical effectiveness studies that we identified concerned with revision rates were based on low counts and/or on small trials with a great deal of uncertainty. Overall, across the THR/THR and THR/RS comparisons, trials were often based on selective populations or interventions and provided data on revision rates that were inconclusive with often wide CIs.

Comparison of the results from randomised controlled trials and systematic reviews

The findings of the RCTs and systematic reviews could be compared only with regard to implant fixation methods (cemented vs. cementless) and femoral head-on-cup articulations (e.g. metal-on-metal vs. metal-on-polyethylene, ceramic-on-ceramic vs. metal-on-polyethylene, ceramic-on-ceramic vs. ceramic-on-polyethylene). In summary, the effect estimates for differences between the above-mentioned THR groups in risk of revision, mortality and complications reported in RCTs and systematic reviews were statistically non-significant and had wide uninformative CIs around them. Therefore, the evidence from both RCTs and systematic reviews was rendered as inconclusive because of the wide variability around the estimates and/or missing data. The reviewed evidence from RCTs suggested that there was no difference in postoperative HHS between cemented and cementless THR groups. The evidence for HHS reported in the included systematic reviews was ruled as inconclusive.

Our update search identified four new relevant systematic reviews.[242–245] Of these four systematic reviews, three compared the effectiveness of THRs using different articulations (metal-on-metal vs. metal-on-polyethylene),[242] implant fixation methods (cemented vs. cementless)[245] or femoral stem coating materials (hydroxyapatite coated vs. non-hydroxyapatite coated)[244] for risk of revision,[245] HHS,[242,244,245] mortality[245] and complications.[242,245] The remaining systematic review compared THR with RS for risk of revision.[243]

Briefly, the review by Voleti *et al.*[242] presented a meta-analysis based on three RCTs and found no significant difference in HHS between the two articulations (metal-on-metal vs. metal-on-polyethylene) at 6 years post-surgery follow-up (pooled MD –1.05; $p = 0.37$). However, the risk of complications (dislocation, aseptic loosening, trochanteric/iliopsoas bursitis, femoral fracture and wound dehiscence) was greater in the metal-on-metal articulation group than in the metal-on-polyethylene articulation group (OR 3.37, 95% CI 1.57 to 7.26).[242] Similarly, another review[245] presented a meta-analysis of seven RCTs showing a statistically non-significant difference in the mean postoperative HHS between the cemented and the cementless THR groups (pooled MD 1.12, 95% CI –1.17 to 3.41). In the same review, the meta-analytic estimates for risk of revision (six RCTs; pooled RR 1.44, 95% CI 0.88 to 2.36), mortality (five RCTs; pooled RR 1.06, 95% CI 0.73 to 1.52) and complications (four RCTs; pooled RR 1.54, 95% CI 0.21 to 11.03) between the cemented and the cementless groups of THR were also statistically non-significant. In the review by Li *et al.*,[244] the postoperative pooled mean HHS was not statistically significantly different between the hydroxyapatite-coated and the non-hydroxyapatite-coated THR groups (four RCTs; pooled MD 3.04, 95% CI –4.47 to 10.54). The review by Pailhe *et al.*[243] included a qualitative synthesis of three RCTs and eight non-RCTs, providing no definitive conclusions regarding the differences between THR and RS in terms of implant survival or risk of revision.

In summary, the findings from the newly identified systematic reviews[242–245] are in agreement with those of this review in showing no difference in postoperative HHS between the cemented and the cementless THR groups. Also in agreement with our findings, the pooled estimates for revision, mortality and complications were statistically non-significantly different between the groups, with sufficiently wide 95% CIs (because of low event counts and the small sample size of trials) that were compatible with a moderate-to-large effect size in either direction, rendering these findings inconclusive.[245] Future well-designed RCTs need to corroborate or refute the finding of one systematic review[242] which suggests that there is an increased risk of complications in the metal-on-metal articulation group compared with the metal-on-polyethylene articulation group.

Strengths and limitations of the review

One of the strengths of this review is the fact that the reviewers used systematic and independent strategies to minimise bias in searching, identifying, selecting, extracting and appraising the relevant evidence. The search strategy was applied to multiple electronic sources. Apart from the limitations of the evidence itself, the scope of this review was limited to a predefined set of outcomes ascertained from recently published evidence (2008 or later); evidence from non-English publications was not included. Given the wide scope and large amount of evidence identified, we limited inclusion to studies with a

sample size of at least 100 that were published since 2008. The rationale for the size limitation was that smaller studies tend to be underpowered to detect meaningful differences in outcomes.[244,245] The results of such studies are usually rendered inconclusive because of statistically non-significant estimates with wide CIs that include large treatment effect size values compatible with both a better and a worse outcome for any given treatment compared with the control treatment. Therefore, to minimise this problem we calculated the minimum sample size for a study that would have 90% power at a two-tailed test significance level of 0.05 to detect a MD of 10 on the HHS (we selected a SD of 15 based on external sources).[107,246] This calculation yielded a total sample size of 100 participants.

Future research

Because the evidence for any given comparison of two types of THR was sparse (maximum of two trials), the observed findings need to be replicated in larger, long-term pragmatic trials comparing the same THRs with each other (or with RS) before more definitive conclusions or recommendation are made. Large, multicentre, long-term pragmatic trials would help to reliably evaluate relative treatment effects and their variation(s) across patients, as well as manufacturer-based subgroups, and maximise generalisability of the findings to larger populations in clinical practice settings. For a more complete picture to aid health-care policy decisions, trials are also needed to investigate the cost-effectiveness of alternative THR (or RS) techniques. Study authors are encouraged to specify MCIDs and power calculations for their primary outcome(s). This information would help in the interpretation of the study findings in both clinical and statistical terms. Better reporting of future trial results is also warranted.

Methods for the review of cost-effectiveness

Identification of studies

Initial scoping searches were undertaken in MEDLINE in October 2012 to assess the volume and type of literature relating to the assessment question. These scoping searches also informed development of the final search strategies (see *Appendix 1*). An iterative procedure was used to develop these strategies with input from clinical advisors and previous HTA reports (e.g. Vale *et al.*,[19] de Verteuil *et al.*[11]). The strategies have been designed to capture generic terms for arthritis, THR and RS. Searches were limited by the addition of economic and quality of life terms, which were selected with reference to previous research.[247,248]

Searches were date limited from 2002 (the date of the most recent NICE guidance in this area[25]). The searches were undertaken in November 2012 (for exact search dates see *Appendix 1*).

All bibliographic records identified through the electronic searches were collected in a managed reference database.

The following main sources were searched to allow for identification of relevant published and unpublished studies and studies in progress:

- electronic bibliographic databases, including research in progress
- references of included studies.

The following databases of published studies were searched: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, Science Citation Index and Conference Proceedings Citation Index – Science, The Cochrane Library (specifically CDSR, CENTRAL, DARE, NHS EED and HTA database) and the Cost-effectiveness Analysis Registry (CEA Registry) (Articles).

The following databases of research in progress were searched: Current Controlled Trials, ClinicalTrials.gov, UKCRN Portfolio Database and NLM Gateway (HSRProj).

The reference lists of included studies were checked for additional studies.

Inclusion and exclusion criteria

The following inclusion and exclusion criteria were used to identify eligible studies reporting costs and/or effects of THR and RS useful for the economic model and decision analysis:

Inclusion criteria

Study design

- RCTs.
- Observational designs, cohort studies and registry-based studies.
- Decision-analytic modelling studies.
- Systematic reviews.
- Meta-analyses.

Population

- People with pain or disability resulting from end-stage arthritis of the hip for whom non-surgical management has failed.

Intervention

- Elective primary THR.
- Primary hip RS.

Comparator

- Different types of primary THR compared with RS for people in whom both procedures are suitable.
- Different types of primary THR compared with each other for people who are not suitable for hip RS.
- Studies reporting costs or utilities without a comparator were also included.

Record

- Full-text articles of completed or in-progress studies (protocols) published in English.

Outcomes

- Cost-effectiveness outcomes were costs (cost of resources/devices, quantitative use of resources reported) and clinical effectiveness measures or utility measures (utility, EQ-5D score or QALYs), incremental cost-effectiveness ratios (ICERs), uncertainty measures, the ceiling willingness-to-pay (WTP) ratios and probabilities of cost-effectiveness from cost-effectiveness acceptability curves (CEACs).

Exclusion criteria

- Non-English-language publications.
- Abstract/conference proceedings, letters and commentaries.
- Quality of life reported without utilities or QALYs.
- Hip/knee data not reported separately.
- Studies including only patients aged < 35 years.

Assessment of eligibility

All retrieved records were collected in a specialist database and duplicate records were identified and removed. An initial sift was undertaken by one reviewer to exclude clearly non-relevant records using the following exclusion criteria:

- non-hip only papers
- papers on animals
- papers on children
- papers on surgery for hip fracture only
- non-English full-text papers.

This was followed by a formal sift by title and abstract by two reviewers using the inclusion/exclusion criteria. All identified relevant studies were read in full by two reviewers to identify eligible studies. Disagreement was resolved by a third reviewer. Reasons for exclusion of full-text papers were documented. The study flow was documented using a PRISMA diagram.[99]

Data extraction

Data extraction was carried out in two stages by one reviewer using the data extraction sheets (see *Appendix 4*) and was checked by a second reviewer. Stage one considered all eligible studies and stage two considered studies assessed for usefulness for populating the economic model and decision analysis. Data extracted during stage one included the following:

- study characteristics [i.e. author names, country, design, study aim, type of economic evaluation (cost-effectiveness analysis, cost–utility analysis), perspective (e.g. societal, health-care payer, patient) and study currency]
- patient characteristics (i.e. number of participants, age, sex, OA)
- outcomes [i.e. utilities, resources use and costs (both direct and indirect), ICERs]

Data extraction also included the overall study conclusion and a comment on the type of data included in the study that are relevant for the economic model. Studies were subsequently categorised by topic (THR or RS) and outcomes (costs or utilities) and cost studies were also ordered by year and date using the following hierarchy:

1. UK study published in 2008 or later
2. UK study UK study published before 2008
3. non-UK study published in 2008 or later
4. non-UK study UK study published before 2008.

Utility studies were ordered by study size and 'patient-reported utility data' (utilities derived prospectively using patient questionnaires or from databases that prospectively collected utilities) using the following hierarchy:

1. > 100 THR/RS patients and primary data
2. < 100 THR/RS patients and primary data
3. > 100 THR/RS patients and secondary data
4. < 100 THR/RS patients and secondary data.

Data extracted during the second stage considered the costs of THR (cost of the device, cost of surgical time/hospital stay), follow-up for successful THR, revision THR, follow-up for successful revision THR, RS (cost of the device, cost of surgical time/hospital stay), follow-up for successful RS, revision RS, follow-up for successful revision RS and utilities at baseline, post surgery up to 12 months and > 12 months post surgery. Information on definition of costs, source of costs, cost year and currency was also extracted.

Quality assessment

The key cost-effectiveness papers that were identified as relevant for the economic model were assessed by one reviewer and checked by a second reviewer using the Consensus on Health Economic Criteria (CHEC);[249] cost-effectiveness studies with decision-analytic models were also assessed using the criteria of Philips *et al.*[250]

Results of the review of cost-effectiveness

Identification of studies

The flow chart outlining the process of identifying relevant literature can be found in *Figure 18*. The database search identified 1650 records, with an additional 14 records identified through screening of reference lists of included studies. Removal of duplicates left 913 studies to be screened for inclusion. The initial sift excluded 283 studies that were clearly not relevant, with a further 525 records excluded on title and abstract ($\kappa = 0.89$). The remaining 105 full-text articles were assessed for eligibility, of which 35 were excluded with reasons (see *Appendix 13*). This resulted in a total of 70 eligible articles,[8,11,19,37,38,40,43,44,120,130,148,208,251–308] in which 66 studies were reported and subsequently included in the review. Of these, 35 were observational studies with or without an economic analysis,[37,208,251,252,254,255,258,264,266–268,270–272,274,276,277,279–282,284–287,290,294,295,297,298,300–302,305,306,308] 22 were economic analyses[11,19,38,44,148,253,256,257,259–262,269,273,275,278,288,291–293,299,304,307] including three HTAs,[11,19,148,299] four were

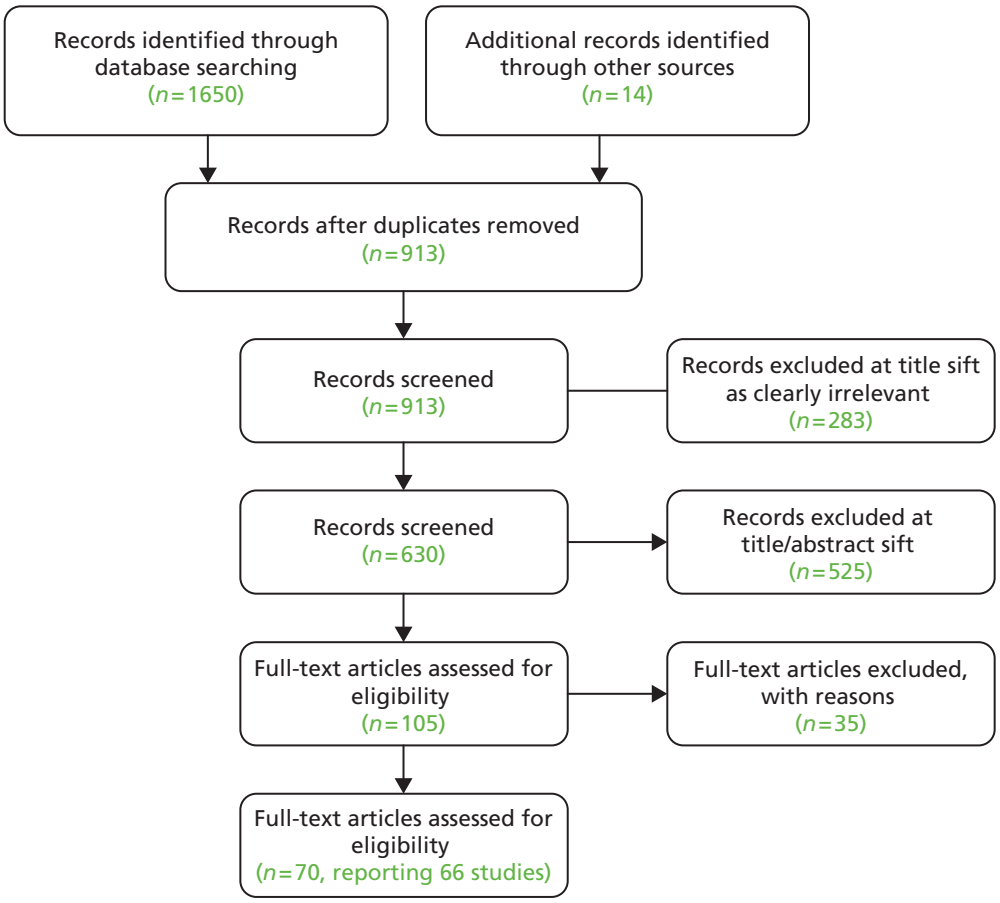


FIGURE 18 Flow diagram of study identification for the cost-effectiveness review.

reviews[8,43,289,296] (three non-systematic[43,289,296] and one systematic[8]), four were RCTs[40,120,130,263,283,303] and one was a before-and-after trial.[265] Study location covered the UK ($n = 13$[8,11,19,37,38,40,43,44,130,251,252,257,292,295,299,304]), other European countries ($n = 22$[256,258,260,261,263,265,266,271,274–276,278,280,281,283,287,288,297,298,302,303,305,306]), North America ($n = 21$[120,148,208,253–255,259,262,267,268,277,284,285,289,291,293,294,296,300,301,307]), Australia and New Zealand ($n = 6$[262,267,271,280,284,288]) and Asia ($n = 4$[270,272,279,308]). Costs/resource use were reported by 30 studies,[43,148,254,256,261,264,267,268,270,271,273,274,276–280,283,285–293,300,304,308] utilities/QALYs by 15 studies[122,251,252,258,266,272,284,294–298,301,302,305,306] and both costs/resource use and utilities/QALYs by 21 studies.[8,11,19,37,38,40,44,130,208,253,255,257,259,260,262,263,265,269,275,281,282,299,303,307] Seven of the 14 economic models reported transition probabilities.[8,11,19,253,259,261,275,299]

A separate search (December 2012) of the ClinicalTrials.gov, Current Controlled Trials, UKCRN Portfolio and HSRProj Databases retrieved 511 potential trials or health services research projects. After screening titles and full records (if available), eight clinical trials were identified as potentially relevant from the cost-effectiveness point of view (see *Appendix 7*). All were either ongoing or completed since 2009.

Description of included studies

Resurfacing arthroplasty

Evidence on RS was scarce with only five of the 66 included studies investigating hip RS (see *Appendix 10*). A 2012 UK RCT including 126 OA patients suitable for RS investigated the cost-effectiveness of RS compared with THR.[40,130] At the end of this 12-month trial small benefits of RS in terms of QALYs could be shown for a selected patient group, resulting in an ICER of £17,451 per QALY. This evidence was stronger for male than for female patients. In a comparison between ceramic-on-ceramic THR and RS at 3 months post surgery, evidence was not as strong, favouring THR over RS.[208] However, longer-term follow-up in a study comparing hybrid THR with RS confirmed that, after 5 and 9 years, the revision rates for RS were lower than for hybrid THR (9.3% and 16.7% at 9 years post surgery, respectively) and patients were more active.[251,252]

A retrospective economic decision analysis of published data over a 30-year time horizon showed the cost-effectiveness of RS compared with THR for women aged < 55 years and men aged < 65 years.[253] The main drivers of cost-effectiveness were the cost of the implant and length of hospital stay.[40,208] However, Vale *et al.*[19] reported in their HTA that RS would be cost-effective compared with THR only if RS revision rates could be shown to be 80–88% lower than revision rates for THR. They further concluded that RS could be cost-effective compared with 'watchful waiting' followed by THR or an extended period of 'watchful waiting' over 20 years.

Total hip replacement

The majority of studies investigated THR ($n = 61$) (see *Appendix 10*). Of these, five compared minimally invasive techniques with standard THR, reporting perioperative advantages, better short-term outcomes and reduced costs in favour of minimally invasive techniques.[11,148,254–256] However, Coyle *et al.*[148] concluded that there is little evidence of a difference between the two surgical techniques in the long term, mainly because of lack of data.

Ten of the THR studies focused on the comparison of different types of THR or specific components/brands of THR. Briggs *et al.*,[38] Davies *et al.*,[43] Fordham *et al.*[257] and Hulleberg *et al.*[258] assessed different brands of THR, Bozic *et al.*[259] investigated alternative bearings including metal-on metal, ceramic-on-ceramic and ceramic-on-polyethylene and Laupacis *et al.*,[120] Marinelli *et al.*,[260] Pennington *et al.*[44] and di Tanna *et al.*[261] compared cemented, cementless and hybrid THR more generally and reported inconsistent findings. The most recent economic model by Pennington *et al.*[44] used PROMs and showed that (1) cemented prostheses were the least costly type for THR, (2) hybrid prostheses were the most cost-effective and (3) cementless prostheses did not provide sufficient improvement in health outcomes to justify their additional costs. Similarly, Davies *et al.*[43] identified cemented prostheses as the least costly type of prosthesis in their review. However, they concluded that there is a lack of observed long-term prosthesis

survival data and particularly limited up-to date evidence for the UK, which led them to call for more trials with longer-term follow-up. Cummins *et al.*[262] reported that use of antibiotic-impregnated bone cement can result in an overall decrease in costs. For more detail on the studies investigating the different types of THR see *Appendix 12*.

Patient management and rehabilitation was the focus of four studies,[263–266] which reported that perioperative management and rehabilitation programmes could improve patient outcomes and reduce costs.

The majority of the THR studies (30/61[8,37,267–285,298,300–302,304–308]) assessed the costs and/or effectiveness of THR without a specific focus on a rehabilitation programme, surgical intervention, implant brand or prosthesis type. Of these, two US studies[267,268] concentrated on obese patients and reported that, even though operative costs are higher for obese patients, overall care costs and in-hospital outcomes for THR are comparable across all BMI groups. Eleven studies[269–279] evaluated the cost-effectiveness of THR in a specific country, and two multicentre studies[280,281] aimed to assess the costs and outcomes of THR comparatively across a number of European member states. These two studies concluded that improvement after surgery is associated with high preoperative expectations. Stargardt *et al.*[280] reported further that the total cost of treatment ranged from €1290 (Hungary) to €8739 (the Netherlands) and that the two main cost drivers were the cost of the implants and ward costs.

The overall findings of the cost-effectiveness studies were that (1) THR resulted in greater benefits than conservative treatment and (2) longer waiting times incurred greater costs and resulted in physical deterioration.[271,282,283] Further, agreement was reached on the long-term cost-effectiveness and sustained benefits of THR.[37,120,257,273,275] However, Bozic *et al.*[284] stated that, although THR improved quality of life, failed THR could lead to health states worse than chronic OA. Resource use might be increased as patients with a THR were shown to have a 10% increase in hospital stay compared with patients pre surgery.[285]

In contrast, two studies[286,287] that took a patient perspective rather than a health-care perspective concluded that out-of-pocket costs (including hospital costs, medication costs, rehabilitation costs, costs of health professional visits, costs of tests, costs of special equipment, costs of household alterations, use of private and community services and transportation costs that are not paid for by the health system), as well as use of health services, fell dramatically in the first year post surgery and that costs as well as resource use depended on pre-surgery health status.

Studies that focused on revision THR concluded that revision THR seems cost-effective but that it is resource intensive and has important implications for the allocation of health-care funding as the number of revisions is expected to increase with increasing demand for THR.[288–293] Vanhegan *et al.*[292] evaluated the costs associated with revision THR for different indications and reported that costs vary significantly by indication and that these variations were not reflected in the NHS tariffs. Durable implants and reduction in complications such as early dislocations have been suggested to be the solutions to reduce revision rates.[289] However, the highest revision costs were reported for revision as a result of infection,[292] with infections caused by methicillin-resistant strains of bacteria (41% of periprosthetic joint infections) incurring significantly higher costs than infections with sensitive strains of bacteria.[293]

Four studies evaluated the usefulness of different outcome measures for measuring quality of life after THR or revision THR, which showed that there was no consistency in the tools used to assess quality of life. Feeny *et al.*[294] reported that there is low agreement between certain outcome measures [SF-36, standard gamble, Health Utilities Index (HUI)-2 and HUI-3]. Dawson *et al.*[295] and Jones *et al.*[296] found that disease-specific measures reported larger changes than generic and utility measures. Ostendorf *et al.*[297] recommended the use of the OHS and the SF-12 in the assessment of THR and the EQ-5D in situations in which utility values are needed.

Overall, studies confirmed the long-standing claims that THR and RS are cost-effective interventions for patients with OA of the hip. However, there is little evidence from long-term trials on differences between implant brands and types of prostheses. This limits the conclusions that can be drawn with regard to the most cost-effective type of prosthesis. Studies used different methodologies to estimate costs (reference costs vs. prices actually paid by health-care centres) and definitions of costs included varied extensively, and many studies did not clearly report how costs were broken down. Although this review concentrates on clinical outcomes measured by the EQ-5D, the included studies tended to use more than one outcome measure with great variation across studies. In summary, THR, more so than RS, is a widely researched topic and receives great interest in many countries; however, further research should set out to include an assessment of the cost-effectiveness of different treatments.

Core studies for the cost-effectiveness analysis

Ranking eligible cost studies by year and country (most recent UK studies on top) and utility studies by number of participants, 11 studies were identified that were potentially useful to inform the decision model. These included one HTA and a further four cost-effectiveness studies. The HTA assessed the cost-effectiveness of hip RS compared with watchful waiting and THR.[19] The cost-effectiveness studies included three models that compared the cost-effectiveness of RS and THR,[253] the cost-effectiveness of cemented, cementless and hybrid prostheses[44] and the cost-effectiveness of two particular prosthesis types,[38] respectively. One cost-effectiveness study was included that evaluated THR and RS but did not use a model.[40]

The remaining six studies included partial economic evaluations that examined either costs or consequences but not both. Vanhegan *et al.*[292] reported costs for revision THR; Baker *et al.*[252] and Hulleberg *et al.*[258] reported medium- to long-term utilities in small populations; Dawson *et al.*[295] investigated quality of life post revision THR; and Bozic *et al.*[284] measured health state utilities for chronic OA of the hip, successful primary THR, failed primary THR, successful revision THR, failed revision THR and chronically infected THR. Rolfson *et al.*[298] evaluated the Swedish patient-reported outcomes data, reporting utilities for close to 35,000 THR patients.

Of the 11 studies three reported costs for THR,[19,40,44] two reported costs for follow-up of successful THR[19,40] and three reported costs of revision THR[19,44,292] (see *Appendix 12*). Costs for RS were reported in three studies.[19,40,253] Of these, Edlin *et al.*[40] and Vale *et al.*[19] also reported follow-up costs after successful RS and Bozic *et al.*[253] reported costs for revision RS (see *Appendix 11*).

The studies reporting the most useful data on utilities following THR were those by Pennington *et al.*,[44] Rolfson *et al.*,[298] Hulleberg *et al.*,[258] Dawson *et al.*[295] and Bozic *et al.*[284] (see *Appendix 14*). Utilities for RS were reported in only three studies[40,252,253] (see *Appendix 15*). No data were identified on quality of life at > 12 months post RS or for post-revision RS. Follow-up costs reported by Vale *et al.*[19] were the same for THR, RS and revision THR. Similarly, Bozic *et al.*[253] made no distinction between revision following THR or RS in terms of costs.

Quality assessment of core studies

Of the 11 core studies, five[252,253,258,295,298] provided useful information on EQ-5D utility scores only and one[292] provided useful data on costs only. These partial economic evaluations were not included in the critical appraisal.[309]

Five studies[19,38,40,44,253] were full economic evaluations and have been critically appraised using the CHEC-list.[249] Of these five studies, four[19,38,44,253] included models. These studies have also been critically appraised using an adapted checklist for models developed by Philips *et al.*[250]

Table 46 shows that all studies met ≥ 16 of the 19 criteria in the CHEC-list.[249]

Table 47 shows that all studies met ≥ 20 of the 32 criteria for economic models provided by Philips *et al.*[250] All studies correctly reported the time horizon and the perspective of the model, and the inputs used within the models were consistent with the perspectives that were chosen. In terms of costs and outcomes used in the model, these were appropriate to the specific study data set that was used. All studies conducted subgroup analyses. None of the studies applied a half-cycle correction and no justification was given for its exclusion. In addition, Pennington *et al.*[44] did not provide a clear definition of all of the options under evaluation and Briggs *et al.*[38] did not specify the cycle length of the model.

TABLE 46 Critical appraisal of the economic evaluation studies using the CHEC-list[249]

CHEC-list	Bozic 2010[253]	Briggs 2004[38]	Edlin 2012[40]	Pennington 2013[44]	Vale 2002[19]
1. Is the study population clearly described?	Y	Y	Y	Y	Y
2. Are competing alternatives clearly described?	Y	Y	Y	Y	Y
3. Is a well-defined research question posed in answerable form?	Y	Y	Y	Y	Y
4. Is the economic study design appropriate to the stated objective?	Y	Y	Y	Y	Y
5. Is the chosen time horizon appropriate to include relevant costs and consequences?	Y	Y	Y	Y	Y
6. Is the actual perspective chosen appropriate?	Y	Y	Y	Y	Y
7. Are all important and relevant costs for each alternative identified?	Y	Y	Y	Y	Y
8. Are all costs measured appropriately in physical units?	Y	Y	Y	Y	Y
9. Are costs valued appropriately?	Y	Y	Y	Y	Y
10. Are all important and relevant outcomes for each alternative identified?	Y	Y	Y	Y	Y
11. Are all outcomes measured appropriately?	Y	Y	Y	Y	Y
12. Are outcomes valued appropriately?	Y	Y	Y	Y	Y
13. Is an incremental analysis of costs and outcomes of alternatives performed?	Y	Y	Y	Y	Y
14. Are all future costs and outcomes discounted appropriately?	Y	Y	NA	Y	Y
15. Are all important variables whose values are uncertain appropriately subjected to sensitivity analysis?	Y	Y	Y	Y	Y
16. Do the conclusions follow from the data reported?	Y	Y	Y	Y	Y
17. Does the study discuss the generalisability of the results to other settings and patient/client groups?	Y	N	Y	UN	N
18. Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?	UN	Y	Y	Y	UN
19. Are ethical and distributional issues discussed appropriately?	N	N	N	UN	N

N, no; NA, not applicable; UN, unclear; Y, yes.

TABLE 47 Critical appraisal of the economic models using an adapted checklist from Philips *et al.*[250]

Philips criteria	Bozic 2010[253]	Briggs 2004[38]	Pennington 2013[44]	Vale 2002[19]
Structure				
1. Is there a clear statement of the decision problem?	Y	Y	Y	Y
2. Is the objective of the model specified and consistent with the stated decision problem?	Y	Y	Y	Y
3. Is the primary decision-maker specified?	N	Y	N	Y
4. Is the perspective of the model stated clearly?	Y	Y	Y	Y
5. Are the model inputs consistent with the stated perspective?	Y	Y	Y	Y
6. Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	Y	Y	Y	Y
7. Are the sources of the data used to develop the structure of the model specified?	Y	Y	Y	Y
8. Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	UN	Y	UN	UN
9. Is there a clear definition of the options under evaluation?	Y	Y	UN	Y
10. Have all feasible and practical options been evaluated?	Y	N	Y	Y
11. Is there justification for the exclusion of feasible options?	UN	N	UN	UN
12. Is the chosen model type appropriate given the decision problem and specified casual relationships within the model?	Y	Y	Y	Y
13. Is the time horizon of the model sufficient to reflect all important differences between the options?	Y	Y	Y	Y
14. Do the disease states (state transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of interventions?	Y	Y	Y	Y
15. Is the cycle length defined and justified in terms of the natural history of disease?	Y	UN	Y	Y
Data				
16. Are the data identification methods transparent and appropriate given the objectives of the model?	N	Y	Y	Y
17. Where choices have been made between data sources are these justified appropriately?	Y	UN	Y	Y
18. Where expert opinion has been used are the methods described and justified?	NA	NA	NA	Y
19. Is the choice of baseline data described and justified?	N	Y	Y	Y
20. Are transition probabilities calculated appropriately?	UN	Y	UN	Y
21. Has a half-cycle correction been applied to both costs and outcomes?	N	N	N	N
22. If not, has the omission been justified?	N	N	N	N
23. Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	UN	Y	Y	Y
24. Are the costs incorporated into the model justified?	Y	Y	Y	Y
25. Has the source for all costs been described?	Y	Y	Y	Y
26. Have discount rates been described and justified given the target decision-maker?	Y	Y	Y	Y

continued

TABLE 47 Critical appraisal of the economic models using an adapted checklist from Philips *et al.*[250] (*continued*)

Philips criteria	Bozic 2010[253]	Briggs 2004[38]	Pennington 2013[44]	Vale 2002[19]
27. Are the utilities incorporated into the model appropriate?	Y	Y	Y	Y
28. Is the source of utility weights referenced?	Y	Y	Y	Y
29. If data have been incorporated as distributions, has the choice of distributions for each parameter been described and justified?	N	Y	N	NA
30. If data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	NA	NA	NA	N
31. Has heterogeneity been dealt with by running the model separately for different subgroups?	Y	Y	Y	Y
32. Have the results been compared with those of previous models and any differences in results explained?	Y	N	N	N

N, no; NA, not applicable; UN, unclear; Y, yes.

Core studies for the economic model

Of the 11 core studies, Edlin *et al.*,[40] Pennington *et al.*,[44] Vale *et al.*[19] and Vanhegan *et al.*[292] provided data for the model in *Chapter 9* (see *Chapter 9* for the rationale of the selection procedure). This section will provide a brief description of the four core studies (*Table 48*).

Edlin *et al.*[40] reported a cost–utility analysis of RS compared with THR as part of a RCT of 126 adult patients with severe arthritis of the hip. Patients were randomised on a 1 : 1 basis between THR and RS. All RS patients received a Cormet™ (Corin Group, Cirencester, UK) metal-on-metal RS prosthesis. The THR patients received one of three types of prosthesis (ceramic-on-ceramic, metal-on-metal or metal-on-polyethylene) depending on the surgeon's preference. The study took the NHS perspective and considered the within-trial period without any extrapolation past the 12-month trial period. The costs were reported in 2009/10 UK pounds and EQ-5D 3 Levels (EQ-5D-3L) outcomes were measured as secondary outcomes of the trial.

The study used Healthcare Resource Group v4 (HRG4) reference costs combined with NHS trust finance department list prices for implants and IPD on length of stay (LOS). Resource use data and personal costs were obtained from patient-reported data. Univariate sensitivity analyses included an assessment of the impact of using the cheapest THR type (metal-on-metal) for all THR operations. The study reported NHS and Personal Social Services (PSS) costs after 12 months by type of hip replacement (THR vs. RS), including the costs of the initial operation/care, subsequent inpatient, outpatient, primary and community care, aids and medication [THR £7217 (£1320); RS £6653 (£917)], as well as private and social costs. The main results of this analysis included a difference in QALYs of 0.033 in favour of RS after 12 months and a greater cost of RS (difference of £564) in the first 12 months following surgery. This resulted in an ICER for RS of £17,451 per QALY. These results are based on a short-term trial using a single RS prosthesis type. The study did not explore variation in costs within for each type of prosthesis used in THR. Variation in prosthesis costs by hospital, a change in current practice regarding the choice of THR implant, longer follow-up (including higher revision rates for RS than for THR) and use of different RS implants may affect the reported cost-effectiveness in this study.

Pennington *et al.*[44] used IPD from three data sources (national PROMs programme, the NJR and Hospital Episode Statistics) to compare the cost effectiveness of cemented, cementless and hybrid THR in adult patients with hip OA. They conducted a probabilistic Markov model over patients' lifetime taking the NHS perspective. Implant prices were based on prices paid by English NHS centres. Costs for surgery plus hospital stay were taken from the literature and adjusted for LOS by prosthesis type and costs of revision were varied by reason for revision. Costs were reported as 2010/11 prices. The national data sources provided data on quality of life, LOS, rates of revision and rerevision and mortality for 30,203 patients.

TABLE 48 Characteristics of key cost-effectiveness studies informing the Markov model

Study and country	Study design	Methods	Results	Main conclusion	Information provided in the study
Edlin 2012,[40] UK Costa 2012,[130] UK	Type: RCT and economic (cost–utility) analysis Aim: to report on the relative cost-effectiveness of THR and RS in patients with severe arthritis suitable for hip joint RS	Population: patients aged > 18 years with severe arthritis of the hip joint suitable for RS ($n = 126$): THR $n = 66$, RS $n = 60$ Outcomes: primary: hip function (12 months post-surgery OHS and HHS); secondary: quality of life (EQ-5D), disability rating, physical activity level, complications, cost-effectiveness, incremental costs, ICERs Economic analysis: NHS perspective, 12-month time horizon, cost year 2009/10 (£), univariate sensitivity analyses	Hip function: mean OHS: effect size 2.23 (95% CI –1.52 to 5.98, $p = 0.070$); mean HHS: effect size 6.04 (95% CI –0.51 to 12.58, $p = 0.242$) Complication rates did not differ ($p = 0.291$) Quality of life at 12 months: RS 0.795, THR 0.727; RS vs. THR: incremental QALYs 0.032, incremental cost £564, ICER £17,451 per QALY	No evidence of a difference in hip function between groups was seen in patients with severe arthritis of the hip, 1 year post surgery RS appears to offer very short-term efficiency benefits over THR within a selected patient group	1(a) Resource use ☒ (b) Costs ☒ 2(a) Utilities ☒ (b) QALYs ☒ 3 Transition probabilities
Pennington 2013,[44] UK	Type: retrospective economic (cost–utility) and decision analysis Aim: to evaluate the relative cost-effectiveness of cemented, cementless and hybrid prostheses for elective THR surgery	Population: patients undergoing primary THR for OA ($n = 30,203$ for quality of life analysis) Male: cemented 35.1% ($n = 4195$), cementless 44.6% ($n = 6548$), hybrid 38.0% ($n = 1350$) Age (years), mean (SD): cemented 72.4 (6.7), cementless 67.8 (7.2), hybrid 70.4 (7.2) Outcomes: quality of life 6 months post surgery (OHS, EQ-5D), lifetime cost-effectiveness, costs (£), ICERs Economic model: health service perspective, cost year 2010/11 (£); sensitivity analysis of QALYs post 2 years, revision rates using different hazard function, failed hip category without revision, excluding metal-on-metal prostheses	Lifetime costs: lowest with cemented prostheses Postoperative quality of life and lifetime QALYs: highest with hybrid prostheses Women aged 70 years: mean costs for cemented prosthesis £6900, mean costs for cementless prosthesis £7800, mean costs for hybrid prosthesis £7500 Mean postoperative EQ-5D scores: cemented 0.78, cementless 0.80, hybrid 0.81 Lifetime QALYs: cemented 9.0, cementless 9.2, hybrid 9.3 ICER: hybrid vs. cemented £2500 per QALY	Cemented prostheses were the least costly type for THR. For most patient groups hybrid prostheses were the most cost-effective. Cementless prostheses did not provide a sufficient improvement in health outcomes to justify their additional costs	1(a) Resource use (b) Costs ☒ 2(a) Utilities ☒ (b) QALYs 3 Transition probabilities Comment: initial costs (including prosthesis, operating theatre and hospital stay costs); utilities and revision rates; costs and utilities by sex, year group and prosthesis type

continued

TABLE 48 Characteristics of key cost-effectiveness studies informing the Markov model (*continued*)

Study and country	Study design	Methods	Results	Main conclusion	Information provided in the study
Vale 2002,[19] UK McKenzie 2003,[299] UK	Type: systematic review and retrospective economic (cost–utility) analysis Aim: to assess the effectiveness and cost-effectiveness of metal-on-metal hip RS compared with watchful waiting, THR, osteotomy, arthrodesis and arthroscopy of the hip joint	Population: patients with hip disease Age (years): 45–50 and 65–70 Outcomes: costs (£), QALYs, ICERs Economic model: Markov model, 20-year time horizon, NHS perspective, cost year 2000 (£); subgroup analysis considering those who would not outlive a THR; sensitivity analyses for revision rates, operation times, watchful waiting costs, time horizon and quality of life	Revisions: RS over 3-year follow-up: 0–14%, THR over 10-year follow-up: ≤ 10%, osteotomy over 10- to 17-year follow-up: between 2.9% and 29% Patients pain free: RS: 91% at 4 years, THR: 84% at 11 years, arthrodesis: 22% at 8 years Costs: RS for a patient aged < 65 years £5515, THR £4195, revision £6027, arthroscopy £951, osteotomy £2731, watchful waiting £642 annually Cost-effectiveness: for patients aged < 65 years, RS dominated by THR; RS dominated watchful waiting within 20-years' follow-up Incremental cost per QALY: RS vs. osteotomy £3039, RS vs. arthroscopy £366 For patients aged > 65 years, THR dominated RS	Metal-on-metal RS had lower revision rates than THR over an extended time period and resulted in better outcomes overall for those who are likely to outlive a primary THR. If metal-on-metal RS has lower revision rates than THR over an extended period and results in better outcomes from subsequent THR, then metal-on-metal RS could possibly be considered cost-effective or even dominant	1(a) Resource use (b) Costs ☒ 2(a) Utilities (b) QALYs ☒ 3 Transition probabilities ☒ Comment: revision rates for metal-on-metal RS and THR; costs including prosthesis costs; broken-down costs for watchful waiting
Vanhegan 2012,[292] UK	Type: retrospective economic analysis Aim: to evaluate the costs associated with revision THR for different indications	Population: patients undergoing revision THR ($n = 286$; $n = 305$ procedures) Male: aseptic loosening ($n = 194$): 34% ($n = 65$), deep infection ($n = 76$): 42% ($n = 32$), periprosthetic fracture ($n = 24$): 25% ($n = 6$), dislocation ($n = 11$): 28% ($n = 3$) Age (years), mean (range): aseptic loosening 67 (20–89), deep infection 62 (29–83), periprosthetic fracture 76 (31–88), dislocation 79 (54–90) OA: aseptic loosening 69%, deep infection 48%, periprosthetic fracture 80%, dislocation 54% Outcomes: LOS, costs (£)	Mean total costs for revision surgery: aseptic cases £11,897 (SD 4629), septic revision £21,937 (SD 310,965), periprosthetic fracture £18,185 (SD £9124), dislocation £10,893 (SD £5476) Surgery for infection and periprosthetic fracture: longer operating times, increased blood loss, increase in complications, longer LOS	Financial costs vary significantly by indication. Variation is not reflected in current NHS tariffs	1(a) Resource use ☒ (b) Costs ☒ 2(a) Utilities (b) QALYs 3 Transition probabilities

LOS, length of stay.

Patients receiving different prosthesis types were matched by age, sex, number of comorbidities, ASA grade, BMI, deprivation, preoperative quality of life, surgeon experience and hospital type. The study reported data on the combined cost of the prosthesis, operating theatre and hospital stay, quality of life at 6 months post surgery and 5- and 10-year revision rates by prosthesis type, age group and sex. Overall, the study concluded that in patients aged 70 years the ICER for a hybrid prosthesis compared with a cemented prosthesis was £2100 for men and £2500 for women, with hybrid prostheses resulting in higher quality of life in all subgroups except women aged 80 years and cemented prostheses being the least costly option. The initial costs of a cementless prosthesis were highest in all subgroups. One of the limitations of the study was that it assumed that the observed quality of life at 6 months post surgery would remain unchanged for the patients' lifetime. Furthermore, the study did not consider different revision rates by brand for the three different THR types.

Vale *et al.*[19] undertook an assessment of the clinical effectiveness and cost-effectiveness of RS compared with watchful waiting (i.e. patient monitoring, drug-based treatment and supportive activities including physiotherapy), THR and other bone-conserving treatments. The HTA comprised a systematic review of the clinical effectiveness and cost-effectiveness of RS compared with any of the treatments above and a Markov model comparing the comparators from the NHS perspective for patients suitable for RS for up to 20 years. Cost data (in 2000/1 UK pounds) for THR and revision THR were taken from the literature (£4195 and £6027, respectively) and prostheses costs for RS were obtained from manufacturers. The model considered the lower of the two RS implant costs obtained (£1730 vs. £1890), resulting in an overall cost of £5515 for RS. LOS was estimated to be 10 or 12 days for THR and 8 or 10 days for RS. All other costs including use of the operating theatre and staff, radiography, outpatient visits and first-year follow-up costs were assumed to be the same for RS and THR. First-year follow-up included two outpatient visits with one radiography scan, totalling £118.74. Quality-of-life estimates considered pain levels and quality-of-life scores for mild, moderate and severe OA and were combined with revision and mortality rates to generate QALYs.

The main conclusion from the systematic review was that evidence from the literature on the effectiveness of RS was limited. Revision rates were reported to range between 0% and 14% over a 3-year follow-up period for RS compared with $\leq 10\%$ over 10 years for THR. Patients with RS experienced less pain than patients managed by watchful waiting. Results from the model showed that RS was dominated by THR based on assumptions about revision rates for RS and the lower cost of THR. In subsequent sensitivity analyses the revision rates for RS had to be reduced to < 80–88% of the THR revision rates before RS was no longer dominated by THR. However, RS dominated watchful waiting within the 20-year follow-up. The study was limited because of the lack of data for the parameters of the model, particularly revision rates for different RS brands and effectiveness data for revision THR following RS. Furthermore, available data for RS originated from a small number of surgeons.

Vanhegan *et al.*[292] investigated the costs of 305 consecutive revision THRs by reason for revision in 286 patients, with a diagnosis of hip OA in 64% of revisions ($n = 195$). Revision THR was carried out in a single tertiary centre by one of three experienced surgeons. Costs were obtained from the finance department of the tertiary centre (in 2007/8 UK pounds) and included costs of the implant, materials and augmentation, use of the operating theatre and recovery room, the inpatient stay and laboratory tests, radiology, pharmacy, physiotherapy and occupational therapy. The study provided cost data on 13 different implants and data on resource use and costs by reason for revision (aseptic loosening, deep infection, periprosthetic fracture and dislocation).

The mean costs of revision for aseptic loosening, deep infection, periprosthetic fracture and dislocation were reported to be £11,897 (SD £4629), £21,937 (SD £10,965), £18,185 (SD £9124) and £10,893 (SD £5476), respectively. Higher complication rates as well as reoperation rates were associated with revisions for deep infection, periprosthetic fracture and dislocation. However, the numbers of revisions for these three indications were relatively small ($n = 76$, $n = 24$ and $n = 11$, respectively). Although the cost estimates can be assumed to be very accurate, they are limited by their lack of generalisability as they were

based on one single tertiary centre. Furthermore, the study did not consider the cost of readmission for complications and other direct and indirect medical and social costs.

Summary of the cost-effectiveness evidence

We found that four[19,40,44,292] of the 11 core cost-effectiveness studies were able to provide utility and cost data for the model. We assessed these using the checklists developed by Evers *et al.*[249] and Philips *et al.*[250] and found them to be of varying quality. All studies met ≥ 16 of the 19 criteria for economic analyses provided by Evers *et al.*[249] and ≥ 20 of the 32 criteria for economic models provided by Philips *et al.*[250]

Methods for the review of registries

Identification of studies

Initial scoping searches were undertaken in MEDLINE in October 2012 to assess the volume and type of literature relating to national joint registries for hip replacement procedures. These scoping searches informed the development of the final search strategy (see *Appendix 1*). The registry search strategy was designed to capture the generic terms for 'arthritis', total hip replacement' and 'resurfacing arthroplasty' in addition to the word 'registry'. Searches were not date limited for the registry search and were undertaken in November 2012 (see *Appendix 1*). All bibliographic records identified through the electronic searches were collected in a managed reference database.

The following databases of published studies were searched: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, Science Citation Index and Conference Proceedings Citation Index – Science, The Cochrane Library (specifically CDSR, CENTRAL, DARE, NHS EED, HTA database) and CEA Registry (articles).

Inclusion and exclusion criteria

The following inclusion and exclusion criteria were used to identify eligible papers reporting joint replacement studies. The aim was to identify any studies that reported survival, utilities and outcomes that would potentially be useful for the economic model and survival analysis.

Inclusion criteria

Study design (registries)

- Reporting of the results of joint replacement registry data collection.
- All study designs.
- Most recent publication in the series.

Population

- People with pain or disability resulting from end-stage arthritis of the hip for whom non-surgical management has failed.

Intervention

- Elective primary THR.
- Primary hip RS.

Comparator

- Different types of primary THR compared with hip RS for people in whom both procedures are suitable.
- Different types of primary THR compared with each other for people not suitable for hip RS.

Record

- Full-text articles of completed studies published in English and annual reports of national registries.

Outcomes

- All reported outcomes.

Exclusion criteria

- Abstract/conference proceedings, letters and commentaries.
- Non-English-language publications.
- < 1000 patients included in the registry study at the time of publication.
- Hip/knee data not reported separately.

Assessment of eligibility

All retrieved records were collected in a referencing database and all duplicate records were identified and removed. The search returned 541 records. An initial sift was undertaken by one reviewer to exclude clearly non-relevant records using the following exclusion criteria:

- non-hip only papers
- papers on animals
- papers on children
- non-registry papers
- papers on surgery for hip fracture only
- non-English full-text papers.

This was followed by a formal sift of 329 papers by title and abstract by two reviewers using the inclusion/exclusion criteria. All identified relevant studies were read in full by one reviewer to identify eligible studies, with cross-checking by a second reviewer. Disagreement was resolved by a third reviewer. Reasons for exclusion of full-text papers were documented.

Data extraction

Data extraction was carried out on the final eligible papers by one reviewer in two stages. In stage one all eligible studies were considered and in stage two the studies that would provide useful input to the economic model and survival analysis were identified. Data extracted in stage one included the following:

- author surname
- publication year
- country of registry
- year that registry data were collected
- type of registry data collected
- size of the registry database
- description of the patient population
- results of key outcomes.

Data extraction of the overall aim and conclusion of each paper was also conducted to help identify inputs for the economic model and survival analysis. During stage two data extraction, registry studies were ordered by their publication year to ensure that the most recent data were extracted. Stage two extraction included the following additional exclusion criteria:

- not the most recent paper in a publication series
- not the most recent annual report from a national joint registry.

Results of the registry review

Identification of studies

The PRISMA flow diagram outlining the identification of registry studies is shown in *Figure 19*.[99] The database search for registry studies identified 538 publications, with an additional record identified through other sources. A total of 326 papers remained once duplicates were removed and these were screened for relevance. This process resulted in the exclusion of a further 230 papers, with 96 papers screened at title and abstract level. A further 47 studies were excluded with a reason provided (see *Appendix 16*), resulting in the inclusion of 49 studies in the review.[15,16,49,261,298,310–353]

Of the 49 papers included in the review, 44 were carried out in the following 10 countries: Japan ($n = 1$[310]), Australia ($n = 5$[311–315]), the UK ($n = 7$[15,16,316–319,353]), Italy ($n = 2$[261,320]), Finland ($n = 10$[321–330]), Norway ($n = 5$[331–335]), the USA ($n = 4$[49,336–338]), Denmark ($n = 4$[339–342]), Sweden ($n = 3$[298,343,344]), and Slovakia ($n = 1$[345]). In addition, seven papers[346–352] reported outcomes from multinational registries.

In stage two, 19[49,298,310,312,316,319,321,322,324–326,331,332,336,338,340,342,346,352] of the 49 papers were excluded (not most recent paper publication in a series or not most recent annual report from a national joint registry). Therefore, 30 papers were included in the narrative review, reflecting the most recent publication in a series from each particular registry for both THR and RS.[15,16,261,311,313–315,317,318,320,323,327–330,333–335,337,339,341,343–345,347–351,353]

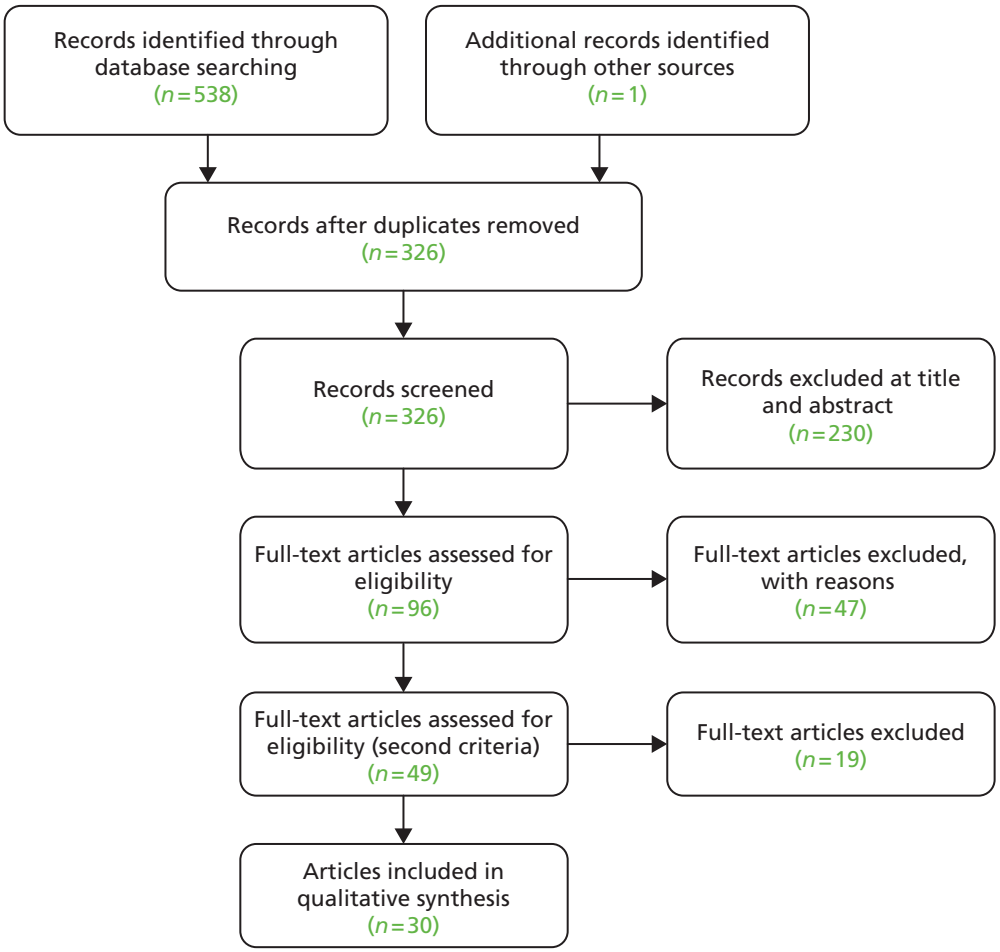


FIGURE 19 Flow diagram of study identification for the registry review.

Review of included studies following stage two exclusion

A narrative review of the included papers by intervention type (THR, RS) and country is given in the following sections. The 30 papers did not report similar patient populations, interventions, comparator groups or outcomes and therefore they are reported separately. For the purposes of the economic model and survival analysis, revision rate and implant survival were the key outcomes to be extracted.

Resurfacing arthroplasty

Eight registry studies provided evidence on RS.[15,311,313,318,329,349,351,353] The majority of these studies investigated various comparisons between THR and RS. *Table 49* provides a summary of the RS studies.

England and Wales

Jameson *et al.*[353] conducted a retrospective cohort study and reported survival time to revision for RS procedures from 2003 to 2013. The study explored the risk factors independently associated with failure. Mean time to revision for each group was not reported. Data were taken from the NJR for England and Wales. The study concluded that women were at greater risk of revision than men (HR 1.30, 99% CI 1.01 to 1.76; $p = 0.007$), independent of age. Smaller femoral head components were also significantly more likely to require revision than medium (≤ 44 mm: HR 2.14, 99% CI 1.53 to 3.00; $p < 0.001$) or large heads (45–47 mm: HR 1.48, 99% CI 1.09 to 2.00; $p = 0.001$), as was surgery performed by low-volume surgeons (HR 1.36, 99% CI 1.09 to 1.71; $p < 0.001$).

McMinn *et al.*[318] examined mortality and revision rates among patients with OA undergoing THR, both cemented and uncemented procedures, or RS. The authors used data from the NJR database for the analysis [154,996 patients receiving cemented THR, 120,017 receiving uncemented THR and 8352 receiving RS (in particular, Birmingham hip RS)]. The baseline characteristics recorded include age (cemented mean 73.2 years, uncemented mean 66.7 years), sex (cemented: men 53,409, women 101,587; uncemented: men 50,529, women 69,488) and ASA grade. The analysis took into account the age of patients at primary surgery and their length of follow-up. Survival analysis was used to compare the cemented and uncemented procedures with adjustment for sex, age at primary surgery, ASA grade before the operation, complexity of the procedure and 'both sides' (surgery on both hips at the same time).

The multivariable survival analyses demonstrated a higher mortality rate for patients undergoing cemented THR than for those undergoing uncemented THR (adjusted HR 1.11, 95% CI 1.07 to 1.16). There was a lower revision rate for cemented procedures (unadjusted HR 0.53, 95% CI 0.50 to 0.57). The authors stated that these findings translate into small predicted differences in the population-averaged absolute survival probability at all time points. At 8 years post surgery the predicted probability of death in the cemented group was 0.013 higher (95% CI 0.007 to 0.019) than that in the uncemented group and the predicted probability of revision was 0.015 lower (95% CI 0.012 to 0.017). In multivariable analyses that included only men, there was a higher mortality rate in the cemented group and the uncemented group than in the RS group. RS had a similar revision rate to uncemented THR and both had a higher revision rate than cemented THR. The authors concluded that there was a small but significant increased risk of revision with uncemented THR compared with cemented THR, and a small but significant increased risk of death with cemented procedures.

A study from Smith *et al.*[15] reported that, in women, RS resulted in worse implant survival than THR, regardless of head size. The predicted 5-year revision rates in 55-year-old women were 8.3% (95% CI 7.2% to 9.7%) for a 42-mm RS head, 6.1% (95% CI 5.3% to 7.0%) for a 46-mm RS head and 1.5% (95% CI 0.8% to 2.6%) for a 28-mm cemented metal-on-polyethylene stemmed THR. In men with smaller femoral heads, RS resulted in poor implant survival. Predicted 5-year revision rates in 55-year-old men were 4.1% (95% CI 3.3% to 4.9%) for a 46-mm RS head, 2.6% (95% CI 2.2% to 3.1%) for a 54-mm RS head and 1.9% (95% CI 1.5% to 2.4%) for a 28-mm cemented metal-on-polyethylene stemmed THR. Of the male RS patients, only 23% (5085/22,076) had a head size ≥ 54 mm. The authors concluded that RS resulted in similar implant survival to other surgical options in men with large femoral heads, and worse implant survival in other patients, particularly women.

TABLE 49 Summary table of registry studies on RS

Study	Registry	Implant type/comparator	Outcomes	Results
Jameson 2012[353]	NJR	Men vs. women undergoing RS	Survival time to revision for RS procedures	Women were at greater risk of revision than men (HR 1.30, 99% CI 1.01 to 1.76; $p = 0.007$)
McMinn 2012[318]	NJR	Cemented vs. uncemented THR procedures, and cemented and uncemented THR procedures vs. RS in men only	Mortality and revision rates (8 years)	Higher mortality rate for patients undergoing cemented than for patients undergoing uncemented THR (adjusted HR 1.11, 95% CI 1.07 to 1.16)
Smith 2012[15]	NJR	Men vs. women undergoing RS by femoral head size	Revision rate (5 years)	Revision rate: women, 55 years: 8.3% (95% CI 7.2% to 9.7%) for 42-mm RS head, 6.1% (95% CI 5.3% to 7.0%) for 46-mm RS head and 1.5% (95% CI 0.8% to 2.6%) for 28-mm cemented metal-on-polyethylene stemmed THR; men, 55 years: 4.1% (95% CI 3.3% to 4.9%) for 46-mm RS head, 2.6% (95% CI 2.2% to 3.1%) for 54-mm RS head and 1.9% (95% CI 1.5% to 2.4%) for 28-mm cemented metal-on-polyethylene stemmed THR
Seppanen 2012[329]	Finnish Arthroplasty Register	RS vs. THR	Risk of revision (3.5–3.9 years)	No statistically significant difference in risk of revision between RS and THR (risk of revision 0.93, 95% CI 0.78 to 1.10)
Buergi 2007[311]	Australian National Joint Replacement Registry	RS vs. THR	Risk of revision (3 years)	Revision rates after RS and THR were 2.8% and 2.0%, respectively
Corten 2010[313]	Multinational	RS vs. THR	Revision rate (3 years)	Revision rate for RS was 1.8% in England and Wales and 3.4% in Sweden
Johanson 2010[349]	Nordic Arthroplasty Registry	RS vs. THR	RR	RS had a threefold increased revision risk compared with THR (RR 2.7, 95% CI 1.9 to 3.7)
Schuh 2012[351]	Multinational	RS reported in registry vs. clinical studies from specialist centres	Revision rates (difference in revisions per 100 observed component-years)	Specialist clinical centres (defined by the number of patients treated, staff training and personal expertise): 0.27 (95% CI 0.14 to 0.40) per 100 observed component-years; register data: 0.74 (95% CI 0.72 to 0.76) per 100 observed component-years. Average revision rate from the register data was 3.41% (SD 1.79%)

Finland

Seppanen *et al.*[329] analysed the risk of revision of 4401 RS procedures in the Finnish Arthroplasty Register compared with the risk of revision of 48,409 THRs performed during the same time period. The median follow-up time was 3.5 (range 0–9) years for RS and 3.9 (range 0–9) years for THRs. The study reported no statistically significant difference in risk of revision between RS and THR (risk of revision 0.93, 95% CI 0.78 to 1.10). The 4-year unadjusted Kaplan–Meier survival rate was 96% (95% CI 96% to 97%) for both the RS group and the THR group. Female patients had about double the risk of revision as male patients (risk of revision 2.0, CI 1.4 to 2.7).

Australia

Buergi *et al.*[311] reported the use of RS based on the Australian National Joint Replacement Registry. A total of 7205 RS procedures were carried out between 1999 and 2005. The study concluded that, in the database, early revision rates were higher for RS than for THR. At 3 years, the revision rate after RS was 2.8% and that after THR was 2.0%.

Multinational

Corten *et al.*[313] compared RS survivorship reported by registries in Australia, England and Wales and Sweden with the failure of THR between 2006 and 2009. RS was associated with an overall increased failure rate compared with THR. The cumulative revision rates in the Australian registry were 3.7% for RS and 2.7% for THR. The 3-year revision rate for RS was 1.8% in England and Wales and 3.4% in Sweden.

A study using data from the Nordic Arthroplasty Registry compared the outcome of RS ($n = 1638$) with that of THR ($n = 309{,}290$) between 1995 and 2007.[349] Results indicated that RS had a threefold increased revision risk compared with THR (RR 2.7, 95% CI 1.9 to 3.7). The difference was greater when RS was compared with cemented THR (RR 3.8, 95% CI 2.7 to 5.3). In men aged < 50 years the difference in revision risk was less (RS vs. THR: RR 1.9, 95% CI 1.0 to 3.9; RS vs. cemented THR: RR 2.4, 95% CI 1.1 to 5.3). However, the difference in revision risk was higher in women of the same age group (RS vs. THR: RR 4.7, 95% CI 2.6 to 8.5; RS vs. cemented THR: RR 7.4, 95% CI 3.7 to 15). In the Cox regression analysis, RS showed an increased risk of early aseptic revision compared with THR (RR 2.7, 95% CI 1.9 to 3.7; $p < 0.001$) and cemented THR (RR 3.8, 95% CI 2.7 to 5.3; $p < 0.001$).

The purpose of one recent study[351] was to evaluate the outcome of Birmingham hip RS using revision rates as reported in national joint replacement registry studies (categorised as from the UK, Australia, Asia and the USA). In total, 9806 RS procedures were analysed (reported as 44,294 observed component-years). The analysis revealed a significant difference in revisions per 100 observed component-years between studies authored by specialist clinical centres (defined by the number of patients treated, staff training 4and personal expertise) (0.27, 95% CI 0.14 to 0.40) and the register data (0.74, 95% CI 0.72 to 0.76). The average revision rate from register data was 3.41% (SD 1.79%).

Summary of resurfacing arthroplasty in registry studies

In summary, the eight studies that reported data from joint registries had mixed results. There is little evidence from long-term studies; generally, 5-year revision rates (or less) were reported. No two studies had the same comparators for analysis, which makes drawing conclusions from the eight studies difficult. The reported benefits of RS include preservation of the bone on the femoral side, greater physiological stress transfer at the proximal femur and lower risk of dislocation because of the larger femoral head compared with conventional THR.[351] However, the majority of studies included in this review found that RS had a higher revision rate than THR, particularly in female patients. Only one study found no significant difference between the procedures.[329] No studies were included that reported RS implant survival as better than that for THR. One study of men only reported that RS had a similar revision rate to that of uncemented THR, but that both had a higher revision rate than that of cemented THR.[318]

Total hip replacement

In total, 22 registry studies reported evidence on THR, with the majority of these studies investigating various types of THR surgery or demographic differences regarding the specific countries. *Table 50* provides a summary of the THR studies.

England and Wales

Jameson *et al.*[317] reported survival time to revision following primary cemented THR in 34,721 THRs recorded in the NJR for England and Wales between 2003 and 2010. The authors reported the 7-year rate of revision for any reason as 1.70% (99% CI 1.28% to 2.12%). The overall risk of revision was independent of age, sex, ASA grade, BMI, surgeon volume, surgical approach, brand of cement/presence of antibiotic, femoral head material (stainless steel/alumina) and stem taper size/offset.

Smith *et al.*[16] assessed the use of metal-on-metal bearing surfaces in the NJR between 2003 and 2011. They reported that metal-on-metal THR failed at high rates and that this was linked to head size. Analysis of the 31,171 metal-on-metal THRs showed that larger heads failed earlier (cumulative incidence of revision: 3.2%, 95% CI 2.5% to 4.1% for 28-mm heads and 5.1%, 95% CI 4.2% to 6.2% for 52-mm heads at 5 years in men aged 60 years). The 5-year revision rates in younger women were 6.1% (95% CI 5.2% to 7.2%) for 46-mm metal-on-metal THR and 1.6% (95% CI 1.3 to 2.1) for 28-mm metal-on-polyethylene THR. This finding contrasted with findings for ceramic-on-ceramic bearing surfaces, for which larger head sizes were associated with improved survival (5-year revision rate: 3.3%, 95% CI 2.6% to 4.1% for 28-mm heads and 2.0%, 95% CI 1.5% to 2.7% for 40-mm heads for men aged 60 years).

Denmark

Johnsen *et al.*[339] examined the association between patient-related factors and the risk of initial, short-term and long-term failure after primary THR using data from the Danish Hip Arthroplasty Registry ($n = 36,984$). The study concluded that in Denmark between 1995 and 2002 male sex and comorbidity index score (Charlson Comorbidity Index) were strongly predictive of THR failure. The Charlson Comorbidity Index includes 19 disease categories, which correspond to *International Classification of Diseases*, Eighth Edition (ICD-8) and *International Classification of Diseases*, Tenth Edition (ICD-10) codes used in the national registries. A total of 1132 primary THRs were revised (3.1% of the 36,984 procedures) during this time period.

A more recent study from Denmark[341] evaluated short-term (0–90 days) and longer-term (up to 12.7 years) mortality of patients undergoing primary THR compared with mortality in the general population. THR patients ($n = 44,558$) was matched at the time of surgery with three people from the general population ($n = 133,674$). The findings suggest that there was a 1-month period of increased mortality immediately after surgery among THR patients (adjusted mortality rate ratio 1.4, 95% CI 1.2 to 1.7); however, overall short-term mortality (0–90 days) was significantly lower (adjusted mortality rate ratio 0.8, 95% CI 0.7 to 0.9). THR surgery was associated with increased short-term mortality in subjects aged < 60 years and among THR patients without comorbidity. Long-term mortality was lower among THR patients than in the general population control group (adjusted mortality rate ratio 0.7, 95% CI 0.7 to 0.7).

Sweden

Lazarinis *et al.*[343] analysed patient data ($n = 8043$) on cementless cups with or without a hydroxyapatite coating that had been recorded in the SHAR between 1992 and 2007. The primary end point was revision because of aseptic loosening; the secondary end points were cup revision for any reason and cup revision because of infection. The results reported that the hydroxyapatite coating was a risk factor for cup revision because of aseptic loosening (adjusted RR 1.7, 95% CI 1.3 to 2). Age at primary THR of < 50 years, paediatric hip disease, a cemented stem and the cup brand were also associated with a statistically significantly increased risk of cup revision due to aseptic loosening.

TABLE 50 Summary table of registry studies on THR

Study	Registry	Implant type/comparator	Outcomes	Results
Jameson 2012[317]	NJR	Primary cemented THR	Survival time to revision (7 years)	7-year rate of revision for any reason 1.70%
Smith 2012[16]	NJR	Metal-on-metal THR vs. non-metal-on-metal THR – head size and sex	Survival time to revision (5 years)	Larger heads failed earlier: cumulative incidence of revision 3.2% (95% CI 2.5% to 4.1%) for 28-mm heads and 5.1% (95% CI 4.2% to 6.2%) for 52-mm heads at 5 years in men aged 60 years. The 5-year revision rates in younger women were 6.1% (95% CI 5.2% to 7.2%) for 46-mm metal-on-metal THR and 1.6% (95% CI 1.3 to 2.1) for 28 mm metal-on-polyethylene THR
Johnsen 2006[339]	Danish Hip Arthroplasty Registry	Patient-related factors and the risk of initial, short-term and long-term failure after primary THR	Implant revision	Male sex and comorbidity index score (Charlson Comorbidity Index) were strongly predictive of THR failure. In total, 3.1% of the 36,984 procedures were revised
Pedersen 2011[341]	Danish Hip Arthroplasty Registry	Mortality of patients undergoing primary THR compared with that in the general population	Adjusted mortality rate ratio	Long-term mortality was lower among THR patients than in the general population control group (adjusted mortality rate ratio 0.7, 95% CI 0.7 to 0.7)
Lazarinis 2010[343]	SHAR	Cementless cups with or without HA	Revision because of aseptic loosening	HA coating was a risk factor for cup revision because of aseptic loosening (adjusted RR 1.7, 95% CI 1.3 to 2)
Weiss 2012[344]	SHAR	Monoblock cups vs. modular cups	Implant survival (5 years)	Implant survival 95% (95% CI 91% to 98%) for monoblock cups and 97% (95% CI 96% to 98%) for modular cups ($p = 0.6$)
Luo 2012[314]	AOANJRR	Identification of implants with higher than expected failure rates between 2003 and 2007	NR	Results state that if the poor-performing THRs had been conducted using average longevity designs, the number of THR revisions could have been reduced by 47%
Sexton 2009[315]	AOANJRR	Metal-on-polyethylene vs. ceramic-on-ceramic THR	Rate of revision	Higher rate of revision for dislocation in ceramic-on-ceramic THR than in metal-on-polyethylene THR when smaller head sizes (≤ 28 mm) were used in younger patients (< 65 years) (HR 1.53, $p = 0.041$) and also with larger head sizes (> 28 mm) in older patients (≥ 65 years) (HR 1.73, $p = 0.016$)
Di Tanna 2011[261]	Emilia-Romagna Regional Registry on Orthopaedic Prosthesis	Cementless vs. hybrid prostheses	Numbers of revisions expected	243 revisions would be expected in the cementless group vs. 300 in the hybrid group. This was equal to a 19% difference and a NNT of 18
Stea 2009[320]	Emilia-Romagna Regional Registry on Orthopaedic Prosthesis	Survival rates for THR in Italy between 2000 and 2006	Implant survival rate (7 years)	7-year implant survival rate was 96.8% (95% CI 96.4% to 97.1%)

continued

TABLE 50 Summary table of registry studies on THR (*continued*)

Study	Registry	Implant type/comparator	Outcomes	Results
Eskelinen 2005[323]	Finnish Arthroplasty Register	Population-based survival of cementless THR	Implant survival rate (10 years)	Survival rate of > 90% at 10 years for cementless THR
Makela 2011[327]	Finnish Arthroplasty Register	Cemented vs. cementless THR	Implant survival rate (15 years)	15-year survival rate for cementless THR (80%) was comparable with rates in the cemented groups (86%)
Makela 2011[328]	Finnish Arthroplasty Register	Cemented vs. cementless THR for OA patients	Implant survival rate (15 years)	Implant survival rates for the cementless THR groups (62%, 95% CI 57% to 67% and 58%, 95% CI 52% to 66%) were worse than that of the cemented THR group (71%, 95% CI 62% to 80%)
Nečas 2011[345]	Slovakia	Operations performed between 2003 and 2010	Revision rate (7 years)	Revision rate in period 2003–10 was 9.15%
Espehaug 2011[332]	Norwegian Arthroplasty Register	Differences by county and regional health authority over a 20-year period (1989–2008)	Numbers of THR procedures performed	Increase in number of THR procedures performed from 109 operations per 100,000 inhabitants in 1991–5 to 140 in 2006–8
Fevang 2010[333]	Norwegian Arthroplasty Register	Risks of revision during the time periods 1993–7, 1998–2002 and 2003–7 were compared to that in the reference period 1987–92	Revision risk	Reduced risk of revision in the time periods 1993–7, 1998–2002 and 2003–7 compared with the reference period
Schrama 2010[335]	Norwegian Arthroplasty Register	THR in RA patients vs. OA patients	Implant survival (5 years)	5-year survival was 99.5% in RA patients and 99.4% in OA patients (RR 0.98, 95% CI 0.65 to 1.48 for RA vs. OA patients)
Namba 2012[337]	Kaiser Permanente Total Joint Replacement Registry	Factors associated with deep SSI following THR	Incidence of SSI	155 deep SSIs (0.51%, 95% CI 0.43% to 0.59%) occurred at a mean of 72 days (median 28, SD 93.3 days) after the procedure
Sadoghi 2012[350]	Multinational	Compared primary THRs between different countries in terms of THR number per inhabitant, age and procedure type	Implant survival	THRs performed in Denmark showed the lowest survival rate within the first 15 years; however, THRs performed in Norway had similar low survival rates
Graves 2011[347]	Multinational	The use of metal-on-metal THR across three registries	NR	All registries reported an increased revision rate associated with larger femoral head sizes when using metal-on-metal bearing surfaces
Havelin 2009[348]	Nordic Registry	Compared demographics, choice of implant, fixation techniques and results between countries	Implant survival (10 years)	10-year survival rate was 92% (95% CI 91.6% to 92.4%) in Denmark, 94% (95% CI 93.6% to 94.1%) in Sweden and 93% (95% CI 92.3% to 93.0%) in Norway
Kadar 2012[334]	Nordic Registry	Metal femoral heads made from various materials (cobalt–chromium, aluminium, zirconium)	Implant survival (12 years)	The survival rate was 88.1% with cobalt–chromium heads and 74.8% with zirconium heads

HA, hydroxyapatite; NR, not reported; NNT, number needed to treat; RiR, risk of revision; RR, revision rate; SSI, surgical site infection.

A more recent study from Sweden reported data from 1999 to 2010.[344] The authors investigated revision rates of monoblock cups used in primary THR that were registered in the SHAR. Kaplan–Meier and Cox regression analyses with adjustment for age, sex and other variables were used to calculate survival rates and adjusted HRs of the revision risk for any reason. The cumulative 5-year survival rate with any revision as the end point was 95% (95% CI 91% to 98%) for monoblock cups and 97% (95% CI 96% to 98%) for modular cups ($p = 0.6$). The adjusted HR for revision of monoblock cups compared with modular cups was 2 (95% CI 0.8 to 6, $p = 0.1$). The authors concluded that there was not any clinically relevant difference in risk of revision between monoblock and modular acetabular cups in the medium term.

Australia

Luo *et al.*[314] analysed the effect of the AOANJRR on the cost of joint arthroplasty through identification of implants with higher than expected failure rates between 2003 and 2007. A total of 242,454 primary joint arthroplasties were performed in Australia at a cost of AU$4.1B. The authors state that if the poor-performing THRs had been conducted using average longevity designs, the number of THR revisions could have been reduced by 47%.

One study[315] investigated the relationship between the bearing surface and the risk of revision because of dislocation using 110,239 records in the AOANJRR from 1999 to 2007. The authors reported that 2621 (2.4%) primary THRs were revised for any reason; 862 (0.78%) THRs were revised because of dislocation. Ceramic-on-ceramic bearing surfaces had a lower risk of revision for dislocation than metal-on-polyethylene and ceramic-on-polyethylene bearing surfaces at 7 years' follow-up. The authors reported a significantly higher rate of revision for dislocation with ceramic-on-ceramic bearing surfaces than with metal-on-polyethylene bearing surfaces when smaller head sizes (≤ 28 mm) were used in younger patients (< 65 years) (HR 1.53, $p = 0.041$) and also with larger head sizes (> 28 mm) in older patients (≥ 65 years) (HR 1.73, $p = 0.016$).

Italy

Di Tanna *et al.*[261] report data from the Emilia-Romagna Regional Registry on Orthopaedic Prosthesis from 2000 to 2007. This registry collects information on all orthopaedic interventions performed in Emilia-Romagna, Italy. The study assessed the cost-effectiveness of cementless prostheses compared with hybrid prostheses in 41,199 THRs and concluded that there were differences in the revision rate and impact on costs between the two groups. The authors concluded that, considering two cohorts of 100 subjects, 243 revisions would be expected in the cementless group compared with 300 in the hybrid group. This was equal to a 19% difference and a number needed to treat of 18.

A second paper reporting on the Emilia-Romagna Regional Registry on Orthopaedic Prosthesis[320] conducted survival analysis using the Kaplan–Meier method to analyse survival rates for THRs in Italy between 2000 and 2006 (35,042 THRs, 5878 revisions). The reported cumulative survival rate for THR at 7 years was 96.8% (95% CI 96.4% to 97.1%). Multivariate analysis demonstrated that THR survival was affected by pathology, for example the presence of RA. Women comprised 66.4% of patients and > 54.0% of patients were overweight (BMI > 25 kg/m^2). Mean age at primary surgery was 66.9 years (range 16–101 years) and at revision was 70.0 years (range 22–98 years).

Finland

Eskelinen *et al.*[323] evaluated the population-based survival of cementless THR in patients aged < 55 years using data from the Finnish Arthroplasty Register. All cementless stems studied showed a survival rate of > 90% at 10 years.

Makela *et al.*[327] analysed population-based survival rates for cemented and cementless THRs in patients aged ≥ 55 years in Finland between 1980 and 2006. The 15-year survival rate for cementless THR (80%) was comparable with the rates for the cemented groups [86% in cemented group 1a (cemented, loaded-taper stem combined with a cemented, all-polyethylene cup) and 79% in cemented group 2 (a cemented, composite-beam stem with a cemented, all-polyethylene cup)] when revisions for any reason

were used as the end point. The authors concluded that both cementless stems and cementless cups, analysed separately, had a significantly lower risk of revision for aseptic loosening than cemented implants.

The same authors reported revision outcomes in primary OA.[328] The 15-year survival rate of group 1 cementless THR (implants with a cementless, straight, proximally circumferentially porous-coated stem and a porous-coated press-fit cup) performed in 1987–96 (62%, 95% CI 57% to 67%) and group 2 cementless THR (implants with a cementless, anatomic, proximally circumferentially porous-coated stem, with or without hydroxyapatite, and a porous-coated press-fit cup with or without hydroxyapatite) performed during the same time period (58%, 95% CI 52% to 66%) was worse than that of cemented THR (71%, 95% CI 62% to 80%), although the difference was not statistically significant. The risk of revision for aseptic loosening of group 1 cementless THR (0.49, 95% CI 0.32 to 0.74) was lower than that of cemented THR ($p = 0.001$).

Slovakia

One study[345] reported findings from Slovakia from 2003 to 2010, including a total of 4970 primary THRs and 457 revisions. Cement was used for all components in 35.45% of all arthoplasties, 53.25% were cementless and 11.28% were hybrids. By 2010, the revision rate reached 9.20%, representing an annual increase of 1.1%. The revision rate in the whole observed period from 2003 to 2010 was 9.15%.

Norway

Espehaug *et al.*[332] studied differences by county and regional health authority over a 20-year period (1989–2008) using data from the Norwegian Arthroplasty Register. The authors observed an increase in the number of THR operations, from 109 operations per 100,000 inhabitants in the years 1991–5 to 140 in 2006–8. Variations were found across the four regions studied.

A second study from Norway[333] reported the risks of revision after THR during a 21-year period among hip replacements reported to the Norwegian Arthroplasty Register. The risks of revision during the time periods 1993–7, 1998–2002 and 2003–7 were compared with that of the reference period 1987–92. There was an overall reduced risk of revision in the time periods 1993–7, 1998–2002 and 2003–7 compared with the risk of revision in the reference period. The improved results were due to a reduction in the incidence of aseptic loosening of the femoral and acetabular components in all time periods and in all subgroups of prostheses. The best results were obtained with the use of cemented prostheses. Analyses of revision for any cause were carried out for all prostheses together and separately for cemented, hybrid, reverse hybrid and cementless prostheses. The major cause of revision was aseptic loosening of one or both implant components.

One study used data from the Norwegian Arthroplasty Register (data from 1987 to 2008)[335] to compare the difference in risk of THR revision from infection and change in risk over time. Data was from 1987 to 2008.[333] Of the 84,492 THRs, 534 (0.6%) were revised for infection. Women had a significantly lower risk of revision for infection than men (RR 0.41, 95% CI 0.34 to 0.48). The cumulative 5-year survival rate was 99.5% in RA patients and 99.4% in OA patients (RR 0.98, 95% CI 0.65 to 1.48 for RA vs. OA patients) with revision for infection as the end point. The risk of revision for infection from 6 years postoperatively was higher in patients with RA.

USA

One study reported registry data from the USA.[337] It examined patient and surgical factors associated with deep surgical site infection (SSI) following THR using data from the Kaiser Permanente Total Joint Replacement Registry between 2001 and 2009. A total of 30,491 THRs were included in the analysis, of which 17,474 (57%) were performed on women. The incidence of SSI was 0.51% (155/30,491), equating to a total of 155 deep SSIs, which occurred at a mean of 72 days (median 28, SD 93.3 days) after the procedure. Patient factors associated with SSIs included female sex, obesity and ASA grade ≥ 3.

Multinational

Sadoghi *et al.*[350] compared primary THRs between different countries in terms of THR number per inhabitant, age and procedure type and compared survival curves including all THRs using data from nine registries. On average, the annual number of primary THRs per 100,000 inhabitants was found to be 133 for all ages, 26 for those aged < 55 years, 269 for those aged 55–64 years, 520 for those aged 65–74 years and 531 for those aged ≥ 75 years. The fixation method varied by country, for example in Sweden 67% of THRs are cemented whereas in Emilia-Romagna (Italy) 89% are cementless. Cementless fixation was more popular in Australia, Denmark, Emilia-Romagna, New Zealand and Portugal (50%) and cemented fixation was used more in Sweden and Norway (50%). Cemented and cementless fixations were used equally in England and Wales and Slovakia. The use of hybrid fixation was more uniform across countries and ranged from 8% in Portugal to 34.5% in New Zealand. Denmark showed the lowest survival rate within the first 15 years; however, THRs performed between 2006 and 2009 in Norway had similar low survival rates. All survival curves calculated in the study (except for Danish data) varied by < 1% within the first 9 years. Multivariate or subgroup analyses were not performed to compare the survival curves. The use of primary RS was not reported separately in the registries from Norway and Slovakia. Use of RS in the other countries varied from 1% in Portugal to between 2% and 3% in Denmark, Emilia-Romagna, New Zealand and Sweden to approximately 5% and 6% in Australia and England and Wales, respectively.

Graves *et al.*[347] performed an investigation of the use of metal-on-metal THRs in the National Arthroplasty Registries of Australia, England and Wales and New Zealand. All registries reported an increased revision rate associated with larger femoral head size when metal-on-metal bearing surfaces were used.

The Nordic Registry includes the joint registries of Denmark, Sweden and Norway. One study[348] aimed to compare demographics, choice of implant, fixation techniques and results between the countries, including a total of 280,201 THRs performed between 1995 and 2006. The study reported that 9596 THRs (3.4%) had later been revised. RS accounted for ≤ 0.5% of procedures in all countries. The 10-year survival rate was 92% (95% CI 91.6% to 92.4%) in Denmark, 94% (95% CI 93.6% to 94.1%) in Sweden and 93% (95% CI 92.3% to 93.0%) in Norway.

A second study reporting data from the Nordic Registry compared the survival of cemented THRs with metal femoral heads made from various materials (cobalt–chromium, aluminium and zirconium).[334] The study reported prosthesis survival and relative revision risks adjusting for age, sex and diagnosis between 1987 and 2010. In total, 132,000 cases of THR were included in the analysis. At 12 years the survival rate was 88.1% for cobalt–chromium heads and 74.8% for zirconium heads. Aluminium femoral heads provided no advantage over cobalt–chromium heads for prosthesis survival. The authors concluded that cemented polyethylene THR with aluminium heads had a similar survival rate as the same THR with ceramic-on-ceramic heads when any revision was the end point.

Summary of the total hip replacement studies

The 22 THR studies reported the analysis of registry data from nine countries. These studies examined various aspects of the THR procedure, including revision and survival rates; different implants and combinations of implant bearing surfaces; and outcome measures such as reason for failure and patient differences associated with failure. Four of the 22 THR studies used registry data from multinational databases. Sadoghi *et al.*[350] provided an extensive review of registries worldwide. They stated that fixation methods varied by country, with the cemented THR being most popular in Sweden and Norway and the cementless THR being most common in Emilia-Romagna (Italy) but also popular in Australia, Denmark, New Zealand and Portugal. Cemented and cementless fixations were used equally in England and Wales and Slovakia. In terms of survival rates, THRs carried out in Denmark showed the lowest survival rate within the first 15 years.

Core articles included in the economic model and survival analysis

The prioritisation of the eligible studies resulted in the identification of 30 papers that were deemed to be potentially useful for the economic model and survival analysis. The final number of core papers that helped to inform the survival analysis in this report was three.[15,16,318] This was in addition to the annual reports from the Swedish Arthoplasty Registry,[96] the NJR[36] and the AOANJRR,[95] which were used for comparison of survival analysis methods.

Summary of the registry evidence

Thirty papers were identified in the registry review and were included in the narrative synthesis. Eight of the studies reported registry data investigating the use of RS for the treatment of arthritis. Five of the studies combined findings in three individual countries and three studies used multinational data. The final number of THR papers included was 22. These papers reported various aspects of the THR procedure, including revision and survival rates; however, the time periods over which the analyses were carried out varied between 3 years and 15 years. Comparison of different implants and combinations of implant bearing surfaces was also conducted. Finally, additional outcome measures analysed included reason for failure (e.g. infection) and patient/demographic differences associated with failure.

Chapter 5 Individual patient data set

Introduction to individual patient data analysis

This chapter provides a narrative description of the IPD that were retrieved from the NJR and used for analysis in this report. The data set is known here as the NJR data set; data comes from the 009 data set including primary operations carried out before 1 March 2012. Any revision or notified death up to September 2012 has been included. The NJR is maintained on behalf of the Department of Health and the Welsh government. It was established in 2002 and is updated annually; data on hip and knee joint replacements were collected from April 2003. Northern Ireland joined the registry in 2013, which was after the receipt of the data.[36] Data are collected for all types of implants used in joint replacement and carried out across England and Wales. The NJR also includes data from some of the private operations carried out in independent hospitals.

Method

This is a retrospective cohort study that involves analysis of NJR data to derive time to revision of hip replacement procedures. The data provided by the NJR were divided into two types depending on type of surgery carried out: RS or THR. THR data were separated into five categories on the basis of the frequency of combinations of the components used in the procedures.

Selection of patients

Within this report THR and RS used for hip replacement procedures in England and Wales have been considered. This chapter explains the NJR data used for calculating parameter values to evaluate cost-effectiveness in the THR and RS economic models (see *Chapter 8* and *9*). For the purpose of this report and in line with the scope, information and analyses have been stratified by procedure type (THR and RS).

Structure of the database

The NJR database collects numerous variables relating to the joint affected, outcomes, procedures and implants. For the purposes of this study, 198 variables were requested from both the RS database and the THR database. The extracted data contained the following information:

- patient demographics
- provider type
- lead surgeon grade
- procedure types/patient procedure/side
- indications for primary surgery
- primary thromboprophylaxis
- primary untoward intraoperative events
- primary bone-graft usage
- all primary implant details
- current outcome type
- time from primary operation to outcome
- age at death
- any revision details – date and reasons and implants removed.

All but a few entries for 'indication' included the word 'osteoarthritis'; the few that did not were mostly entered as RA seronegative or RA seropositive. These were excluded from the analysis of time to revision.

Contents of the database

To evaluate the cost-effectiveness of hip replacement procedures in line with the scope, we requested the variables outlined in the previous section separately for the two patient groups (RS and THR):

- RS – this involves removing the damaged surfaces of bones inside the hip joint and cementing a metal surface to the reshaped bone; the socket has a metal surface and is fixed into the pelvis without using cement ($n = 31,222$ excluding RA patients)
- THR – this involves the removal of the entire damaged hip joint and replacement with an artificial joint ($n = 387,667$ including RA patients; 386,556 excluding RA patients).

Results

For statistical and economic modelling the primary outcome was time to revision.

Hip resurfacing arthroplasty

This section describes the data reported for the patients in the NJR RS data set. *Figure 20* shows the outcomes for this group of patients. Of 31,222 patients, 9339 were female and 21,883 were male. Further subdivision according to age and head size is shown in *Tables 51* and *52* (excludes RA patients).

Total hip replacement

The NJR describes the outcomes of patients undergoing THR surgery in England and Wales from April 2003 to December 2012. On date of receipt of the data (6 December 2012) the data set had a total of 387,694 records. From this number only 387,667 records were usable for one of more of the following reasons:

- irrelevant data type reported (negative age, zero age) (22 records)
- missing variable information (11 records).

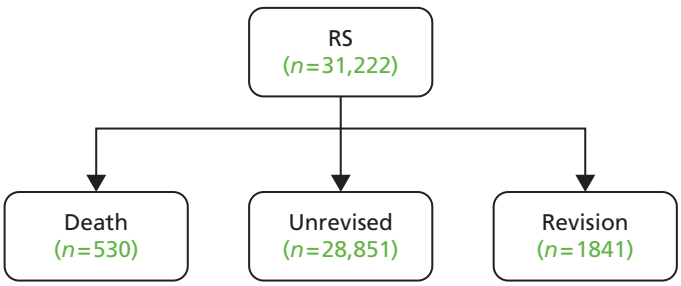


FIGURE 20 End points for all RS patients included in the analysis (excludes RA patients).

TABLE 51 Subdivision of male RS patients by age and head size (excludes RA patients)

Age group (years)	Head size (mm)													Total
	36	38	40	42	44	46	48	50	52	54	56	58	60	
15–24	0	2	0	3	0	8	2	11	4	7	1	0	0	38
25–34	0	1	0	2	7	37	44	69	28	36	6	4	1	235
35–44	0	0	2	12	30	205	300	776	311	405	41	31	0	2113
45–54	0	2	3	13	89	565	936	2516	1109	1312	164	121	3	6833
55–64	1	1	5	22	123	776	1334	3717	1519	1882	204	150	4	9738
65–74	0	0	1	9	24	206	340	1070	404	564	87	47	3	2755
75–84	0	0	1	2	3	15	11	63	20	44	2	5	0	166
85–94	0	0	0	0	0	0	2	1	1	1	0	0	0	5
Total	1	6	12	63	276	1812	2969	8223	3396	4251	505	358	11	21,883

TABLE 52 Subdivision of female RS patients by age and head size (excludes RA patients)

Age group (years)	Head size (mm)												Total
	34	36	38	40	42	44	46	48	50	52	54	58	
15–24	0	0	7	2	10	5	7	1	2	0	1	0	35
25–34	0	0	5	9	46	24	52	10	14	0	0	0	160
35–44	1	0	17	45	245	172	361	72	53	10	0	0	976
45–54	0	0	45	163	769	604	1267	240	225	22	14	1	3350
55–64	0	1	31	133	738	759	1678	355	342	20	9	1	4067
65–74	0	1	6	25	118	119	299	69	74	3	2	0	716
75–84	0	0	1	1	2	5	17	1	4	0	1	0	32
85–94	0	0	0	0	0	1	2	0	0	0	0	0	3
Total	1	2	112	378	1928	1689	3683	748	714	55	27	2	9339

The remaining 387,667 patients could have one of three outcomes (*Figure 21*):

1. death
2. unrevised THR
3. revision surgery.

Of these 387,667 patients, 240,156 (62%) were selected for analysis on the basis of the frequency of use of different THR components and of these 239,089 patients had an OA indication for surgery. Five different types of THR category were selected by looking at the frequency distribution of THR components used in the population of NJR participants using cross-tabulation.

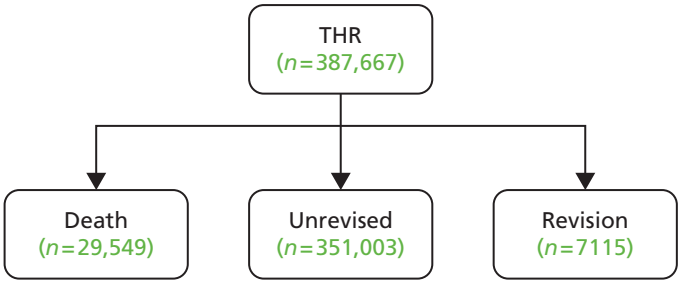


FIGURE 21 End points for all THR patients included in the analysis (includes RA patients).

Total hip replacement category development

The NJR database for non-RS procedures contained 387,694 records. After removing unusable records (this included records with missing entries and in which the primary time to outcome was negative), the database contained 387,667 useable records.

The database contained several key components of THR procedures, which were used to determine the categories that were used in the survival and cost-effectiveness analyses:

- cup component group
- cup component type
- cup composition
- cup fixation
- cup implant type
- head component type
- head composition
- liner component type
- liner composition
- stem component type
- stem fixation
- stem implant type.

We conducted two-way cross tabulations for each of the variables listed above to determine the most frequent combinations. For example, we cross-tabulated the cup component group with liner composition. We then added another component that was the most frequently occurring. For example, looking at the two-way cross-tabulation for cup component group and head composition, we know from the previous two-way cross-tabulation that the most frequent cup component group is shell, so taking this into account we then added the most frequent head composition. The next most frequent combination was then added and so forth and the process was repeated until all of the key components listed above had been taken into account.

This was an iterative process; by adding on the next most frequent combinations, we identified seven mutually exclusive categories. After consulting with our expert clinical advisor, we included four of these categories, which each accounted for > 25,000 operations. Our expert clinical advisor identified a further exclusive category ($n = 12,705$), which is a well-known option consisting of a cemented stem with a ceramic head articulating with a cemented polyethylene cup (*Figures 22* and *23*). Both the cup and stem are cheaper than cementless options and the ceramic femoral head is known to have better wear properties than the metal equivalent. Our advisor suggested that this combination is often used in younger high-demand patients because of its low wear characteristics.

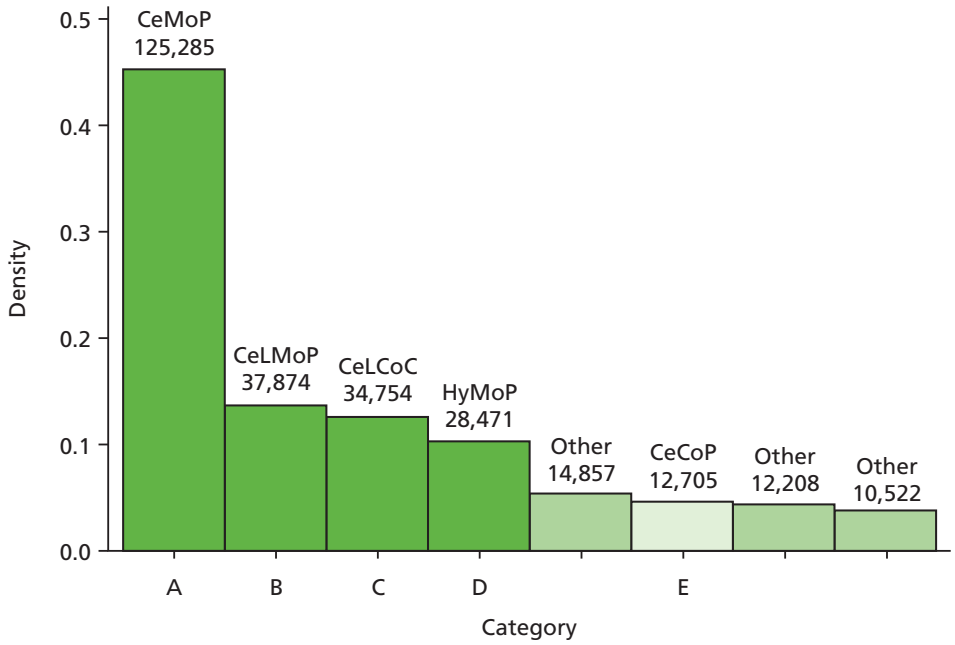


FIGURE 22 Density of patients among the most frequently used combinations of THR (RA patients excluded). Total number of patients in the eight categories = 276,676; total number of patients among the selected categories (A–E) = 239,089. Dark green shading indicates those categories selected on the basis of frequency use only; the intermediate green shading represents unselected next most frequently used combinations of THR component parts; and the light green shading indicates categories selected on the basis of clinical advice plus the frequency of use.

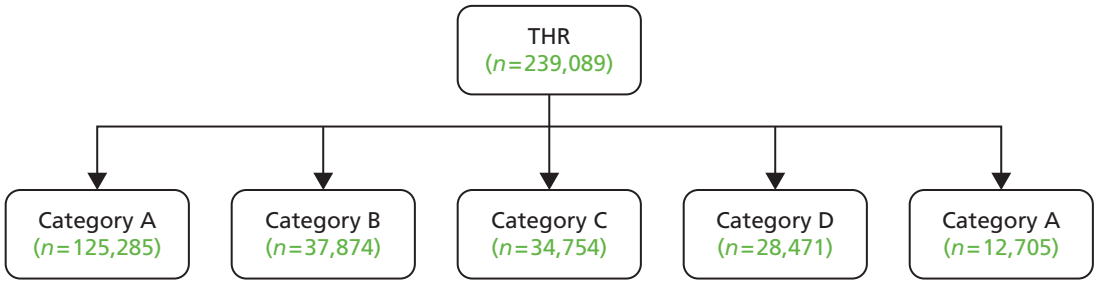


FIGURE 23 Numbers of patients in each of the selected THR categories (RA patients excluded).

Table 53 shows the final five categories that were used in the time to revision and cost-effectiveness analysis and this accounts for 239,089 patients (~62% of patients in the NJR non-RS database). Characteristics of the five THR categories are provided along with their short-form acronyms. Further information on age and sex distribution and technical characteristics of the categories is provided in *Tables 54* and *55*, respectively.

Matching

In health evaluation, data often do not come from RCTs but from (non-randomised) observational studies. Rosenbaum and Rubin[354] proposed propensity score matching as a method to reduce the bias in the estimation of treatment effects using observational data sets. Propensity matching on age and sex was undertaken using the Edwin Leuven procedure.[355]

The rationale for using propensity scores is that, because in observational studies assignment of subjects to the treatment and control groups is not random, estimation of the effects of treatment may be biased by the existence of confounding factors. Using propensity score matching is a way to adjust or correct the estimation of treatment effects, controlling as far as possible for the existence of confounding factors, and is based on the idea that bias is reduced when comparison of outcomes is performed using treated and control subjects who are as similar as possible. We used the IPD retrieved from the 009 NJR data set with primary surgery undertaken before 1 March 2012.

TABLE 53 Characteristics of the five THR categories

Category	Characteristics	Acronym used in the report
A	Metal head (cemented stem) on cemented polyethylene cup	CeMoP
B	Metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)	CeLMoP
C	Ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)	CeLCoC
D	Metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)	HyMoP
E	Ceramic head (cemented stem) on cemented polyethylene (poly) cup	CeCoP

TABLE 54 Composition of the five THR categories by age and sex

Category	Women aged > 65 years	Men aged > 65 years	Women aged < 65 years	Men aged < 65 years	Total
A	75,734	37,018	8079	4454	125,285
B	18,396	11,878	4423	3177	37,874
C	7554	6186	11,698	9316	34,754
D	15,641	8657	2649	1524	28,471
E	4655	2777	3073	2200	12,705
Total	121,980	66,516	29,922	20,671	239,089

TABLE 55 Prosthesis characteristics for the five THR categories

Category	Cup component group	Cup component type	Cup composition	Cup fixation	Cup implant type	Head component type	Head composition	Liner component type	Liner composition	Stem component type	Stem fixation	Stem implant type	Number of patients in category with OA
A	Cup	Monobloc	Polyethylene	Cemented	Cups cemented	Modular	Metal	Null	Null	Modular	Cemented	Stem cemented	125,285
B	Shell	Standard	Metal	Cementless HA coated	Cups cementless	Modular	Metal	Standard	Polyethylene	Modular	Cementless HA coated	Stem cementless	37,874
C	Shell	Standard	Metal	Cementless HA coated	Cups cementless	Modular	Ceramic	Standard	Ceramic	Modular	Cementless HA coated	Stem cementless	34,754
D	Shell	Standard	Metal	Cementless HA coated	Cups cementless	Modular	Metal	Standard	Polyethylene	Modular	Cemented	Stem cemented	28,471
E	Cup	Monobloc	Polyethylene	Cemented	Cups cemented	Modular	Ceramic	Null	Null	Modular	Cemented	Stem cemented	12,705

HA, hydroxyapatite.

We stratified data by sex (RS $n = 31,222$: 21,883 male, 9339 female; THR categories A–E $n = 239,089$: 87,187 male and 151,902 female) and matched by age within each sex stratum. From the RS group 9321 women and 17,322 men were matched by age with 9321 women and 17,322 men, respectively, from the THR group.

Analysis to match the RS and THR groups was performed using the statistical package Stata 12 Special Edition (StataCorp LP, TX, USA).

We used the Stata command 'psmatch2'.[355] We used nearest-neighbour matching using one-to-one matching by identifying the 'nearest neighbour' to each RS patient from the THR database based on closest propensity score; the variable used to construct the propensity score was age within sex strata.

In using these programs it should be kept in mind that they allow us only to reduce, and not to eliminate, the bias generated by unobservable confounding factors.

Assessment of utility and quality of the National Joint Registry for England and Wales database

This section considers the utility and quality of the data set from the perspective of the requirements of the present report. Unsurprisingly, the database structure of this resource was not tailored specifically for the task in hand. The strengths and weaknesses of the data set are briefly summarised below.

Strengths

1. The data set was comprehensive in that it contained information on all patients listed for hip arthroplasty surgery in NHS hospitals in England and Wales between April 2003 and December 2012.
2. A small number of missing variables was present (less than 0.2% for the THR database).
3. The size of the data set was large; this provides narrow CIs for survival analysis and hence more certainty in the evaluation of cost-effectiveness.
4. It was possible to distinguish between THR and RS patients.

Weaknesses

1. The elapsed time to any primary outcome was reported in years rather than number of days or by date.
2. No costs were reported for the procedures.
3. It was not possible to link patients who progressed from treatment with RS to full THR in the database.
4. Our data set was not linked by revision surgery.
5. There was very poor reporting of BMI.
6. There was no linkage to the PROMs data set in our data.

Summary of the individual patient data set

The NJR provides valuable information about patient subgroups and the categorisation of hip replacement procedures for all patients receiving treatment in the NHS in England and Wales. There were insufficiently complete data to estimate linked primary and secondary surgery for each patient or costs or utilities associated with the procedures.

Subsequent chapters describe further analysis of this data set in the cost-effectiveness model.

Chapter 6 Patient-reported outcome measures

Quality of life and utilities

Background

This section provides a brief description of the PROMs data set, which was used to provide utility data for analysis in the Markov model. We obtained quality-of-life data from the PROMs data set for patients who had a THR between January 2009 and December 2012.[97] The variables in the data set included the following: PROMs ID, patient sex, patient death, surgery date, complications (e.g. bleeding, infection and wound problems, readmission, further surgery) and EQ-5D-3L data, which was completed 6 months after surgery.

The EQ-5D-3L is a generic health-related quality of life measure that comprises the following five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels of scoring: no problems, some problems or severe problems. This creates 243 possible health states, to which unconscious and dead have been added, giving a total of 245 health states. These health states are then converted to an index score from 0 (dead) to 1 (perfect health).

Methods

Two analyses of the PROMs data set were undertaken.

Analysis 1

The PROMs data set for patients who had a THR between January 2009 and December 2012 included 207,436 records. After removing records with missing EQ-5D scores or surgery dates the data set contained 117,044 records. No age-specific utilities by sex were available in this data set.

Analysis 2

A second PROMs data set containing EQ-5D-3L data for THR by age and sex for the year 2010/11 was downloaded from the NHS Information Centre website in March 2013 (www.ic.nhs.uk/catalogue/PUB07049) for further analysis. This data set included 38,378 records. After removing patients with missing information with regard to EQ-5D scores, sex and age category, and after excluding patients aged < 40 years, the data set contained 32,577 records.

Overall

For both analyses, mean EQ-5D index results including SDs and 95% CIs were calculated. Linear regression analyses were conducted for EQ-5D index scores by sex for the different age categories. All statistical analyses were conducted in Stata 12.

Results

For all patients, the mean EQ-5D score after their hip operation was 0.767 (*Table 56*). Men had a slightly higher EQ-5D utility index score than women (0.787 vs. 0.753).

TABLE 56 European Quality of Life-5 Dimensions utility index scores for patients who completed the EQ-5D-3L questionnaire after THR (analysis 1)

	All patients	Men	Women
n	117,044	47,745	68,676
Mean (SD)	0.767 (0.256)	0.787 (0.253)	0.753 (0.257)
95% CI	0.765 to 0.768	0.785 to 0.790	0.751 to 0.754

Table 57 shows that the mean EQ-5D utility score for patients who required further surgery after hip replacement was 0.575 for men and 0.553 for women.

Table 58 shows the EQ-5D results for patients after surgery for the period 2010/11 by sex and age band. Overall, men had a slightly higher EQ-5D utility index score than women after their hip operation for all age bands. Men in the age band 60–70 years gave a slightly higher value to their health-related quality of life than men in any other age band; likewise, women in the age band 60–70 years gave a slightly higher value to their health-related quality of life than women in any other age band.

Summary of patient-reported outcome measure data

The PROMs data set has provided valuable EQ-5D data by age and sex for patients who have undergone a THR for use in the economic model. However, there were insufficient linkage data to link the PROMs data set to the NJR data set.

TABLE 57 European Quality of Life-5 Dimensions index scores for patients who completed the EQ-5D-3L questionnaire after THR and who required further surgery (analysis 1)

	All patients	Men	Women
n	3096	1320	1776
Mean (SD)	0.562 (0.341)	0.575 (0.352)	0.553 (0.332)
95% CI	0.550 to 0.574	0.556 to 0.594	0.537 to 0.568

TABLE 58 European Quality of Life-5 Dimensions index scores for patients who completed the EQ-5D-3L questionnaire after THR by age band and sex (analysis 2)

Age band	All patients	Men	Women
40–50 years			
n	794	316	478
Mean (SD)	0.726 (0.297)	0.736 (0.319)	0.720 (0.282)
95% CI	0.706 to 0.747	0.700 to 0.771	0.695 to 0.746
50–60 years			
n	4352	1883	2469
Mean (SD)	0.753 (0.287)	0.767 (0.287)	0.742 (0.286)
95% CI	0.744 to 0.761	0.754 to 0.780	0.731 to 0.753
60–70 years			
n	11,106	4758	6348
Mean (SD)	0.779 (0.259)	0.792 (0.261)	0.769 (0.257)
95% CI	0.774 to 0.784	0.784 to 0.799	0.763 to 0.775
70–80 years			
n	12,308	4841	7467
Mean (SD)	0.764 (0.246)	0.790 (0.235)	0.747 (0.251)
95% CI	0.759 to 0.768	0.783 to 0.797	0.741 to 0.752
80–90 years			
n	4017	1234	2783
Mean (SD)	0.721 (0.253)	0.745 (0.249)	0.710 (0.254)
95% CI	0.713 to 0.729	0.731 to 0.759	0.701 to 0.720

Chapter 7 Methods for modelling revision rates

Introduction

This section describes methods used for modelling revision rates to feed into the economic model. Revision rates found, the justification for using subgroups and findings by age and sex subgroups are included. We also compare here our findings with the previous benchmark generated from NICE TA 2 and TA 4.

Data were extracted from the NJR database (see *Chapter 5*) and patient cohorts were analysed for time to revision. Kaplan–Meier and competing risk analysis were implemented in Stata 11. For Kaplan–Meier analysis, non-revision by end of follow-up and death were censored; for competing risk analysis, the competing risk was death and the risk of interest was revision according to the Statauser-written routine.[356]

Kaplan–Meier analyses were fitted with parametric distributions to allow for extrapolation beyond the observed data. Following the NICE Decision Support Unit (DSU) recommendation (see www.nicedsu.org.uk/Methods-Development(1985316).htm; accessed August 2014), the IPD were fitted with Weibull, Gompertz, log-logistic, log-normal and gamma distributions using the 'streg' command in Stata. It was found that for most cohorts of patients these commonly used distributions predicted decreasing hazard for revision beyond the observed data. As decreasing hazard is unlikely to capture increasing likelihood of revision from wear and tear, particularly for those who are active or of young age, further alternative models (bathtub, Rayleigh and Mitscherlich) were explored to allow for increasing hazard of revision beyond the observed data. An initial analysis of these was carried out using ordinary least squares in Stata 11 or Excel (2010; Microsoft Corporation, Redmond, WA, USA). The Rayleigh model predicts a linearly increasing hazard, the bathtub model a U-shaped hazard and the Mitscherlich model a hazard that increases at a decreasing rate with time to reach an asymptote:[357,358]

$$\text{Rayleigh:} \quad h = a + 2bt \tag{1}$$

$$\text{Bathtub:} \quad h = at + \frac{b}{(1+gt)} \tag{2}$$

$$\text{Mitscherlich:} \quad h = \pi - b\ \exp(lt), \tag{3}$$

where π, a, b, g and l are constant parameters and t is time.

In practice, the Mitscherlich and Rayleigh models generated poor fits and were not pursued. The results from the Weibull, Gompertz, log-logistic, log-normal and bathtub models for each cohort are catalogued in *Appendix 17*, which presents modelled time to revision and hazard for the observation period and for extrapolation to 50 years.

The selection of an appropriate model or models for use in the economic analysis was based on the Akaike information criterion (AIC), judgement of the plausibility of the resulting extrapolations, visual goodness of fit to the IPD-derived Kaplan–Meier plot, and plots of the log-Kaplan–Meier estimated cumulative hazard compared with the log-modelled cumulative hazard.[359] In sex-stratified sensitivity analyses parametric fits were adjusted for age, with age for each cohort centred near the mean. The bathtub models were analysed using the Stata 'stgenreg' package developed by Crowther and Lambert.[360] This provided considerable advantages including the use of IPD, adjustment for age, prediction of hazard and survival and generation of AIC estimates for comparison with other models and of a covariance matrix of parameters that could be employed for probabilistic economic analysis. Flexible parametric models of Royston–Parmar were implemented using the 'stpm2' package in Stata developed by Lambert and Royston.[361,362]

Revision rates

Categories of total hip replacement

We considered five separate categories of THR, which differ from each other with regard to the characteristics of the component parts of each type of prosthesis. The main features of these five categories are detailed in *Table 53*.

Patient populations to be compared

The remit from NICE for this report (see http://guidance.nice.org.uk/TA304; accessed August 2014) specified the following comparisons in people with pain and disability resulting from arthritis of the hip for which non-surgical management has failed:

- different types of primary THR compared with hip RS for people in whom both procedures are suitable
- different types of primary THR compared with each other for people who are not suitable for hip RS.

We considered five separate categories of THR, which differ from each other with regard to the characteristics of the component parts of each prosthesis category. The derivation and main features of these five categories are detailed in *Chapter 5*. The five categories account for ≈62% of all NJR THR recipients.

We used NJR data to investigate revision rates. *Figure 24* shows the age distribution, according to decade, of NJR patients who received THR or RS and the age distribution by sex for those who received RS.

Most RS patients were aged < 65 years at the time of the intervention whereas most THR recipients were aged > 65 years. *Figure 25* is a Kernel density diagram showing the overlap between the two distributions. We found that populations undergoing RS or THR overlapped substantially (for RS 89.7% were aged < 65 years and for all THR categories 22.6% were aged < 65 years).

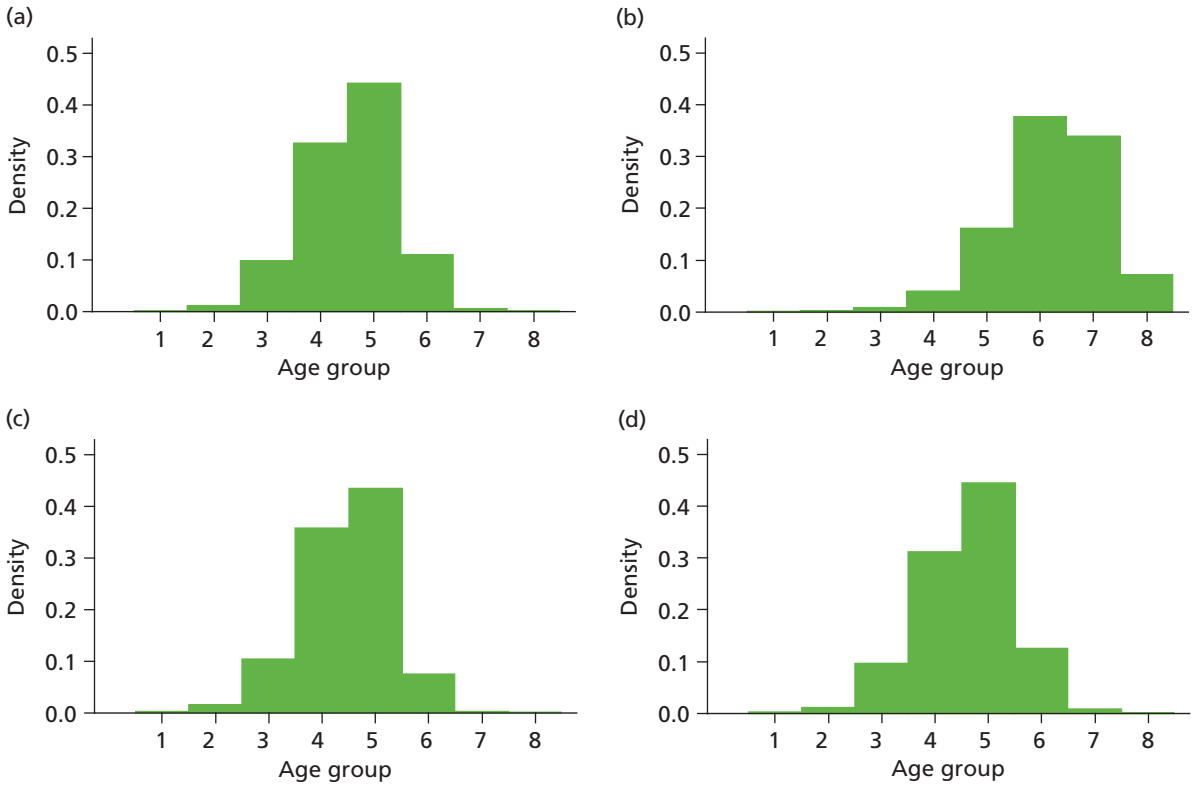


FIGURE 24 Age distribution of NJR patients who received THR or RS and age distribution by sex for those who received RS. (a) RS; (b) THR; (c) RS, female patients; and (d) RS, male patients. Age groups: 1, < 25 years; 2, 25–34 years; 3, 35–44 years; 4, 45–54 years; 5, 55–64 years; 6, 65–74 years; 7, 75–84 years; 8, ≥ 85 years.

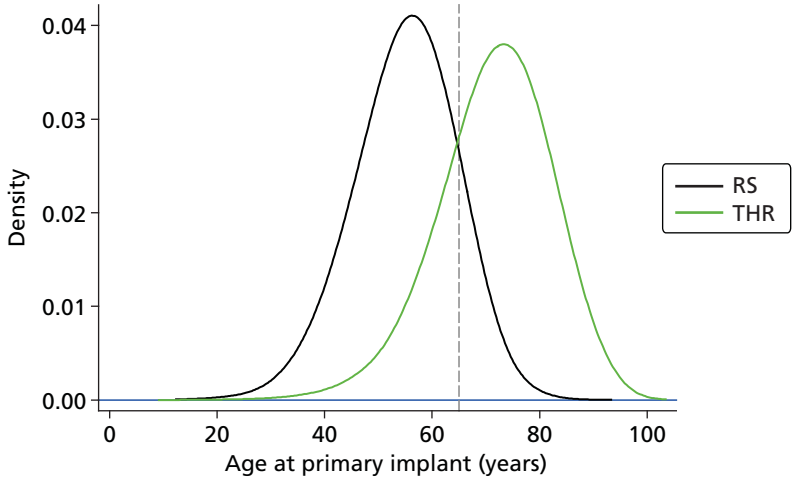


FIGURE 25 Kernel density diagram of the two distributions.

Table 59 summarises the age and sex differences between the population who received RS and the population who received THR. THR interventions outnumbered RS interventions by more than 10 : 1, the proportion of women was twice as large for THR as it was for RS and the mean age of RS recipients was about 15 years less than that for THR recipients.

To compare RS with THR we needed to define patients who were eligible for both interventions. The NJR did not contain information indicating which patients were suitable for both THR and RS, nor was there information on those who might be considered unsuitable for RS. Expert clinical opinion indicated that RS was selected mainly for relatively active younger patients whereas THR was the predominant option for less active older patients. However, the NJR did not provide information on activity levels of patients.

The literature indicates that revision rates after RS are much higher for women than for men,[15] whereas for THR the reverse is the case, a finding that we confirmed in our preliminary analysis (see *Appendix 17*). It is known that revision rates in general are lower for older patients. Because revision rates differ by sex and age it is likely that the cost-effectiveness of interventions will reflect the age and sex mix of the population(s) examined. Given the observed differences in age and sex for RS and THR populations, the following alternative strategies were considered to identify appropriate RS and THR populations for comparison of the interventions:

1. all RS recipients compared with all THR recipients, not matched
2. all RS recipients compared with recipients of the five identified THR categories, not matched (see *Chapter 5*)
3. all RS recipients compared with each of the (different 16+) categories of THR in the NJR data set, separately matched by age and sex
4. all RS recipients compared with THR recipients from each of the five identified categories, separately matched by age and sex
5. all RS recipients compared with all THR recipients from the combined five identified categories, matched by age and sex
6. all RS recipients compared with the total pool of all THR recipients, matched by age and sex.

Options 1 and 2 (without matching) were rejected because of the large age and sex differences between RS recipients and THR recipients; these imbalances influence revision rates and were judged likely to result in an inequitable comparison of the interventions. Options 3–6 avoid age and sex mismatch if age matching is undertaken separately for each sex and then the matched male and female populations combined. Age matching within sexes was in general feasible because of the much larger number of THR recipients than RS recipients. Therefore, we judged options 3–6 to be preferable to options 1 and 2.

TABLE 59 Age and sex of RS and THR recipients

Population	Number	% female	Mean (SD) age (years)	Median age (years)	Interquartile range (years)
All RS recipients	31,222	29.9	55.0 (8.6)	55.7	49.7–60.9
All THR recipients	386,556	61.4	69.5 (10.3)	70.4	63.2–76.8
THR category A–E recipients	239,089	63.5	71.6 (9.6)	72.5	65.8–78.3

Option 3 was considered impractical because of the large number of different THR interventions in the NJR database. Also, for options 3 and 4 the number of recipients within some individual THR categories was too small to allow age and sex matching with a significant proportion of RS recipients. Furthermore, expert clinical advice indicated that the relevant clinical decision was between RS and THR rather than between RS and any one of many THR options and therefore options 3 and 4 were considered less appropriate than options 5 and 6.

For these two important reasons we therefore selected option 5 for the base case. This represents a departure from the comparison specified in the protocol and scope. We selected option 5 to represent the most likely clinical comparison (the selection of THR prosthesis for a patient eligible for both RS and THR is likely to be from the most frequently used prostheses with the lowest revision rates, as represented by the five identified THR categories) (see *Figure 23*).

We therefore used propensity matching to match NJR patients with RS patients for decision problem 1 (see *Chapter 2*). Propensity matching on age and sex was undertaken using the Edwin Leuven procedure.[355]

The comparison of revision rates among these matched individuals was used in the economic analysis.

We undertook subgroup analyses in which the comparison between RS and THR was examined separately for each sex, within which parametric models of revision were controlled for age. Revision rates were then estimated for men and women aged 40, 50 and 60 years. These ages were selected to avoid extremes in the age distribution of patients while capturing age-dependent differences that may exist in revision rates. There were three reasons for undertaking subgroup analyses: (1) the difference between the sexes in mechanical load bearing through the hip joint;[363] (2) the large difference in observed revision rates between men and women (see *Figures 31* and *46*); and (3) expert clinical opinion, which indicated that age represents a reasonable proxy for activity levels.

In the selection of alternative interventions to address our objective (2) (comparison between different types of THR), we were guided by the frequency of use of different prostheses and by clinical advice (see *Chapter 5*). The wording of the scope required identification of THR recipients unsuitable for RS. However, the NJR did not provide information about which THR recipients were unsuitable for RS. Although it can be assumed that all RS patients may also be candidates for THR, the reverse is less likely. The majority of NJR THR recipients were aged > 65 years (see *Figure 24*), consistent with expert clinical opinion that older patients would be more likely candidates for THR than RS. Furthermore, the observed high revision rates that follow RS[15,16] imply that in future fewer younger patients (aged < 65 years) will be considered to be candidates for both procedures. Therefore, for the base case we took the decision to compare THR categories across the whole population who received them (irrespective of age and sex).

However, because of the wide age range of patients who received a THR, and the different proportions of men and women receiving the different types of THR, we conducted sensitivity analysis controlling for age and sex. In addition, as only ≈10% of RS recipients were aged > 65 years, it appears that patients over this age are unlikely to be suitable for RS.

We therefore conducted subgroup analyses in which the THR populations were stratified by age (> 65 years or < 65 years) and examined separately by sex. Parametric models for revision in these subgroups were controlled for age and then revision rates were estimated for men and women aged 40, 50 and 60 years using the population aged < 65 years, and for men and women aged 70 and 80 years using the population aged > 65 years. The ages were selected to avoid extremes in the age distribution of patients while capturing age-dependent differences that may exist in revision rates.

The use of subgroups described above is consistent with NICE consultations for the update of its previous technology assessments of hip replacement interventions (TA2[46] and TA44[25]), which recommended, should evidence allow, that different interventions should be compared in subgroups of patients according to age and sex.[364] However, these subgroup analyses represent an extension from our protocol and scope. *Table 60* summarises the make-up of the THR populations by age and sex.

Overall revision rates, competing risks and rationale for analysis

Revision rates among NJR patients have been the subject of several recent publications.[15,16,318,353] Some investigators have used Kaplan–Meier analysis whereas others have employed competing risk analysis in which the event of interest is revision and death is taken as a competing risk. In Kaplan–Meier analysis death, as well as no revision at the end of follow-up, is censored. We briefly compared overall revision rates in our NJR RS and THR patients according to these methodologies (see *Appendix 17* for results). RS revision rate estimates were very similar for both Kaplan–Meier and competing risk analyses and were

TABLE 60 Age and sex characteristics of patient groups receiving THR prostheses

Population	Number	% female	Mean (SD) age (years)	Median age (years)	Interquartile range (years)
All THR recipients	386,556	61.4	69.5 (10.3)	70.4	63.2–76.8
All THR female recipients	237,436	100	70.2 (10.3)	71.1	63.8–77.6
All THR male recipients	149,120	0	68.45 (10.3)	69.4	62.3–75.6
All THR category A–E recipients	239,089	63.5	71.6 (9.6)	72.5	65.8–78.3
All THR category A–E female recipients	151,902	100	72.1 (9.6)	73	66.4–78.9
All THR category A–E male recipients	87,187	0	70.5 (9.6)	71.5	64.9–77.1
All category A recipients	125,285	66.9	74.6 (7.9)	74.9	69.7–80
All category B recipients	37,874	60.2	71.5 (8.7)	72	65.9–77.5
All category C recipients	34,754	55.4	61.6 (9.9)	62.3	55.9–67.9
All category D recipients	28,471	64.2	73.0 (8.3)	73.4	67.8–78.7
All category E recipients	12,705	60.1	66.2 (9.6)	66.3	60.7–72.5
All category A male recipients	41,472	0	73.9 (7.7)	74.2	69.2–79.0
All category B male recipients	15,055	0	70.9 (8.6)	71.6	65.6–76.7
All category C male recipients	15,502	0	61.6 (9.8)	62.5	56–67.9
All category D male recipients	10,181	0	72.5 (8.1)	72.9	67.6–77.9
All category E male recipients	4977	0	65.5 (9.4)	65.6	60.3–71.6
All category A female recipients	83,813	100	74.9 (8.0)	75.3	70.0–80.5
All category B female recipients	22,819	100	71.8 (8.8)	72.3	66.2–78
All category C female recipients	19,252	100	61.6 (9.9)	62.2	55.8–67.9
All category D female recipients	18,290	100	73.3 (8.5)	73.7	67.9–79.2
All category E female recipients	7728	100	66.7 (9.7)	66.8	60.9–73.1

similar to those reported by Smith *et al.*[15] For THR the Kaplan–Meier analysis generated somewhat higher rates of revision than the competing risk analysis.

Both Kaplan–Meier- and competing risk-estimated revision rates were higher for women than for men for RS whereas revision rates for women were less than those for men for THR. For this reason some sensitivity analyses in the economic analyses that follow have been stratified according to sex. To be consistent with all previous economic analyses of hip replacement technologies, we have used the revision estimates from Kaplan–Meier analysis together with parametric modelling to predict the rate of revision beyond the observed data.

In practice, several parametric models fitted the observed data for revision well. On extrapolation, models generated quite different revision rates, mainly determined by the different modelled hazard during the extrapolation period, with some models predicting an increasing hazard (e.g. bathtub) and others a decreasing hazard (e.g. log-normal); an example is shown in *Figure 26*. An increasing hazard of revision

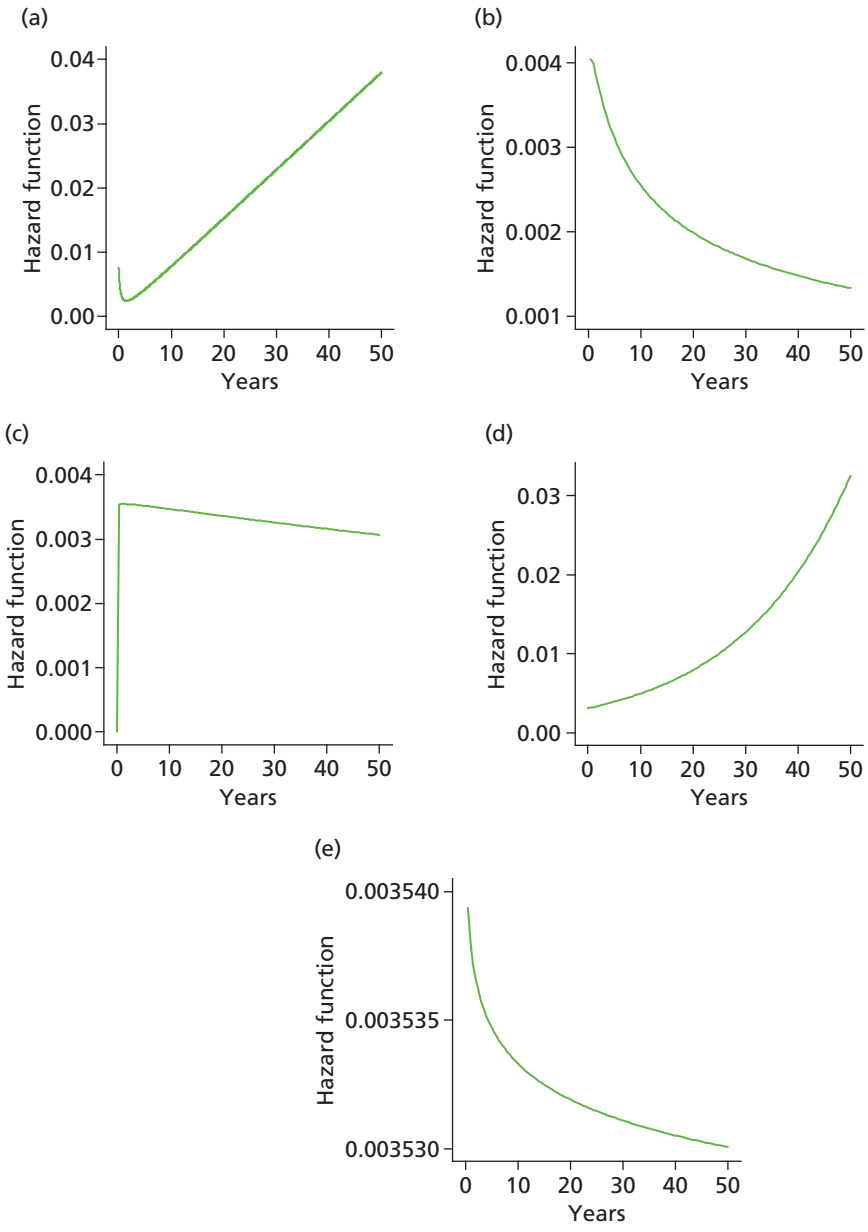


FIGURE 26 Differing modelled hazard on extrapolation beyond observation for THR HyMoP (category D) female patients aged < 65 years. (a) Bathtub model; (b) log-normal model; (c) log-logistic model; (d) Gompertz model; and (e) Weibull model.

appears reasonable for 'younger' patients who are likely to outlive their prosthesis; however, it is clear that for patients of advanced age there is a relative lack of clinical imperative to undertake revision and an extrapolation with increasing hazard becomes less appropriate (*Figure 27*).

In view of these considerations, in the base-case analysis we selected the best fit to the observed data across all of the interventions that we compared. Because in practice the best fit was provided by the bathtub model (increasing hazard on extrapolation), sensitivity analyses were conducted with the best alternative fit that allowed for a decreasing extrapolated hazard. In subgroup analyses according to age and sex a dual approach was adopted in which increasing and decreasing extrapolated hazards were both investigated.

In principle, our approach conforms to NICE DSU guidance for modelling time-to-event IPD. This guidance, however, specifically refers to interventions compared within a single clinical trial and recommends that it is desirable to adopt the same parametric form for the interventions being compared.[365,366] The NJR comprises observational rather than RCT data so parametric fits for different interventions and/or patient groups may not be well described by a single parametric form.

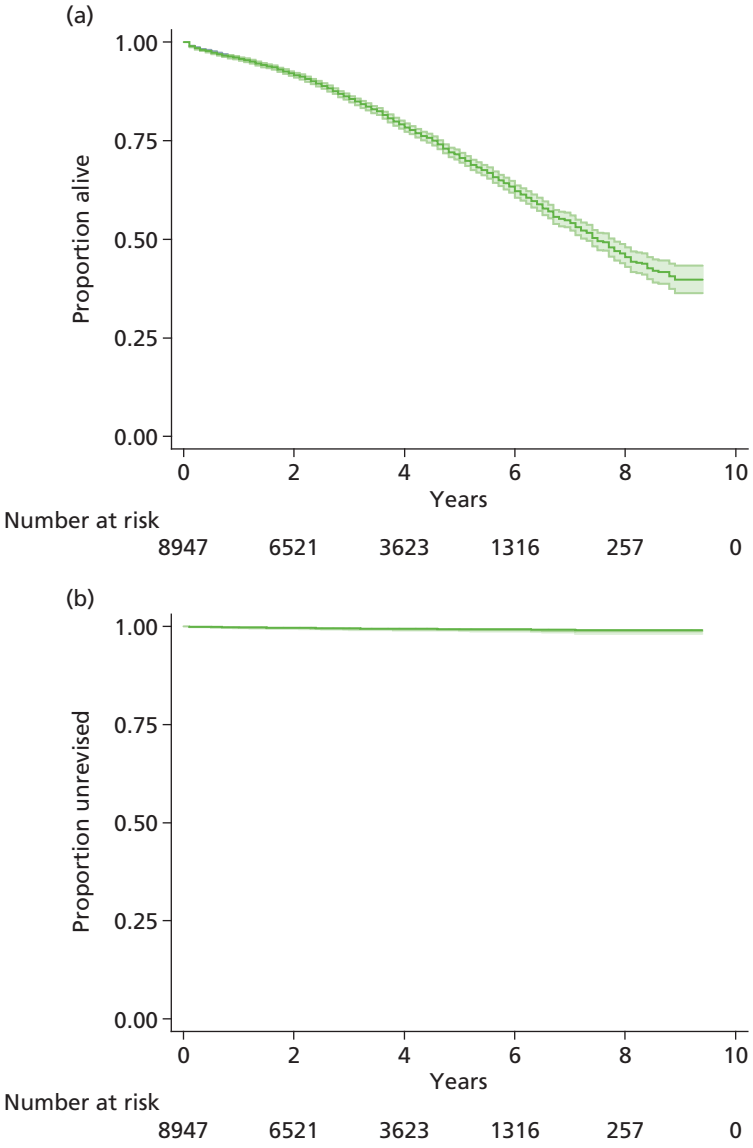


FIGURE 27 Kaplan–Meier analysis for (a) death; and (b) revision for THR CeMoP (category A) female patients aged > 85 years.

Published cost-effectiveness analyses of THR have predominantly adopted a bathtub hazard model for revision rates.[38,44,273,367]

Information criteria [AIC, Bayesian information criterion (BIC)] scores for modelled fits and plots of the modelled log-cumulative hazard compared with the log-Kaplan–Meier estimated hazard were used to judge goodness of fit and are provided in the main text or in *Appendices 19* and *20* respectively.

Results

The parametric modelling results are reported in full in *Appendix 17*.

Proportional hazards tests

The condition of proportional hazards between observed revision rates for compared groups was examined using log-Kaplan–Meier estimated cumulative hazard compared with log-time. The results for RS compared with THR and for the five categories of THR prostheses are shown in *Figures 28* and *29*, respectively.

Cumulative hazard plots for women for the comparison between RS and THR are not parallel (see *Figure 28*); this held also for THR categories when the population was stratified by sex and age (see *Figure 29*). Because there was a lack of general support for proportional hazards for most comparisons, separate models were fitted for each comparison rather than using treatment as a covariate.

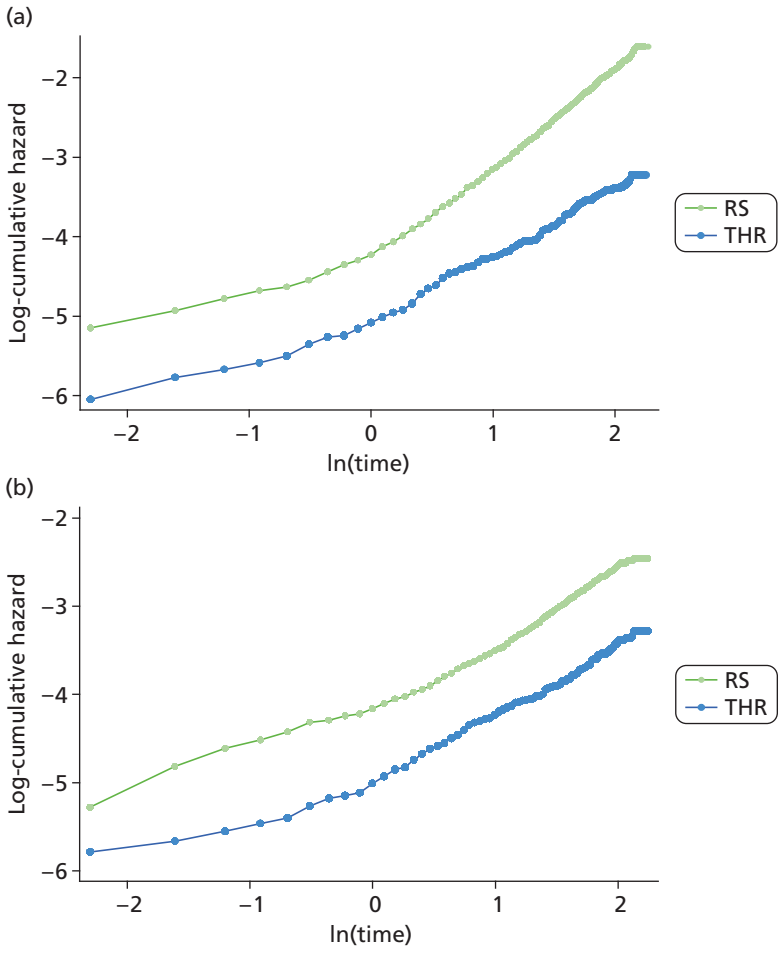


FIGURE 28 Log-Kaplan–Meier estimated cumulative hazard vs. log-time for different THR categories.

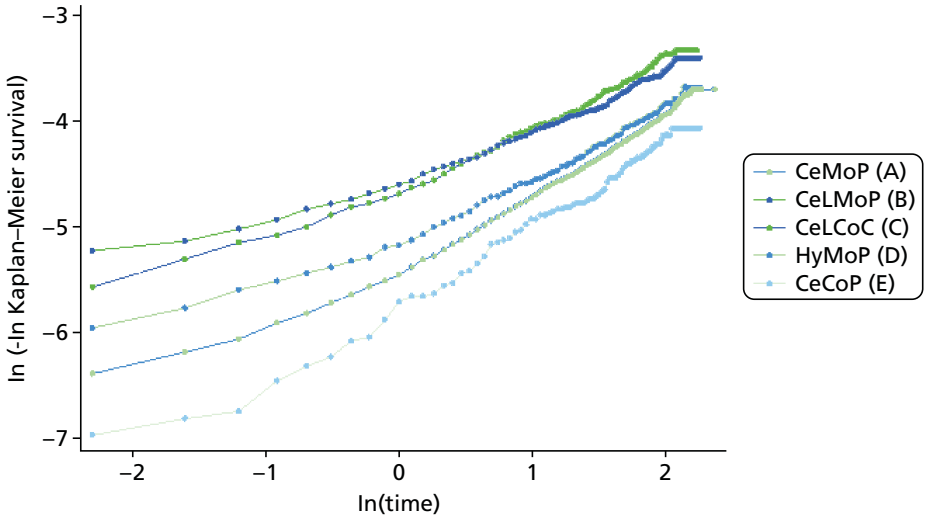


FIGURE 29 Log-Kaplan–Meier estimated cumulative hazard vs. log-time for different THR category populations stratified by sex and age.

For men for the comparison between RS and THR a proportional hazards assumption appears to hold moderately well. For the comparison between different THR prostheses, again the cumulative hazards were not noticeably parallel (see *Figure 28*); this held also for THR categories when the population was stratified by sex and age. As there was a lack of general support for proportional hazards for most comparisons, separate models were fitted for each comparison rather than using treatment as a covariate.

Comparison of resurfacing arthroplasty with total hip replacement

For both sexes many more patients received THR than RS. The observed revision rate for all RS recipients ($n = 31{,}222$) over 9 years of follow-up was about three times that for all THR recipients ($n = 386{,}556$) (*Figure 30*). When the comparison was made by sex the observed revision rate for female RS recipients was more than three times that of female THR patients and the observed revision rate for male RS recipients was about twice that for male THR recipients (*Figure 31*).

When the comparison between RS and THR was restricted to THR recipients of the five prosthesis categories A–E ($n = 239{,}089$), the differences were larger (*Figure 32*) and again held across sexes

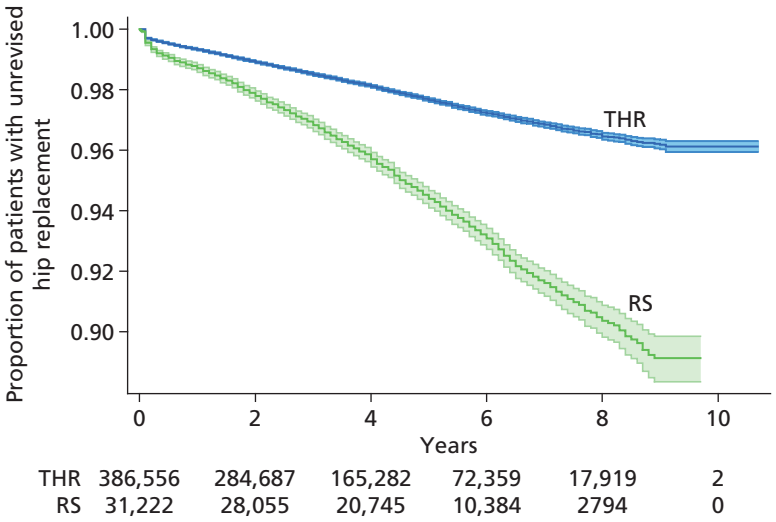


FIGURE 30 Time to revision: all RS patients and all THR patients. Numbers under *x*-axis are numbers at risk.

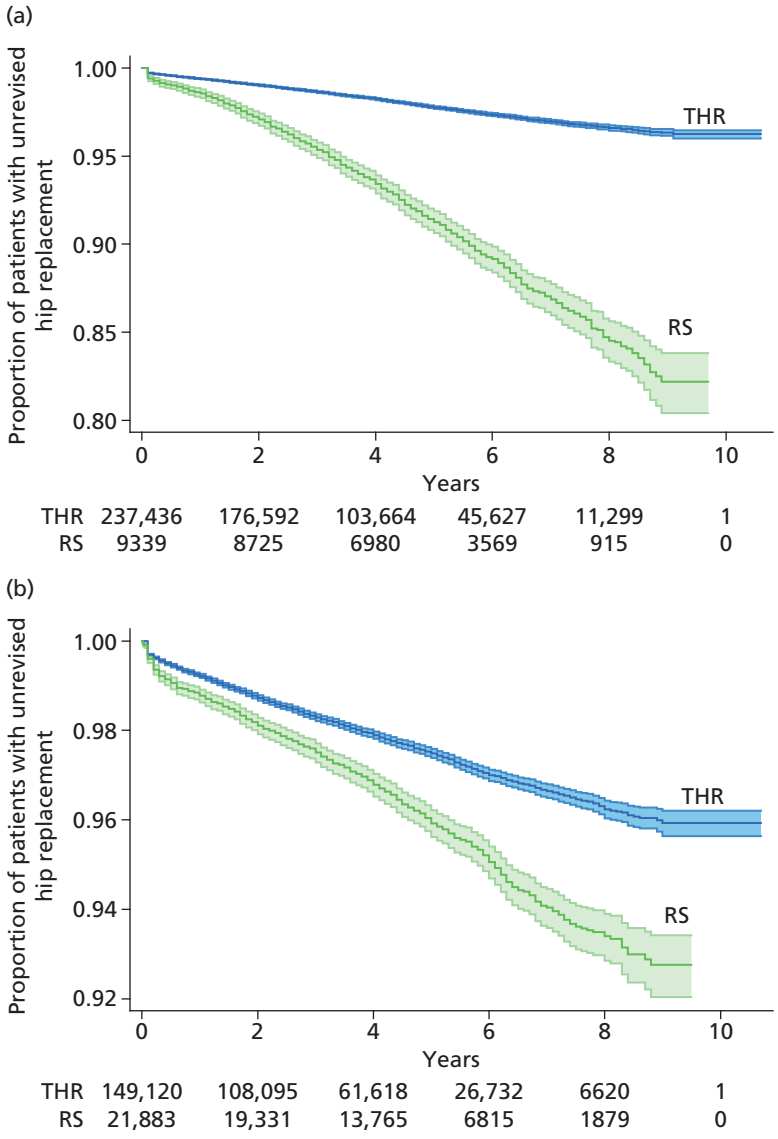


FIGURE 31 Time to revision: all RS and all THR patients according to sex. (a) All female RS patients vs. all female THR patients; and (b) all male RS patients vs. all male THR patients. Numbers under *x*-axis are numbers at risk.

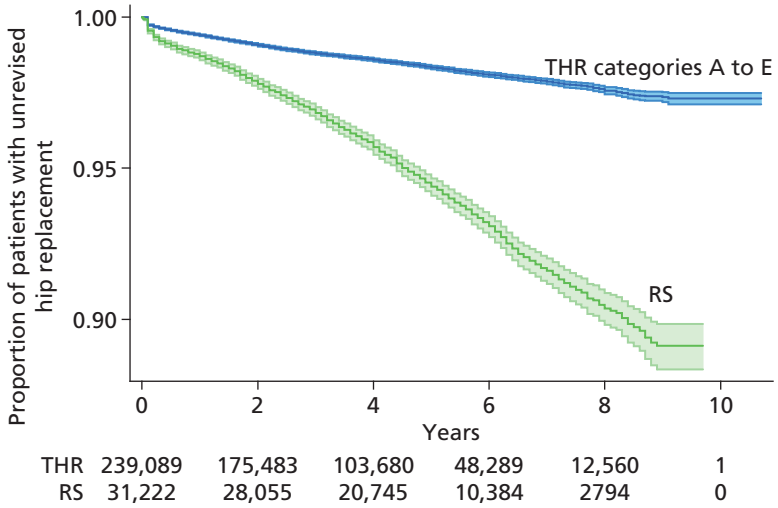


FIGURE 32 Time to revision: all RS and all THR patients (categories A–E). Numbers under *x*-axis are numbers at risk.

(*Figures 33–35*). When revision rates for recipients of the individual categories of THR were compared with all RS recipients the observed revision rates for both sexes were considerably higher for RS than for any single THR category.

It is clear that revision rates after RS are much higher for both sexes than those after THR of any category. However, age and sex differences between the RS and the THR populations (*Table 61*) make these comparisons inequitable. More men than women received RS whereas more women than men received THR, and nearly all RS recipients were aged < 65 years (mean age ≈56 years) whereas most THR recipients were aged > 65 years (mean age ≈72 years). For an equitable comparison of the interventions it is necessary to match populations by sex and age.

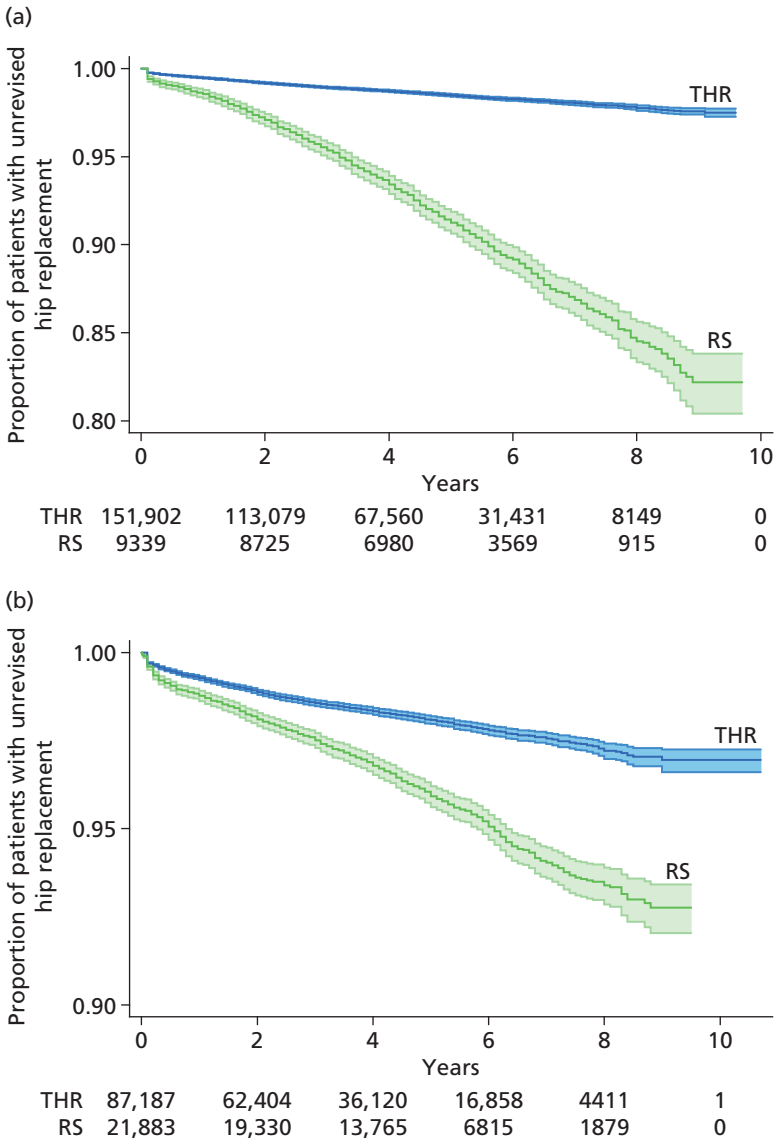
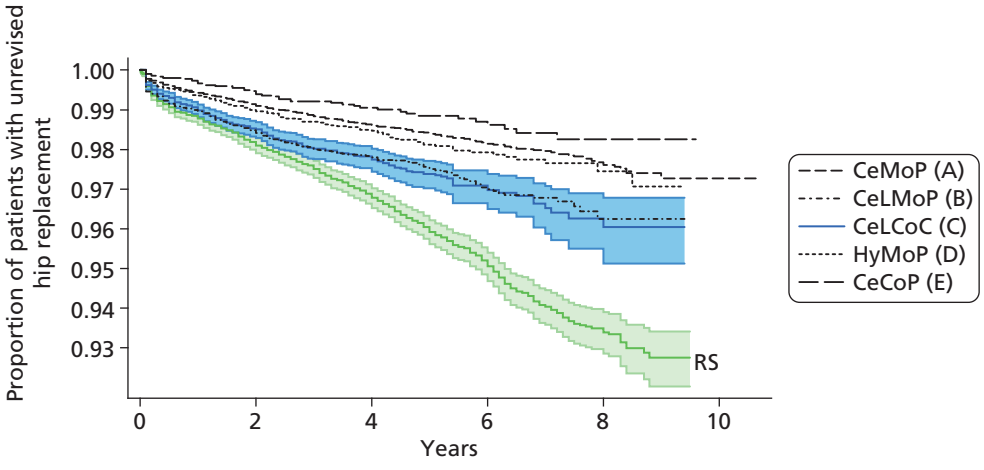
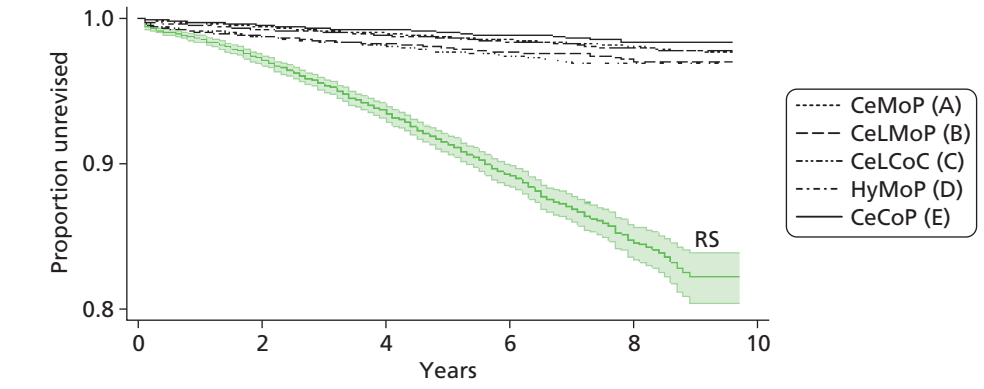


FIGURE 33 Time to revision: all RS and all THR patients (categories A–E) by sex. (a) All female RS patients vs. all female THR category A–E patients; and (b) all male RS patients vs. all male THR category A–E patients. Numbers under *x*-axis are numbers at risk.



RS	21,883	19,330	13,765	6815	1879	0
CeLCoC	15,502	9689	4205	1635	428	0
HyMoP	10,181	7352	4188	1883	446	0
CeLMoP	15,055	10,216	5149	2189	458	0
CeMoP	41,472	31,551	20,343	9980	2783	1
CeCoP	4977	3596	2235	1171	296	0

FIGURE 34 Time to revision: all male RS and all male THR patients (categories A–E). Numbers under *x*-axis are numbers at risk (THR upper, RS lower).



RS	9339	8725	6980	3569	915	0
CeLCoC	19,252	12,524	5602	2011	463	0
HyMoP	18,290	13,420	7626	3334	764	0
CeLMoP	22,819	16,012	8039	3385	700	0
CeMoP	83,813	65,405	42,796	20,927	5727	0
CeCoP	7728	5718	3497	1774	495	0

FIGURE 35 Time to revision: all female RS and all female THR patients (categories A–E). Numbers under *x*-axis are numbers at risk (THR upper, RS lower).

TABLE 61 Age and sex mix of RS and THR populations

Population	Number	% female	Mean (SD) age (years)	Median age (years)	Interquartile range (years)
All RS recipients	31,222	29.9	55.0 (8.6)	55.7	49.7–60.9
All THR recipients	386,556	61.4	69.5 (10.3)	70.4	63.2–76.8
THR category A–E recipients	239,089	63.5	71.6 (9.6)	72.5	65.8–78.3
RS propensity-matched population	26,643	35.0	55.83 (8.3)	54.0	49–59
THR propensity-matched population	26,643	35.0	55.83 (8.3)	54.0	49–59
RS propensity-matched population male	17,322	0	57.1 (8.03)	58	53–62
THR propensity-matched population male	17,322	0	57.1 (8.03)	58	53–62
RS propensity-matched population female	9321	100	53.5 (8.4)	54.0	49–59
THR propensity-matched population female	9321	100	53.5 (8.4)	54.0	49–59

Of the male and female patients who received RS for OA, 17,322 and 9321, respectively, were successfully propensity matched by age with THR patients from THR categories A–E ($n = 239,089$), providing 26,643 matched pairs for comparison (see *Chapter 5, Matching* and *Figure 23*). Age distribution was identical in the RS and THR matched populations (see *Table 61*) but was slightly skewed from normal (*Figure 36*). Kaplan–Meier analysis (see *Figure 36*) revealed that revision rates were much higher for RS than for the matched THR population.

Revision was more frequent among the matched THR population than among the whole THR population (*Figure 37*), demonstrating the importance of the matching process before comparison of RS with THR.

Information criteria (*Table 62*) indicated that bathtub models provided the best fit for both RS and THR, shown in *Figure 36*. Therefore, to compare RS with THR in the base-case economic analysis, transition probabilities were calculated using the bathtub model. Bathtub fits and extrapolations are shown in *Figure 38* and reflect clinical practice as represented by patients in the NJR database. Bathtub fits were supported visually (see *Appendix 17*) and by plots of modelled compared with Kaplan–Meier-estimated cumulative hazards (*Figure 39*).

The bathtub-modelled percentage revision at 10, 20 and 30 years is summarised in *Table 63*.

As the age distributions of the matched populations were somewhat removed from normal (see *Figure 36*) we undertook sensitivity analysis in which bathtub models were controlled for age and sex and extrapolated revision was calculated for an 'average' population aged 55.8 years with 35% women (see *Figure 38*). Because it was evident that revision rates were much higher for women receiving RS than for men receiving RS, and because revision rates likely vary according to the age of patients, subgroup analyses focused on comparing populations stratified by sex and controlled for age. The results of the analysis of revision rates for these subgroups are provided in the following sections and in *Appendix 17*.

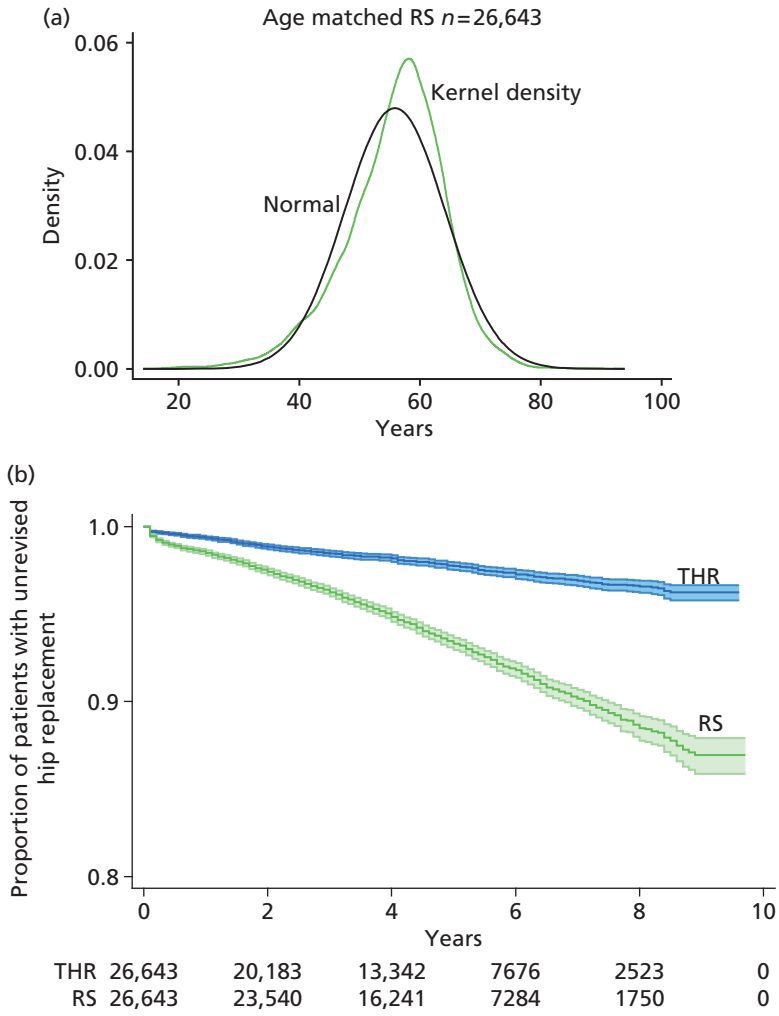


FIGURE 36 (a) Age distribution; and (b) time to revision for RS and THR matched populations. Numbers under *x*-axis are numbers at risk.

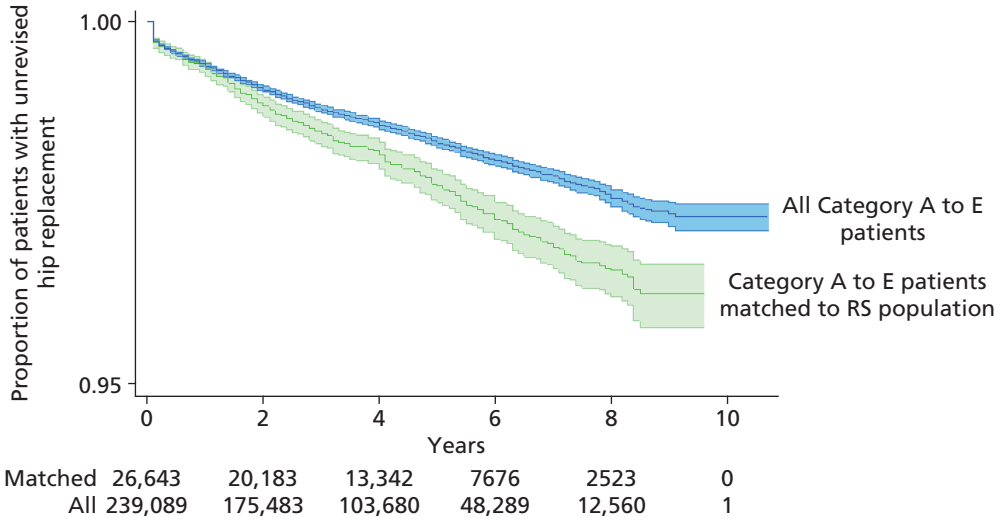


FIGURE 37 Revision rates for matched and whole THR populations. Numbers under *x*-axis are numbers at risk.

TABLE 62 Information criteria scores for models of revision rates (RS and matched THR)

Intervention	Model	Observations	Model likelihood	Parameters	AIC	BIC
THR	Exponential	26,643	–3239.377	1	6480.753	6488.944
THR	Weibull	26,643	–3219.967	2	6443.935	6460.315
THR	Gompertz	26,643	–3230.912	2	6465.825	6482.205
THR	Log-normal	26,643	–3221.913	2	6447.827	6464.207
THR	Log-logistic	26,643	–3220.111	2	6444.222	6460.603
THR	Bathtub	26,643	–3215.51	3	6437.021	6461.592
RS	Exponential	26,643	–8102.451	1	16206.9	16215.09
RS	Weibull	26,643	–8101.688	2	16207.38	16223.76
RS	Gompertz	26,643	–8094.569	2	16193.14	16209.52
RS	Log-normal	26,643	–8162.981	2	16329.96	16346.34
RS	Log-logistic	26,643	–8107.527	2	16219.05	16235.43
RS	Bathtub	26,643	–8037.685	3	16081.37	16105.94

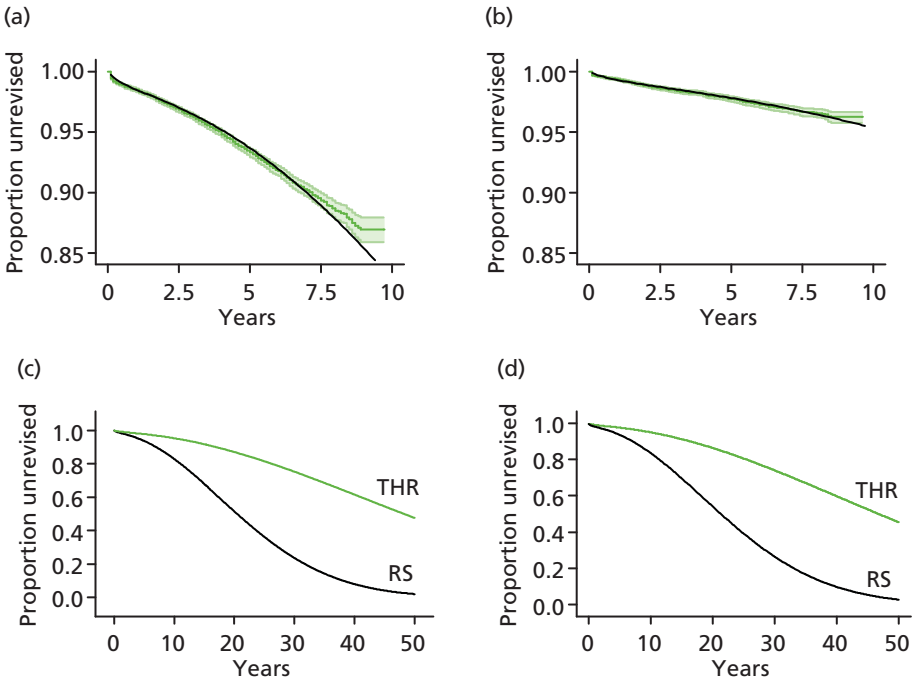


FIGURE 38 Bathtub fits and extrapolations for matched THR and RS populations. (a) Bathtub fit, RS; (b) bathtub fit, THR; (c) uncontrolled bathtub extrapolation; and (d) controlled bathtub extrapolation.

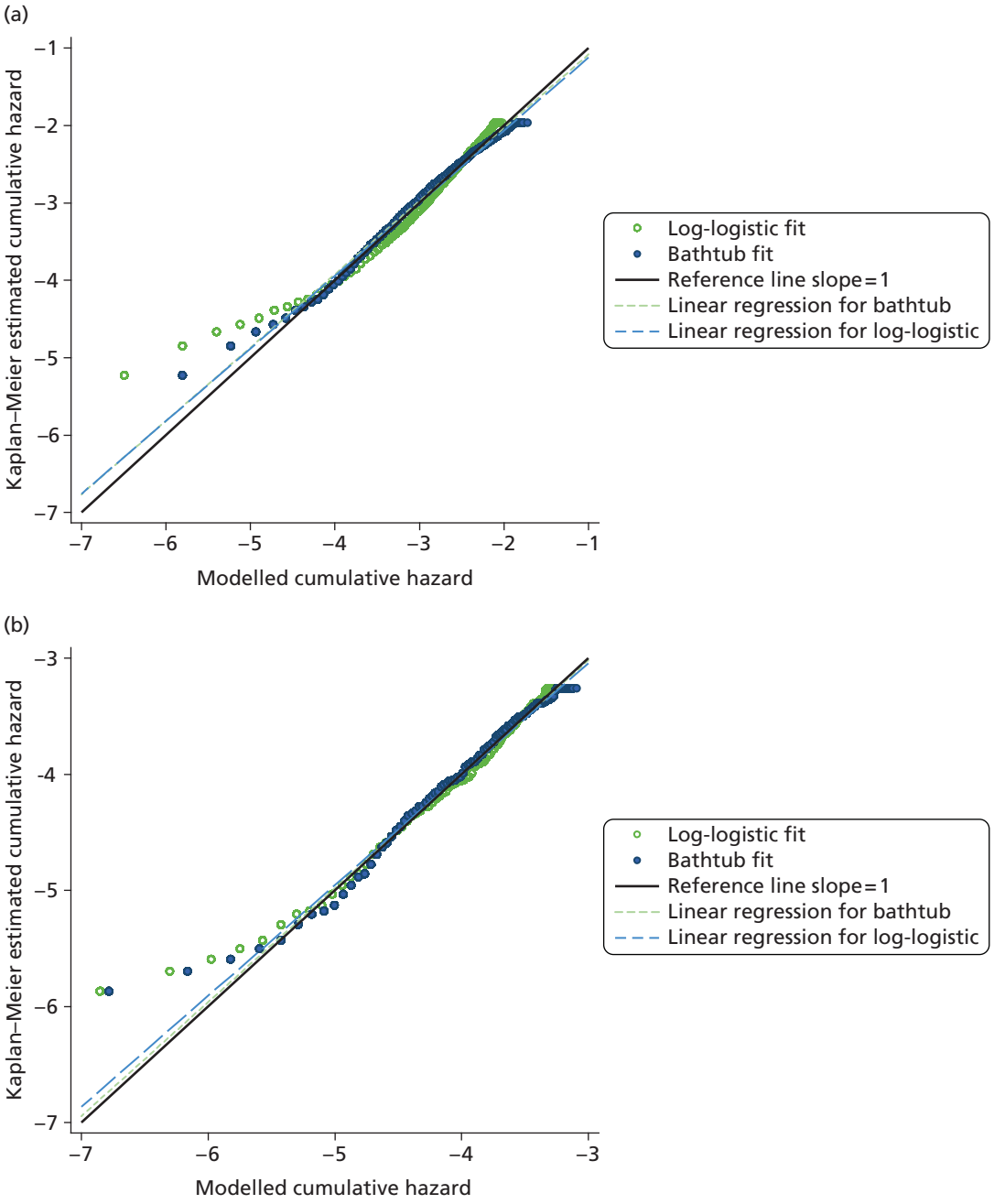


FIGURE 39 Kaplan–Meier vs. modelled cumulative hazard. (a) RS; and (b) THR (matched).

TABLE 63 Bathtub-modelled percentage of patients requiring revision

	Revision (%)		
Intervention	10 years	20 years	30 years
RS	17.2	48.3	76.3
THR	4.6	12.9	24.6

Comparison of the total hip replacement categories

For THR patients encompassed within the five selected categories (A–E, $n = 239{,}089$), the proportion remaining unrevised at 9 years according to Kaplan–Meier analysis was 0.974. The proportion of all 386,556 THR recipients unrevised at 9 years was 0.962 (*Figure 40*). The Kaplan–Meier plot for the five selected THR interventions indicated a relatively high initial hazard for revision that gradually decreased over about 4 years and subsequently gradually increased between 5 and 9 years.

Kaplan–Meier analyses indicated different revision rates across the five categories of THR (*Figures 41* and *42*). Revision rates for patients who received CeLCoC (category C) and CeLMoP (category B) THRs

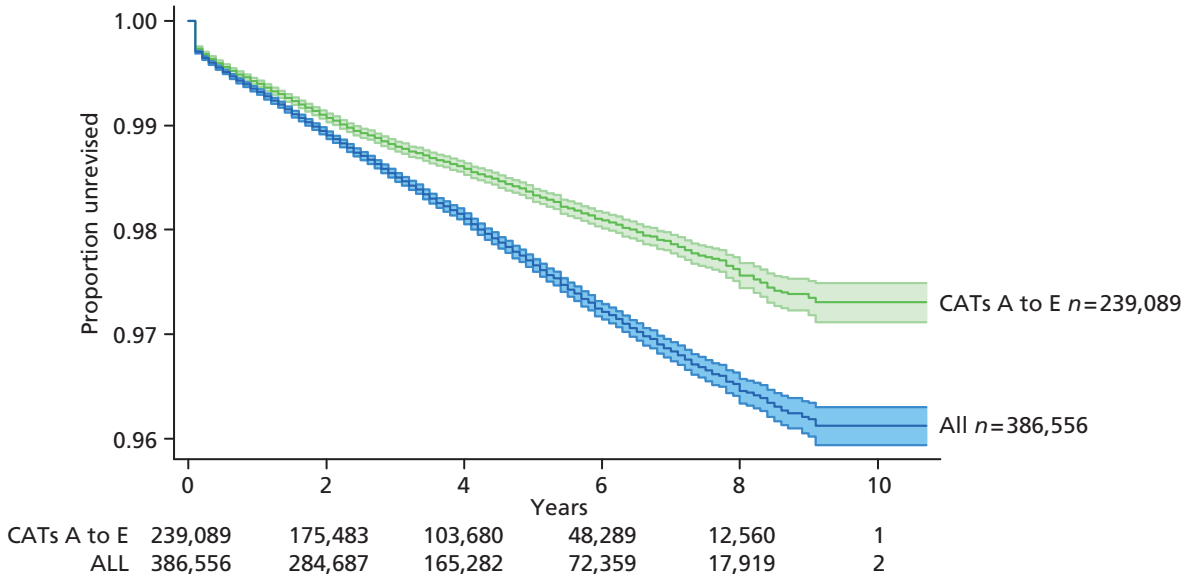


FIGURE 40 Revision estimated for all THR patients and those receiving category A–E THRs. Numbers under *x*-axis are numbers at risk.

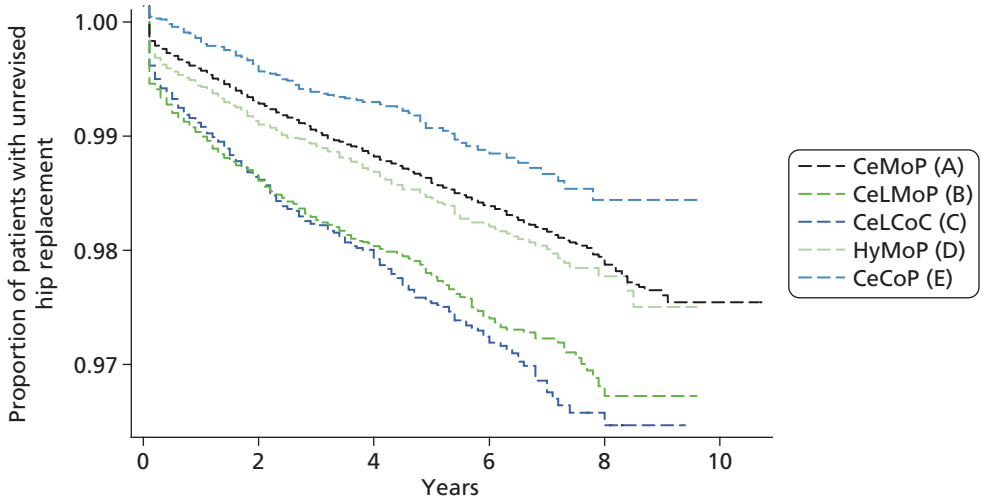


FIGURE 41 Kaplan–Meier analyses across the five categories of THR.

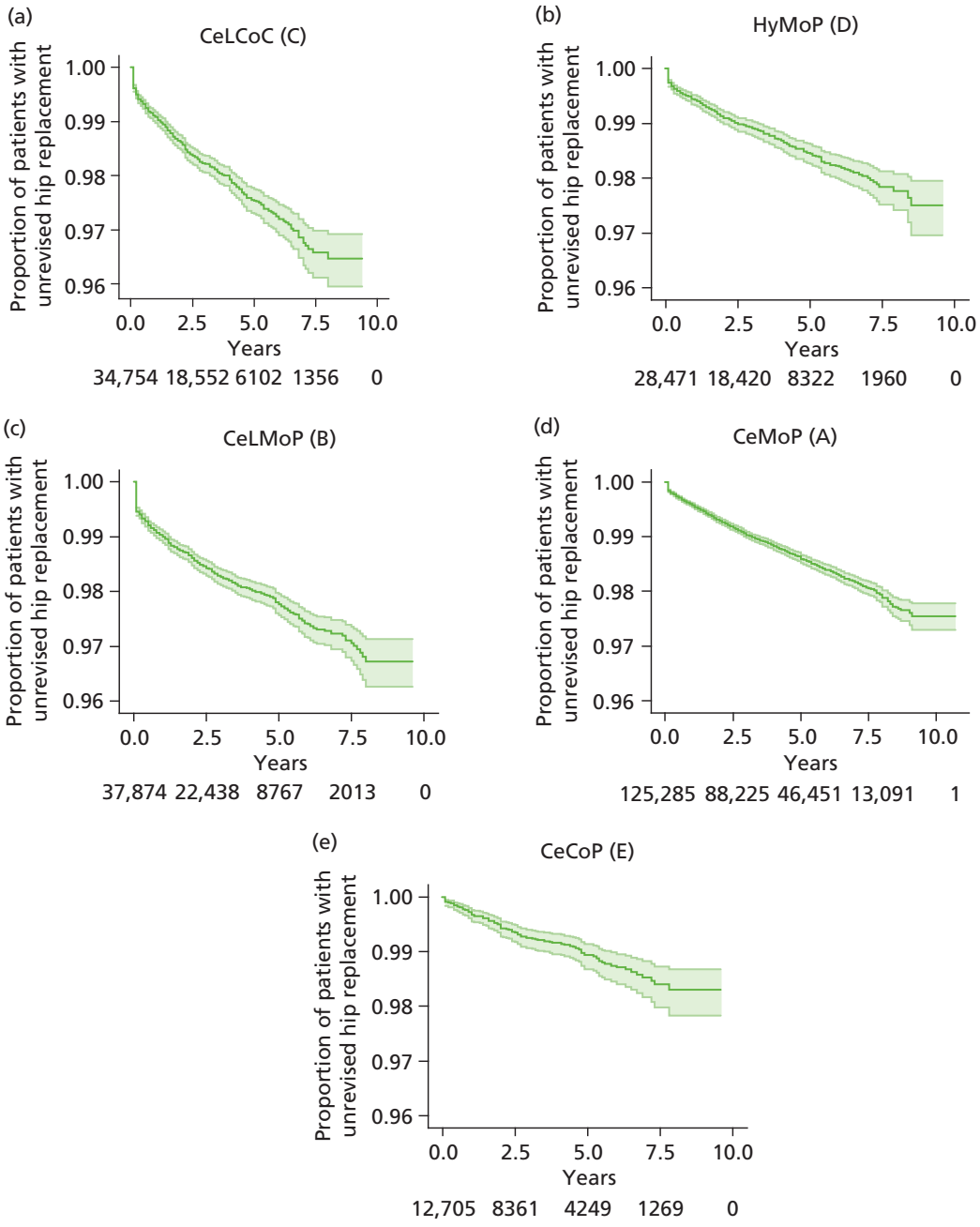


FIGURE 42 Observed time to revision with 95% CIs for compared THR categories. Numbers under *x*-axis are numbers at risk.

were clearly higher than those for patients who received CeCoP (category E) and CeMoP (category A) THRs.

According to information criteria scores (*Table 64*), other than for CeCoP (category E) THR, the bathtub model provided the best parametric fit, followed by the log-normal model. For CeCoP (category E), the log-normal model was marginally superior to the bathtub model. These inferences were supported by visual inspection (see *Appendix 17*) and by comparing modelled with Kaplan–Meier-estimated cumulative hazards for each category (*Figure 43*).

TABLE 64 Information criteria scores for models of revision rates: THR categories

THR	Model	Observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	34,754	–3955.734	1	7913.467	7921.923
CeLCoC (category C)	Weibull	34,754	–3882.115	2	7768.229	7785.141
CeLCoC (category C)	Gompertz	34,754	–3906.282	2	7816.563	7833.475
CeLCoC (category C)	Log-normal	34,754	–3872.162	2	7748.323	7765.235
CeLCoC (category C)	Log-logistic	34,754	–3881.911	2	7767.822	7784.734
CeLCoC (category C)	Bathtub	34,754	–3858.878	3	7723.755	7749.123
HyMoP (category D)	Exponential	28,471	–2428.234	1	4858.468	4866.724
HyMoP (category D)	Weibull	28,471	–2387.427	2	4778.854	4795.368
HyMoP (category D)	Gompertz	28,471	–2405.936	2	4815.872	4832.385
HyMoP (category D)	Log-normal	28,471	–2383.97	2	4771.94	4788.454
HyMoP (category D)	Log-logistic	28,471	–2387.411	2	4778.822	4795.335
HyMoP (category D)	Bathtub	28,471	–2373.646	3	4753.291	4778.061
CeLMoP (category B)	Exponential	37,874	–4535.478	1	9072.955	9081.497
CeLMoP (category B)	Weibull	37,874	–4391.882	2	8787.763	8804.847
CeLMoP (category B)	Gompertz	37,874	–4442.601	2	8889.202	8906.286
CeLMoP (category B)	Log-normal	37,874	–4377.507	2	8759.014	8776.098
CeLMoP (category B)	Log-logistic	37,874	–4391.567	2	8787.133	8804.217
CeLMoP (category B)	Bathtub	37,874	–4345.8	3	8697.601	8723.227
CeMoP (category A)	Exponential	125,285	–10000.51	1	20003.01	20012.75
CeMoP (category A)	Weibull	125,285	–9929.73	2	19863.46	19882.94
CeMoP (category A)	Gompertz	125,285	–9965.745	2	19935.49	19954.97
CeMoP (category A)	Log-normal	125,285	–9927.767	2	19859.53	19879.01
CeMoP (category A)	Log-logistic	125,285	–9929.867	2	19863.73	19883.21
CeMoP (category A)	Bathtub	125,285	–9909.508	3	19825.02	19854.23
CeCoP (category E)	Exponential	12,705	–759.4492	1	1520.898	1528.348
CeCoP (category E)	Weibull	12,705	–757.1662	2	1518.332	1533.232
CeCoP (category E)	Gompertz	12,705	–757.8727	2	1519.745	1534.645
CeCoP (category E)	Log-normal	12,705	–756.8497	2	1517.699	1532.599
CeCoP (category E)	Log-logistic	12,705	–757.163	2	1518.326	1533.226
CeCoP (category E)	Bathtub	12,705	–756.6023	3	1519.205	1541.554

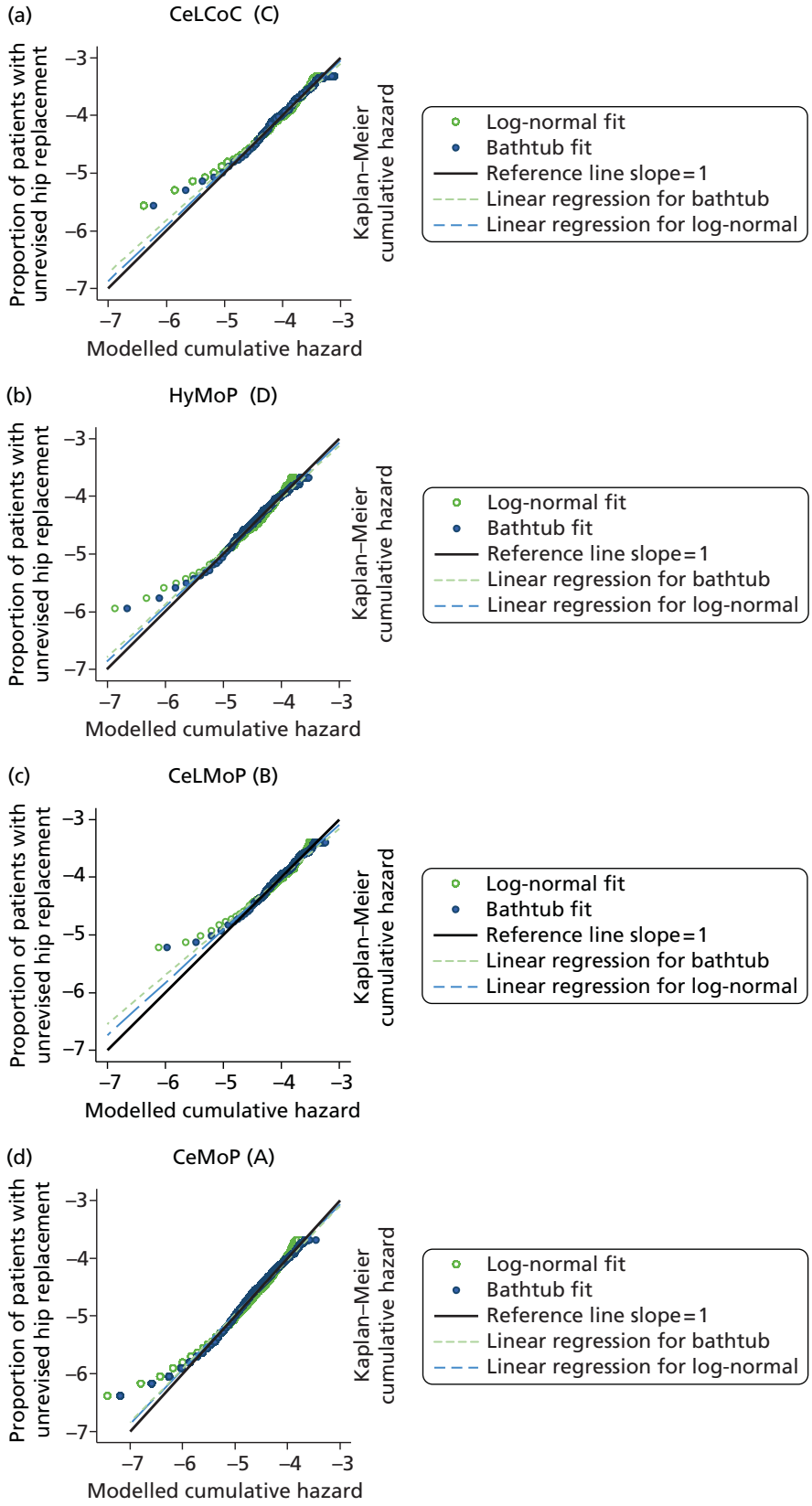


FIGURE 43 Modelled vs. Kaplan–Meier-estimated cumulative hazards. (a) CeLCoC (category C); (b) HyMoP (category D); (c) CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E) (categories B–D, cementless; categories A and E, cemented). (*continued*)

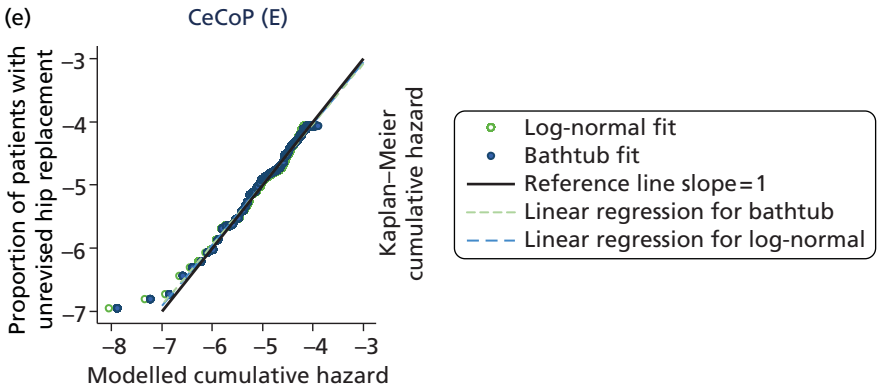


FIGURE 43 Modelled vs. Kaplan–Meier-estimated cumulative hazards. (a) CeLCoC (category C); (b) HyMoP (category D); (c) CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E) (categories B–D, cementless; categories A and E, cemented).

For the base-case economic analysis, transition probabilities were calculated from the bathtub fit for all categories. The fit to the Kaplan–Meier estimates and the extrapolation beyond the observed data are shown in *Figures 44* and *45*, respectively. These analyses reflect the performance of the five types of prosthesis for NJR patients over 9–10 years to 2012.

The lowest and highest revision rates were experienced by CeCoP (category E) and CeLCoC (category C) recipients, respectively (*Table 65*). The bathtub-modelled percentage of patients requiring revision at 10, 20 and 30 years is summarised in *Table 66*.

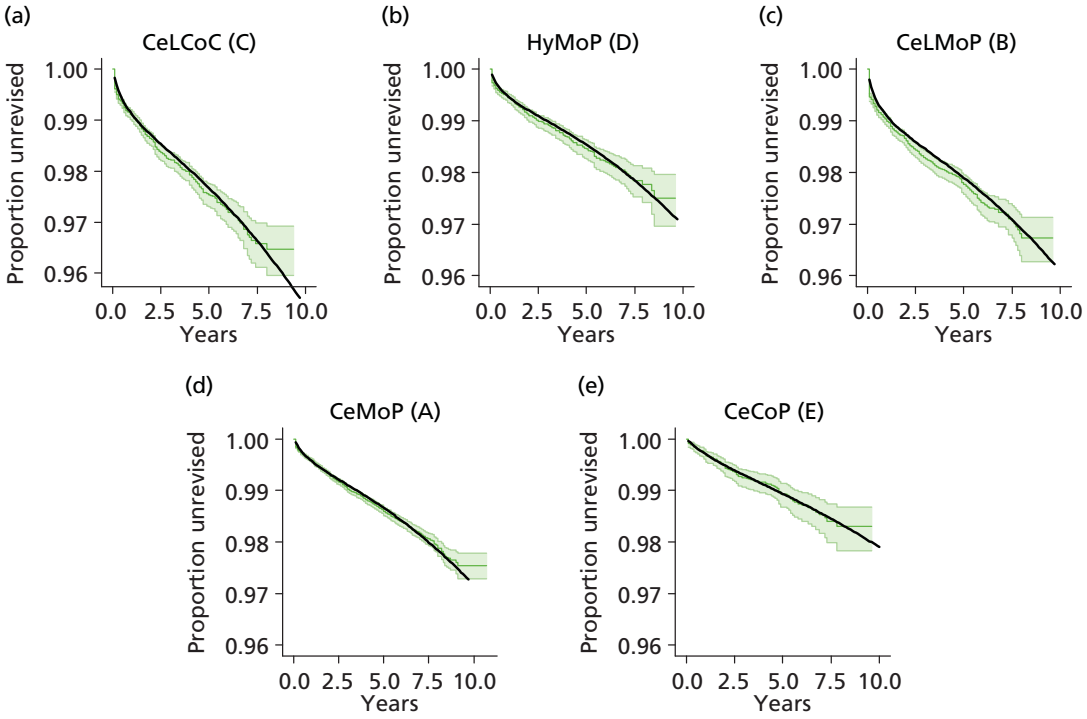


FIGURE 44 Bathtub parametric fits to observed time to revision for THR categories A–E. (a) CeLCoC (category C); (b) HyMoP (category D); (c) CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E).

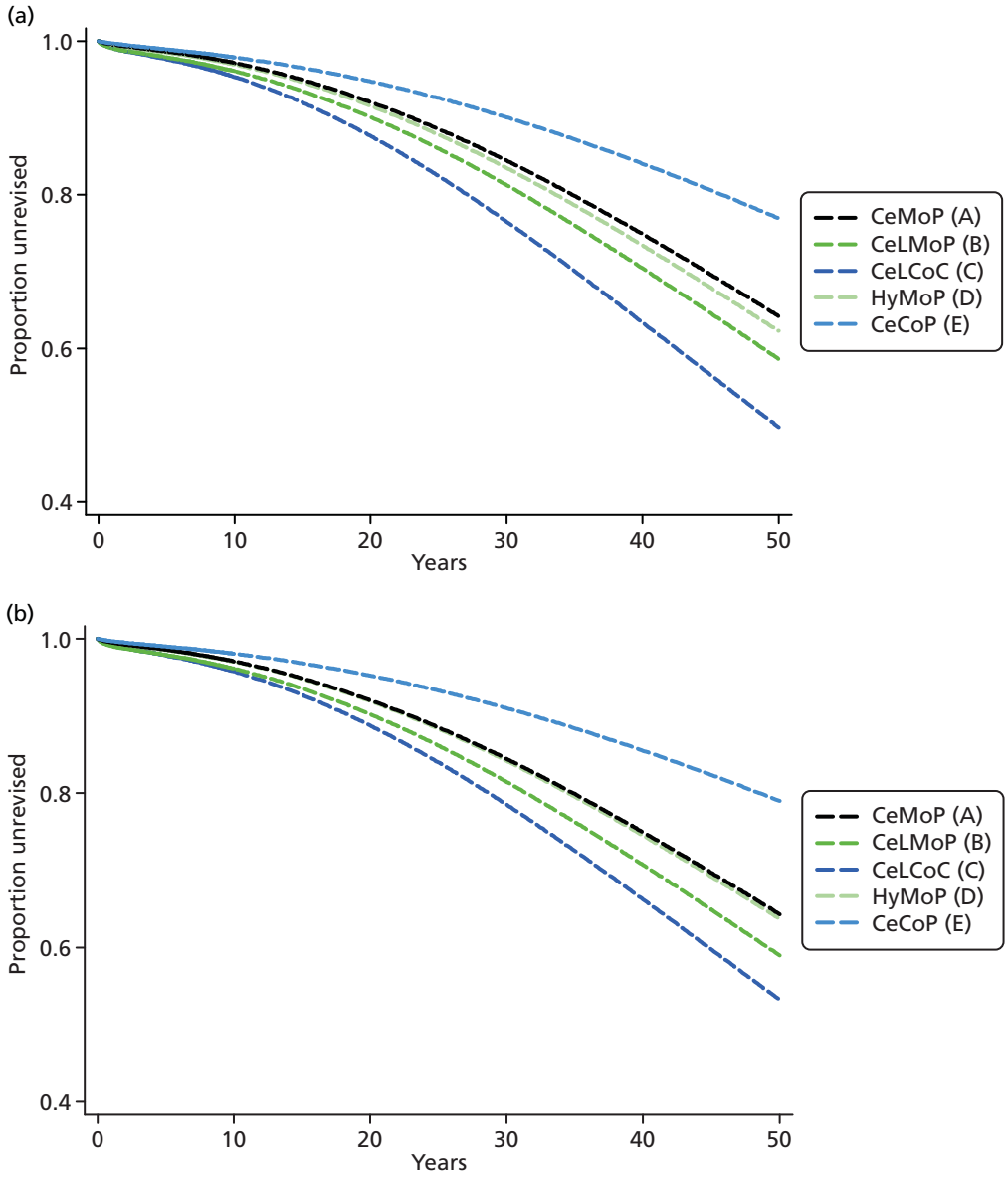


FIGURE 45 Extrapolation of bathtub models of revision for THR categories A–E. (a) Uncontrolled; and (b) controlled for age and sex for a modelled population aged 71.6 years with 63.5% women.

TABLE 65 Age and sex of recipients of THR categories A–E

Population	Number	% female	Mean (SD) age (years)	Median age (years)	Interquartile range (years)
All THR (category A–E) recipients	239,089	63.5	71.6 (9.6)	72.5	65.8–78.3
All CeMoP (category A) recipients	125,285	66.9	74.6 (7.9)	74.9	69.7–80
All CeLMoP (category B) recipients	37,874	60.2	71.5 (8.7)	72	65.9–77.5
All CeLCoC (category C) recipients	34,754	55.4	61.6 (9.9)	62.3	55.9–67.9
All HyMoP (category D) recipients	28,471	64.2	73.0 (8.3)	73.4	67.8–78.7
All CeCoP (category E) recipients	12,705	60.1	66.2 (9.6)	66.3	60.7–72.5

TABLE 66 Bathtub-modelled percentage of patients requiring revision

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	2.8	7.9	15.6
CeLMoP (category B)	3.9	9.9	18.7
CeLCoC (category C)	4.6	12.3	23.5
HyMoP (category D)	3.0	8.4	16.5
CeCoP (category E)	2.1	5.2	9.9

Percentages refer to the mean age of patients in each category.

Across the five THR category recipients 36.5% were men and 63.5% were women but within categories the ratio varied from 1.24 for CeLCoC (category C) to 2.02 for CeMoP (category B). Revision was more frequent for men than for women (*Figure 46*) although this was least pronounced for the CeCoP (category E) prosthesis.

Similarly, the age distribution of patients differed somewhat according to THR category (*Figure 47*). CeLCoC (category C) prostheses were used more for younger patients and CeMoP (category A) prostheses were used more for older patients. Across the five THR categories the mean age was 71.56 years

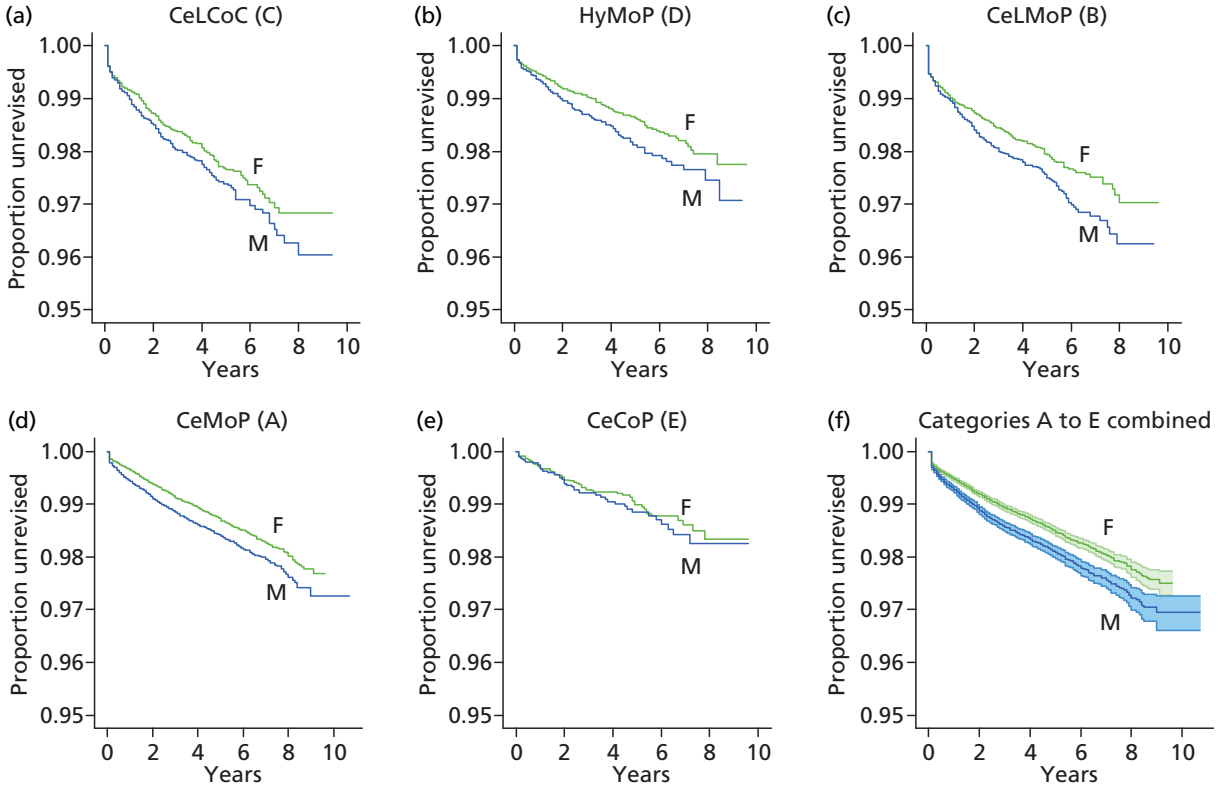


FIGURE 46 Total hip replacement revision rates observed for men and women. (a) CeLCoC (category C); (b) HyMoP (category D); (c) CeLMoP (category B); (d) CeMoP (category A); (e) CeCoP (category E); and (f) all THR categories combined. F, female; M, male.

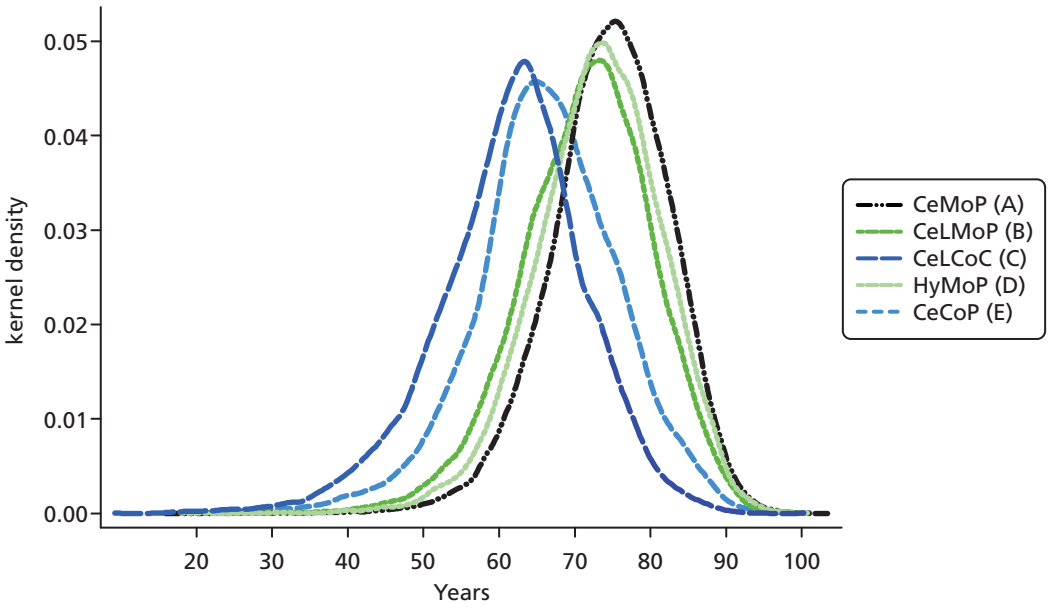


FIGURE 47 Kernel density plots of age at primary THR for category A–E THR prostheses.

In sensitivity analysis the bathtub model was controlled for age and sex to adjust for spurious differences in revision rates because of differing proportions of men and women or of younger or older patients in the different THR categories. The relative performance of the five categories modelled for a population aged 71.6 years with 63.5% women demonstrates that the superiority of the CeCoP prosthesis was somewhat enhanced.

In further sensitivity analysis we used log-normal fits to the Kaplan–Meier-estimated revision rates; these are shown for each of the types of THR (categories A–E) in *Table 67* and *Figure 48*. With a mean age across all categories of nearly 72 years, extrapolation predicting a decreasing hazard for revision may be

TABLE 67 Log-normal-modelled percentage of patients requiring revision

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	2.3	3.5	4.4
CeLMoP (category B)	3.3	4.6	5.5
CeLCoC (category C)	3.7	5.3	6.4
HyMoP (category D)	2.4	3.4	4.2
CeCoP (category E)	1.8	2.9	3.8

Percentages refer to the mean age of patients in each category.

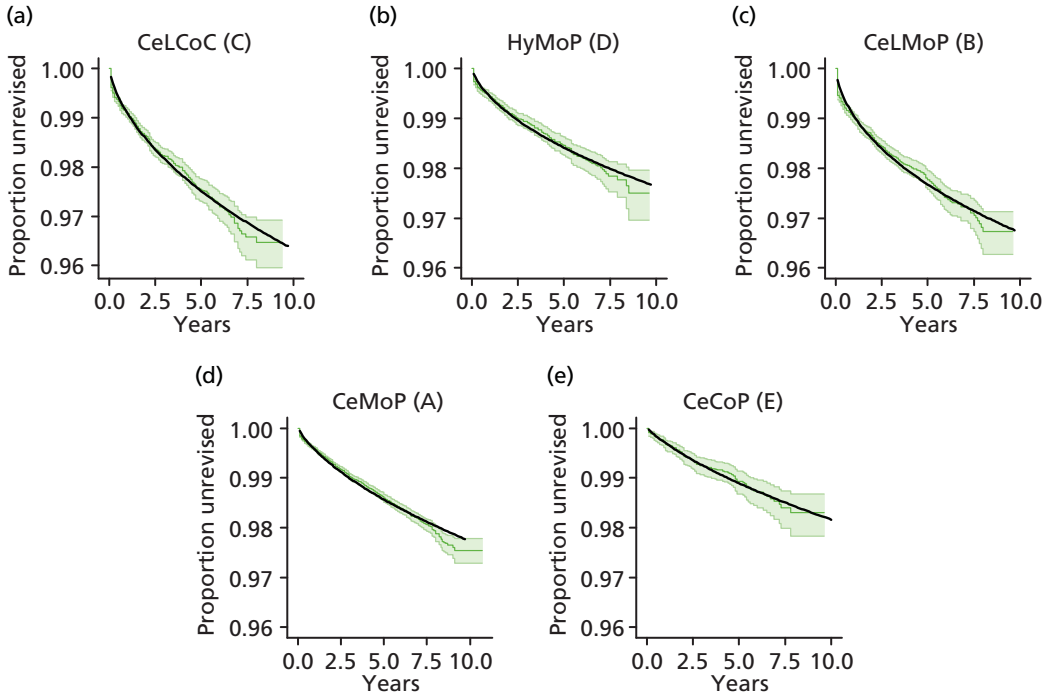


FIGURE 48 Log-normal parametric fits to observed revision rates for THR categories A–E. (a) CeLCoC (category C); (b) HyMoP (category D); (c) CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E).

appropriate. The best-fit model providing this condition was the log-normal model. These fits are shown in *Figure 49*. The relative performance of the prostheses was similar to that with the bathtub model; however, unsurprisingly, extrapolated revision rates were lower than with the bathtub model.

Further sensitivity analysis was carried out in which the log-normal model was controlled for age and sex. With this model the superior performance of the CeCoP (category E) prosthesis was maintained (see *Figure 49*).

Comparison between resurfacing arthroplasty and total hip replacement: subgroup analyses according to sex (women)

Because the use of different categories of THR prostheses differed by age and sex and as recipients of THR interventions aged > 65 years approximate a population unlikely to be considered candidates for RS (see *Figure 24*), we undertook subgroup analyses in which the THR population for each category was stratified by sex and by age (> 65 years and < 65 years) and parametric models were controlled for age. Results from these analyses are presented in *Figure 49*.

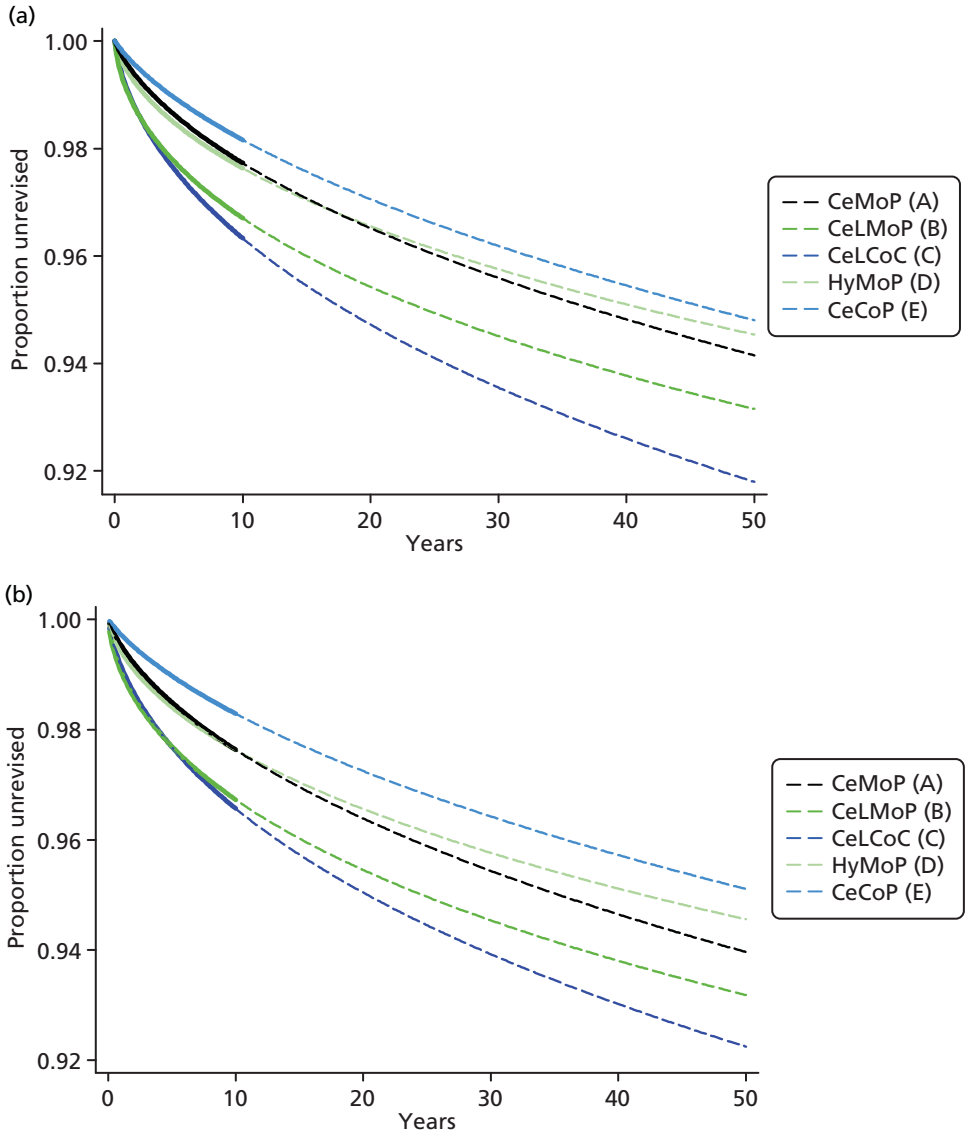


FIGURE 49 Log-normal-modelled revision. (a) Uncontrolled; and (b) controlled for a population of mean age 71.6 years with 36.5% men.

As expected, the matched groups (n =9321) had an identical age distribution [mean 53.5 (SD 8.4) years, range 15–93 years] (*Figure 50*).

The observed time to revision was far shorter for RS than for THR recipients (*Figure 51*).

For RS, Gompertz, bathtub and Weibull models provided good fits and each predicted an increasing hazard beyond the observed data; according to AIC scores and cumulative hazard plots the Gompertz and bathtub models were the better fits (see *Appendices 19* and *20*, respectively) and predicted similar revision beyond the observed data.

For THR patients the bathtub fit was as good as the alternatives (see *Appendix 17*) and was the only model that predicted an increasing hazard beyond the observation period. According to AIC scores and cumulative hazard plots, differences were trivial between the bathtub, log-normal and Weibull models (see *Appendices 19* and *20*, respectively). For the economic analysis the bathtub model was adopted for both the RS group and the THR group. The predicted requirement for revision at 10, 20 and 30 years using the bathtub model is shown in *Table 68*.

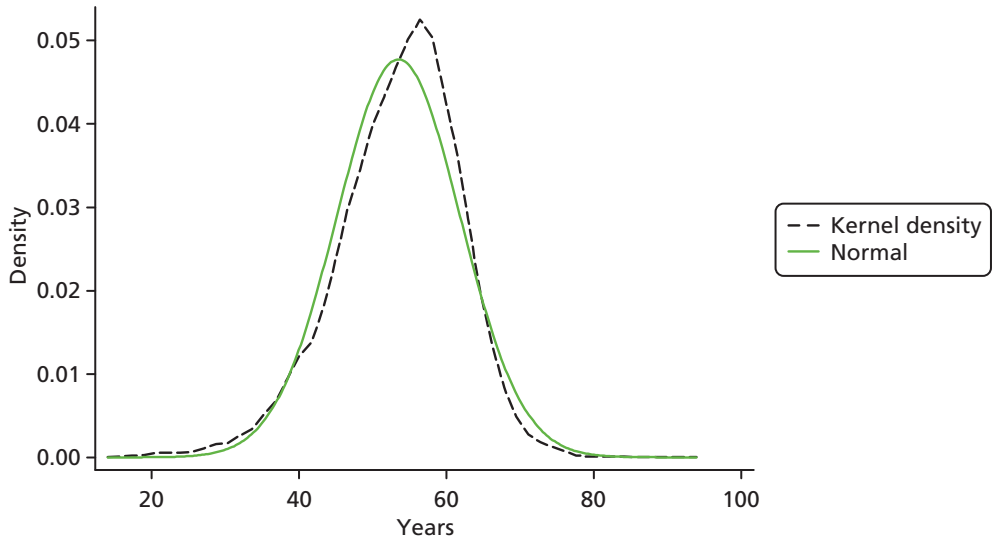


FIGURE 50 Kernel density plot for age distribution in matched RS and THR female groups.

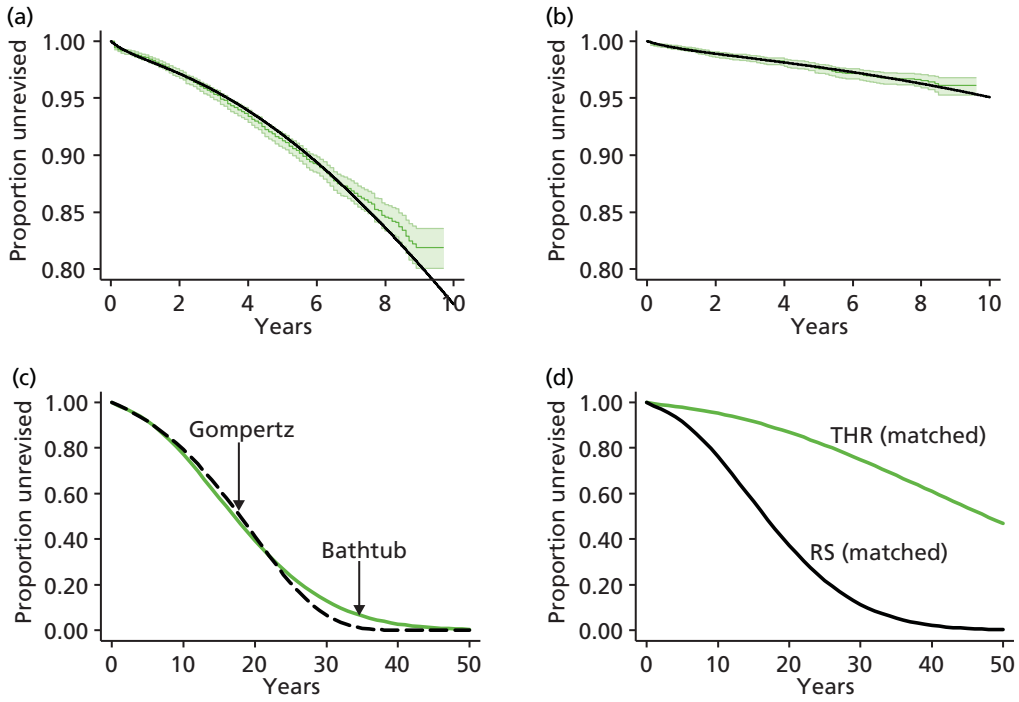


FIGURE 51 Observed revision (95% CI) and bathtub models for RS and THR female groups. Also shown is the Gompertz model for RS. (a) Bathtub model, RS; (b) bathtub model, THR; (c) model extrapolations, RS; and (d) bathtub extrapolations.

TABLE 68 Bathtub-modelled percentage of patients requiring revision (women aged 53.5 years)

	Revision (%)		
Intervention	10 years	20 years	30 years
RS	23.1	61.2	87.6
THR	4.8	13.2	25.2

Percentages refer to the mean age of patients in each category.

Comparison between resurfacing arthroplasty and total hip replacement: subgroup analyses according to sex (men)

Each of the matched groups ($n = 17{,}322$) had a mean age of 57.1 (SD 8.03 years; range 16–89 years) and an identical age distribution (*Figure 52*).

The observed revision rate was higher for RS than for THR (*Figure 53*). Parametric fits are presented in *Appendix 17*. The bathtub distribution produced the lowest AIC scores and visually the superior fit (see *Appendices 17* and *19*); cumulative hazard plots are provided in *Appendix 20*. Apart from the bathtub model, the models predicted a decreasing hazard on extrapolation (see *Appendix 17*). For the economic analysis the bathtub model was adopted for both the RS group and the THR group. The predicted requirement for revision at 10, 20 and 30 years is shown in *Table 69*.

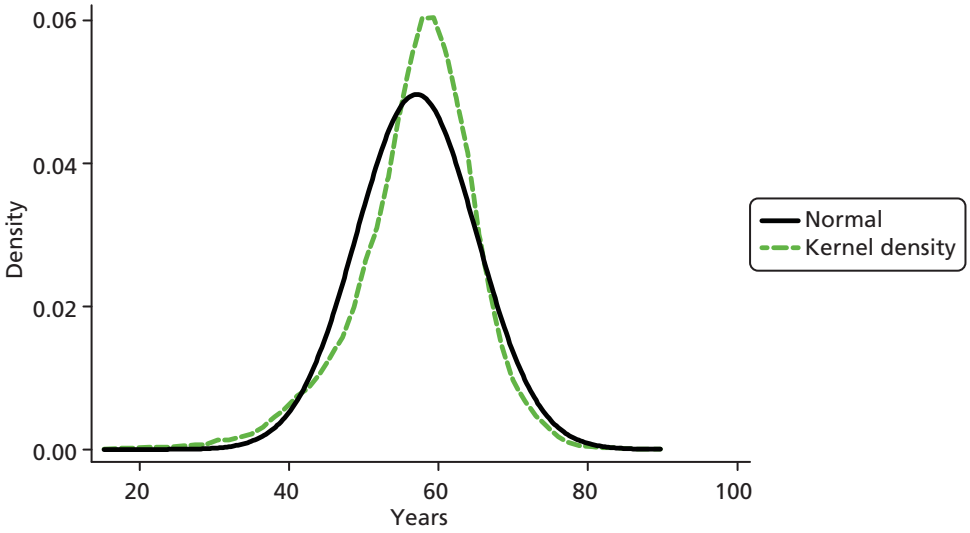


FIGURE 52 Kernel density plot for age distribution in matched RS and THR male groups.

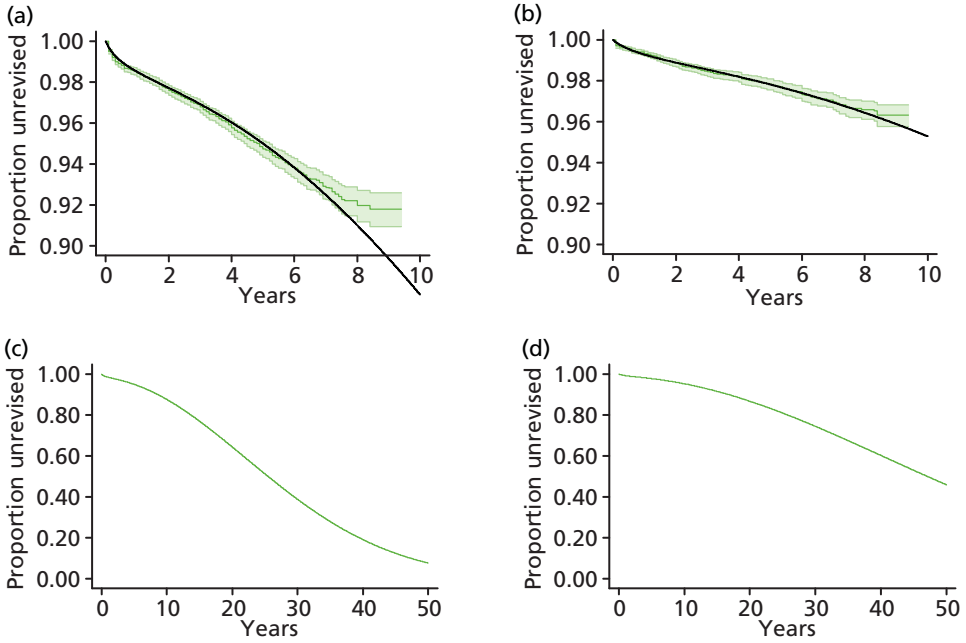


FIGURE 53 Observed revision (95% CI) and bathtub models for RS and THR male groups. (a) Bathtub model, RS; (b) bathtub model, THR; (c) model extrapolation, RS; and (d) model extrapolation, THR.

TABLE 69 Bathtub-modelled percentage of patients requiring revision (men aged 57.1 years)

	Revision (%)		
Intervention	10 years	20 years	30 years
RS	12.4	35.6	61.2
THR	4.7	13.2	25.5

Percentages refer to the mean age of patients in each category.

Comparison of total hip replacement revision rates according to sex and age: men aged more than 65 years

Figure 54 shows the observed time to revision for male patients aged > 65 years according to category of THR prosthesis. Revision was less frequent for CeCoP (category E) than for other categories. Parametric fits to the observed data are shown in *Appendix 17*, AIC values for models in *Appendix 19* and diagnostic plots in *Appendix 20*. Visually and by AIC scores the bathtub and log-normal models generated best fits except for the CeCoP (category E) prosthesis for which the bathtub model did not resolve. In view of the advanced age of these patients, after accumulating 9 years of follow-up data, it was considered that an increasing hazard (bathtub) for revision was unlikely and therefore the log-normal model was used for the economic base case. The extrapolations shown in *Figure 54* apply for patients aged 70 years.

The model-predicted requirements for revision at 10, 20 and 30 years are summarised in *Table 70*.

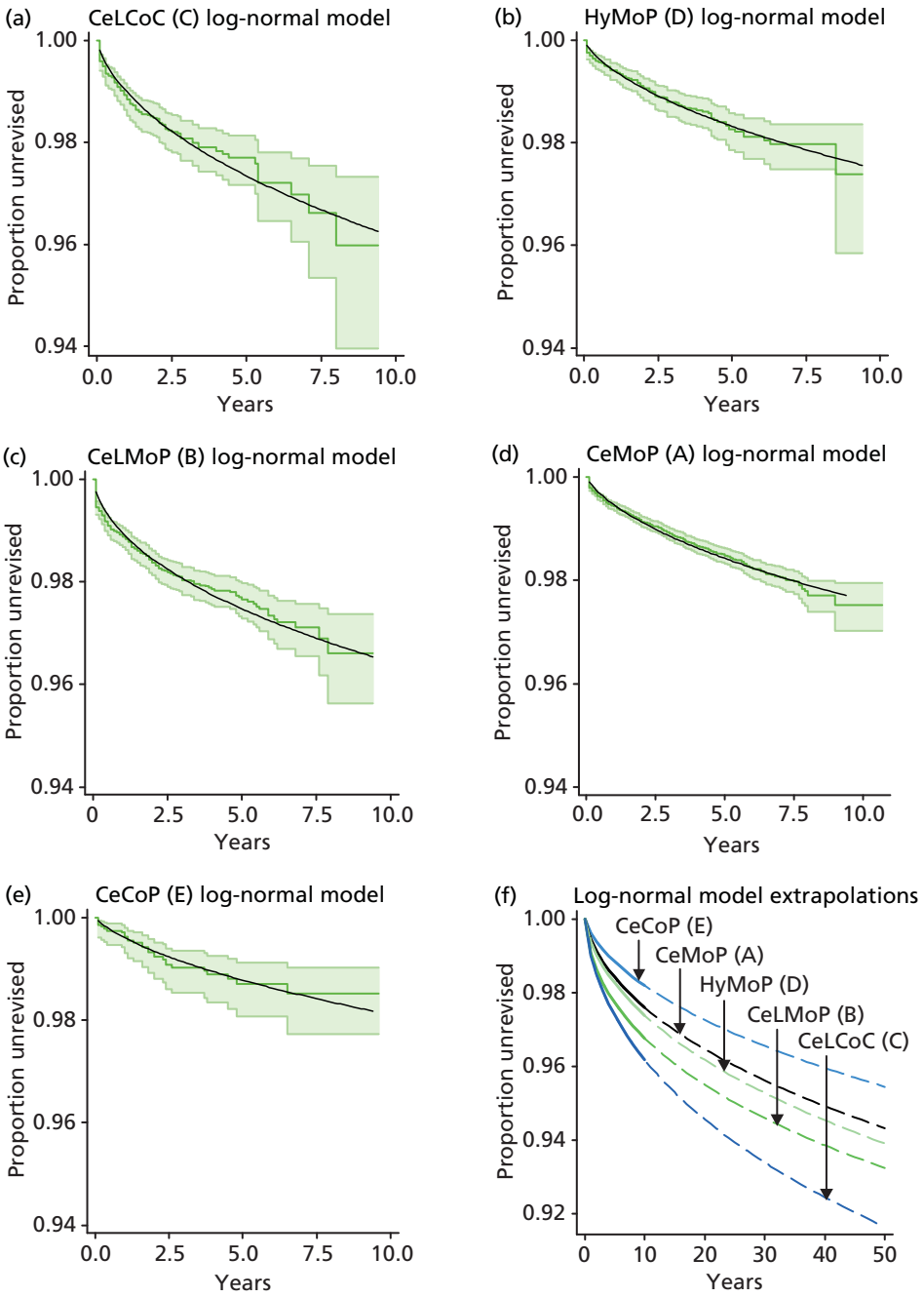


FIGURE 54 Observed revision (95% CI) for men aged > 65 years and log-normal models for THR categories. (a) CeLCoC (category C) log-normal model; (b) HyMoP (category D) log-normal model; (c) CeLMoP (category B) log-normal model; (d) CeMoP (category A) log-normal model; (e) CeCoP (category E) log-normal model; and (f) log-normal model extrapolations.

TABLE 70 Log-normal-modelled percentage of patients requiring revision (men aged > 65 years)

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	2.4	3.5	4.4
CeLMoP (category B)	3.6	4.9	5.9
CeLCoC (category C)	3.9	5.5	6.7
HyMoP (category D)	2.5	3.7	4.6
CeCoP (category E)	1.9	2.9	3.6

Percentages refer to the mean age of patients in each category.

Comparison of total hip replacement revision rates according to sex and age: women aged more than 65 years

Figure 55 shows the observed time to revision for female patients aged > 65 years according to category of THR prosthesis. Revision was less frequent for CeCoP (category E) than for other categories. Parametric fits to the observed data are shown in *Appendix 17*, AIC values for models in *Appendix 19* and diagnostic plots in *Appendix 20*. Visually and by AIC scores the bathtub and log-normal models generated best fits except for the CeCoP (category E) prosthesis for which the bathtub model did not resolve. In view of the advanced age of these patients, after accumulating 9 years of follow-up data, it was considered that an increasing hazard (bathtub) for revision is unlikely and therefore the log-normal model was used for the economic base case. The extrapolations shown in *Figure 55* apply for patients aged 70 years. The predicted requirement for revision at 10, 20 and 30 years is summarised in *Table 71*.

Comparison of total hip replacement revision rates according to sex and age: men aged less than 65 years

Figure 56 shows the observed time to revision for male patients aged < 65 years according to category of THR prosthesis. Parametric fits to the observed data are shown in *Appendix 17* and AIC values for models are summarised in *Appendix 19*. Cumulative hazard plots are shown in *Appendix 20*. Observed revision was less frequent for CeCoP (category E) than for other categories. According to AIC values (and visually), the bathtub model provided a superior fit for categories B, C and D followed by the log-normal model. For categories A and E there were only trivial differences in AIC values between the bathtub and the log-normal models. On extrapolation of the bathtub models the CeMoP category becomes superior to CeCoP after about 25 years' follow-up. Transition probabilities for the economic analysis were based on bathtub models (base case for the subgroup) and log-normal models were used in sensitivity analysis. The extrapolations of the bathtub models shown in *Figure 56* apply to patients aged 50 years.

The bathtub-predicted requirement for revision at 10, 20 and 30 years is summarised in *Table 72*.

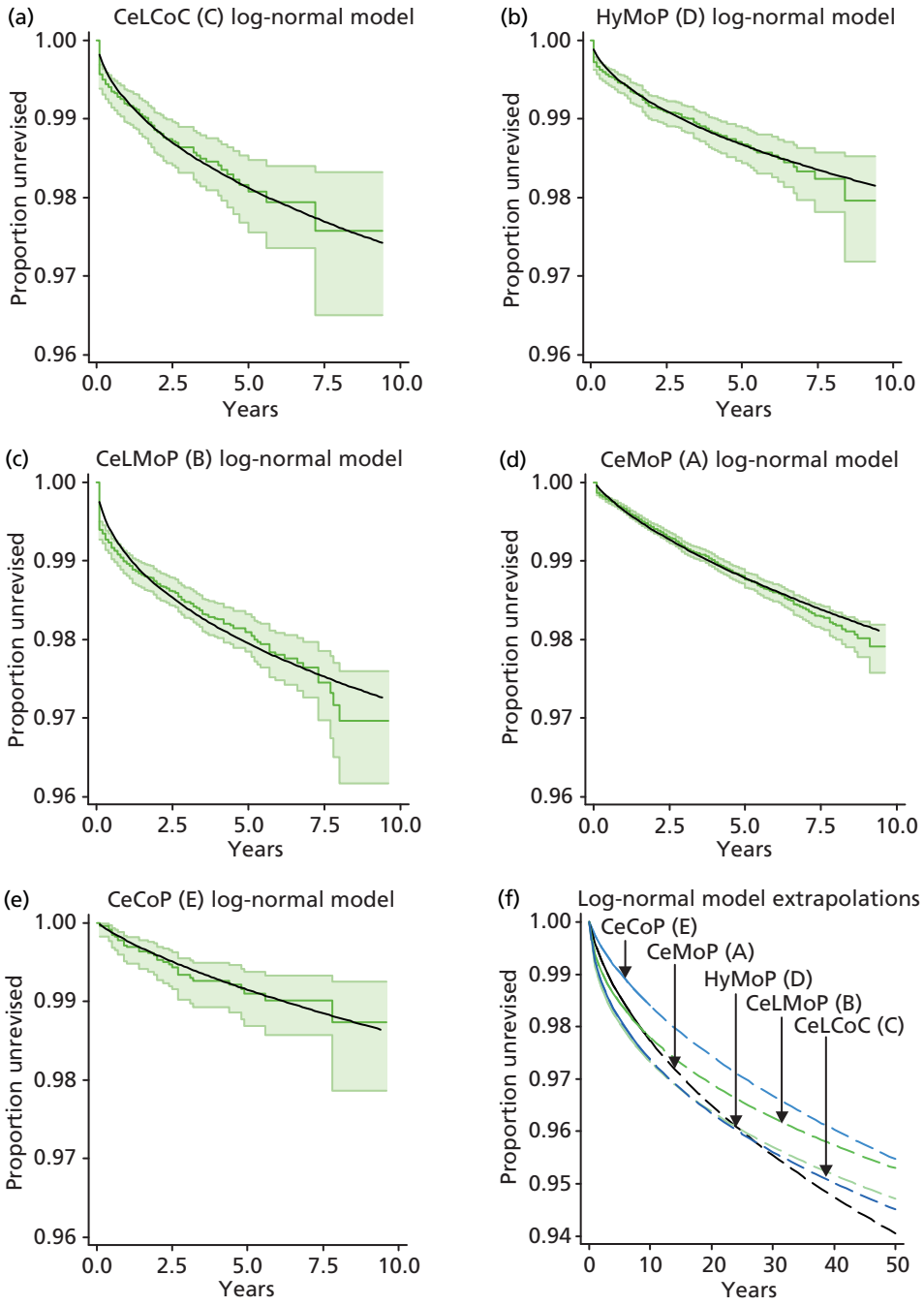


FIGURE 55 Observed revision (95% CI) for women aged > 65 years and log-normal models for THR categories. (a) CeLCoC (category C) log-normal model; (b) HyMoP (category D) log-normal model; (c) CeLMoP (category B) log-normal model; (d) CeMoP (category A) log-normal model; (e) CeCoP (category E) log-normal model; and (f) log-normal model extrapolations.

TABLE 71 Log-normal-modelled percentage of patients requiring revision (women aged > 65 years)

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	2.0	3.1	3.9
CeLMoP (category B)	2.8	3.8	4.5
CeLCoC (category C)	2.7	3.7	4.4
HyMoP (category D)	1.9	2.7	3.3
CeCoP (category E)	1.4	2.3	3.0

Percentages refer to the mean age of patients in each category.

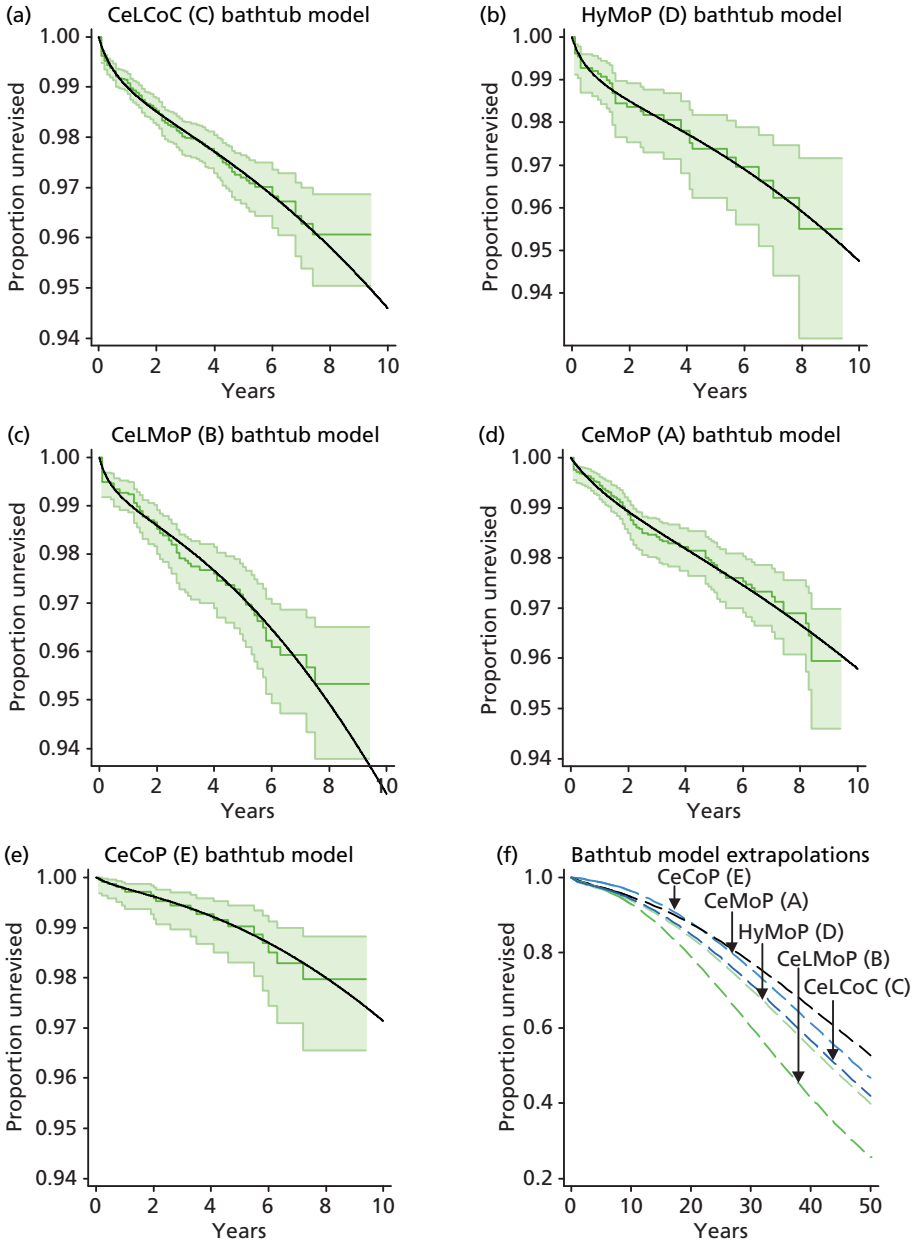


FIGURE 56 Observed revision (95% CI) for men aged < 65 years and bathtub models for THR categories. (a) CeLCoC (category C) bathtub model; (b) HyMoP (category D) bathtub model; (c) CeLMoP (category B) bathtub model; (d) CeMoP (category A) bathtub model; (e) CeCoP (category E) bathtub model; and (f) bathtub model extrapolations.

TABLE 72 Bathtub-modelled percentage of patients requiring revision (men aged < 65 years)

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	4.2	10.3	18.9
CeLMoP (category B)	6.9	20.7	39.0
CeLCoC (category C)	5.4	14.3	27.0
HyMoP (category D)	5.3	13.8	26.0
CeCoP (category E)	2.9	8.5	19.7

Percentages refer to the mean age of patients in each category.

Comparison of total hip replacement revision rates according to sex and age: women aged less than 65 years

Figure 57 shows the observed time to revision for female patients aged < 65 years according to category of THR prosthesis. Observed revision was less frequent for CeCoP (category E) than for other categories. Parametric fits to the observed data are shown in *Appendix 17* and AIC values for models are summarised in *Appendix 19*. Cumulative hazard plots are shown in *Appendix 20*. According to AIC values and visual inspection the bathtub model provided a superior fit to observed data for categories A, C, D and E, but failed to resolve for category B (CeLMoP). Of the tested models for category B, each except for the exponential model generated a decreasing hazard beyond the observed data. For the economic model the bathtub model was selected for all categories except B for which the exponential model was used (this will tend to favour category B over the other categories). The predicted requirement for revision at 10, 20 and 30 years is shown in *Table 73*.

Comparison of revision rates with the National Institute for Health and Care Excellence benchmark

The two previous TA guidance documents (TA44[25] and TA2[46]) suggested a revision rate benchmark of 10% at 10 years for hip replacement interventions. Here we compare the performance of the technologies assessed in this report against this benchmark. It should be noted that the benchmark is derived from an assessment of technologies based on data from approximately 15–20 years ago.

Table 74 summarises our estimates of revision rates at 10 years for the currently examined technologies. It should be noted that these are based on data from the NJR in which follow-up was somewhat short of 10 years so that some extrapolation beyond the observed data was necessary.

It is clear that for each of the THR categories A–E, the revision rate at 10 years is within half the benchmark rate, the CeCoP (category E) prosthesis performing better than the rest. Category A–E THR patients age matched to RS recipients similarly experienced revision rates that were less than half the benchmark rate, and this also nearly applied for the revision rate observed for all THR patients in the NJR.

In contrast, the revision rate for RS recipients as a whole or for RS patients after age matching with THR recipients for both sexes substantially exceeded the benchmark; the rate for women reached 23.1% and the rate for men reached 12.4%.

These results suggest that a new benchmark of < 10% at 10 years would now appear to be appropriate for THR technologies and that RS technologies may require considerable improvement to meet the 10% benchmark.

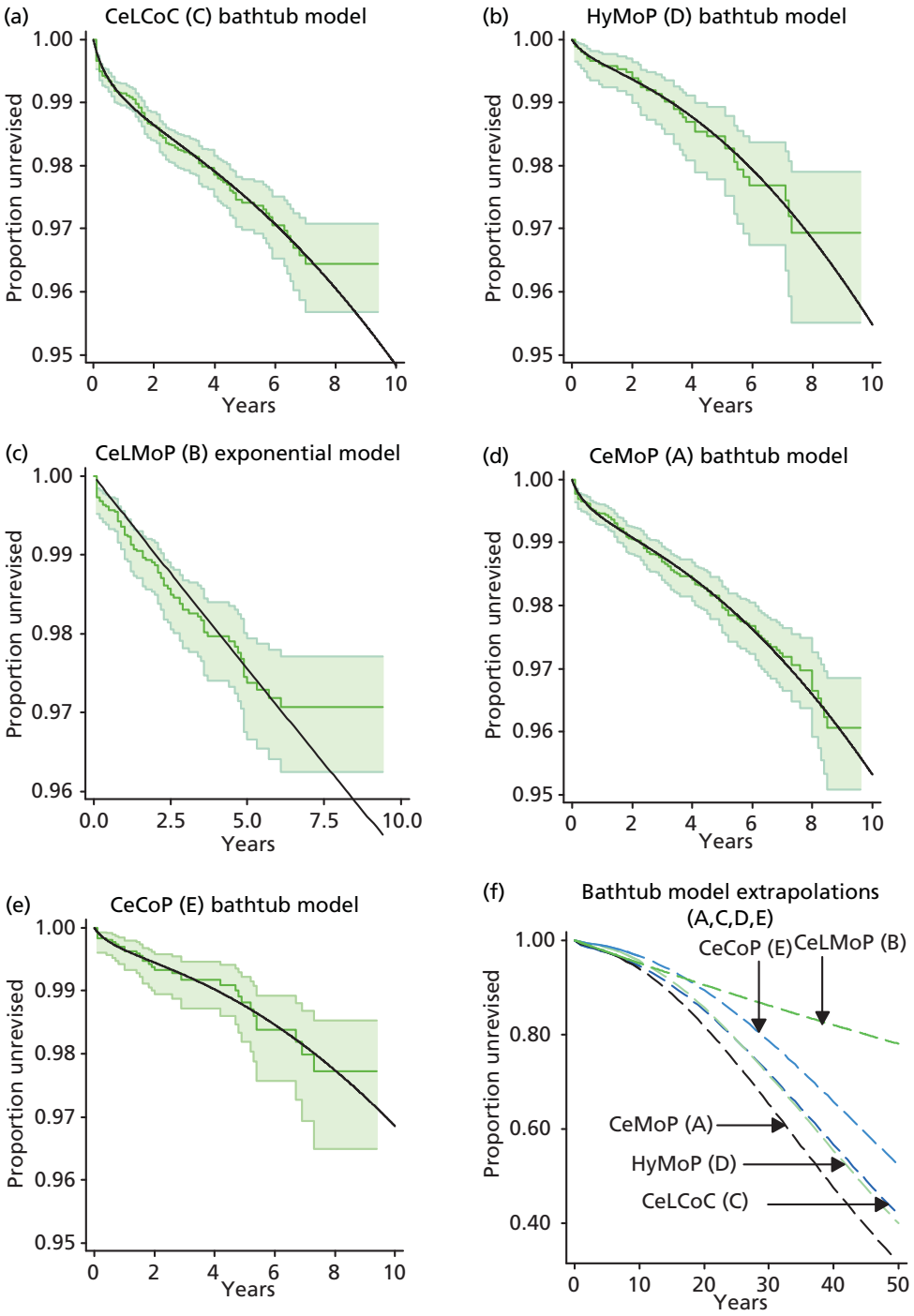


FIGURE 57 Observed revision (95% CI) for women aged < 65 years and bathtub models for THR categories. Note: A bathtub model did not resolve for category B and so an exponential model was used. The extrapolations of models shown apply for patients aged 50 years. (a) CeLCoC (category C) bathtub model; (b) HyMoP (category D) bathtub model; (c) CeLMoP (category B) exponential model; (d) CeMoP (category A) bathtub model; (e) CeCoP (category E) bathtub model; and (f) bathtub model extrapolations.

TABLE 73 Modelled percentage of patients requiring revision (women aged < 65 years)[a]

	Revision (%)		
THR category	10 years	20 years	30 years
CeMoP (category A)	4.7	14.3	28.0
CeLMoP (category B)	4.8	9.4	13.8
CeLCoC (category C)	5.2	14.2	27.1
HyMoP (category D)	4.5	14.9	29.7
CeCoP (category E)	3.1	10.0	20.3

a Bathtub models were used for each category other than category B for which an exponential model was used. Percentages refer to the mean age of patients in each category.

TABLE 74 Estimated percentage of patients requiring revision at 10 years

Intervention	Population	Revision at 10 years (%)
RS	All NJR patients ($n = 31,222$)	14.4
RS	Matched population ($n = 26,643$)	17.2
RS	Female matched ($n = 9321$)	23.1
RS	Male matched ($n = 17,322$)	12.4
THR	Categories A–E matched to RS ($n = 26,643$)	4.7
THR	All NJR patients ($n = 386,566$)	5.2
THR	All CeMoP (category A) ($n = 125,285$)	2.8
THR	All CeLMoP (category B) ($n = 37,874$)	3.9
THR	All CeLCoC (category C) ($n = 34,754$)	4.7
THR	All HyMoP (category D) ($n = 28,471$)	3.0
THR	All CeCoP (category E) ($n = 12,705$)	2.1

Flexible parametric modelling

Several recent analyses of revision rates for patients in the NJR have employed the flexible parametric procedure of Parmar and Lambert.[361] As far as we are aware no economic models for hip replacement have yet employed this approach. We therefore employed flexible parametric modelling in sensitivity analysis of revision rates to determine whether conclusions based on methods described above might be at odds with results from flexible parametric modelling.

In general, flexible parametric models generated good fits to the Kaplan–Meier estimates of observed revision rates (see *Appendix 23*); in some instances, AIC scores were as good as or better than those for alternative models. With regard to different THR categories, revision rates gradually decreased on extrapolation, and rates were sometimes greater and sometimes lesser than those predicted by the Weibull and log-normal models (see *Appendix 23*); as with the base-case bathtub model and the log-normal model, the CeCoP (category E) prosthesis provided the lowest modelled revision rate. With regard to the comparison between RS and THR, for both men and women, as with the base-case bathtub model, flexible modelling yielded considerably higher rates of revision than the log-normal or Weibull model (see *Appendix 23*).

Increasing the number of knots in the flexible parametric modelling improved goodness of fit and modified the extrapolated revision rates such that predicted revision beyond the observed data appeared to be more influenced by the tail of the observed data where the observations were subject to greater uncertainty. This did not necessarily appear to offer an advantage over alternative models. Furthermore, there was no obvious way of determining the number of knots likely to generate the most reasonable extrapolation. Therefore, in sensitivity analysis we used three knots.

Discussion of methods of modelling revision rates

In the NJR twice as many men as women received RS, whereas 1.7 times as many women as men received a category A–E THR, with the mean age for RS recipients nearly 15 years lower than that for THR recipients. The number of THR recipients outnumbered RS recipients by about 10 : 1. When observed revision rates over about 9 years of follow-up were compared between the total THR population and the total RS population they were found to be about three times higher for RS. The difference was greater for women than for men (nearly fourfold and about twofold, respectively). When the comparisons were made between RS and the most frequently used categories of THR, these differences were greater.

All THR categories for both men and women had far lower revision rates than that of RS. Because of the age and sex imbalances between the RS population and the THR population we used propensity matching by age and sex to generate a THR population that would allow an equitable comparison between the RS and the THR interventions. This did not disadvantage RS relative to THR because the younger THR matched population exhibited higher rates of revision than did the whole THR population. The revision rate for RS controlled for age was substantially greater than that for THR. This held for both men and women and, when carried through to the economic analysis, this translated to the association of higher costs with RS than with THR.

The number of unique THR prostheses used for NJR patients was large, even without taking into account the variety of manufacturer brands available for the different prosthesis components. It was necessary to reduce these to a smaller number for economic analysis. Selection was based on the frequency of use of different categories of prosthesis and on expert clinical opinion. The selection of the five THR categories was conducted pre hoc and before all analyses of revision rates. Just over 239,000 patients in the NJR received one of the five selected categories of THR prostheses. The observed revision rates were lowest for CeCoP (category E) and highest for CeLCoC (category C) and CeLMoP (category B) THR. This reflects practice over the last 9–10 years.

Age and sex distributions varied between categories; however, when populations were controlled for differences in age and sex, or were stratified by sex and controlled for age, the lower revision rate for CeCoP (category E) THR relative to the other categories was not diminished. Also, when well-fitting models were used that predicted either increasing hazard or decreasing hazard on extrapolation, the superiority of the CeCoP (category E) revision rate was again upheld. There was insufficient information consistently recorded within the NJR for investigation of other potential confounders. Several potentially influential factors might determine the observed differences in revision rates; these include different prosthesis designs, different patients, different surgical performance and different orthopaedic centres. NJR data were complete for patient age and sex on receipt of THR.

For economic modelling we used the revision estimates from Kaplan–Meier analysis. This conforms with the practice of previous hip replacement cost-effectiveness models found in the literature. McMinn *et al.*[318] aptly define the inference of such analyses as follows: 'inferences about, and comparisons of, revision rates at any time relate to patients who are not already dead at that time'. This was considered appropriate for the structure of the economic model.

To model revision rates we followed NICE DSU guidance in first exploring exponential, Weibull, Gompertz, log-normal and log-logistic models of observed revision rates based on IPD; these commonly used parametric fits are readily available within statistical packages (such as Stata) and an initial consideration of goodness of fit can be obtained, for example from the AIC and BIC.[365] However, most economic analyses of hip replacement, notably those of Briggs *et al.*,[38] Higashi *et al.*[273] and Pennington *et al.*,[44] modelled revision rates on the assumption of a U-shaped hazard. In these analyses an assumed high hazard for failure associated with surgery is followed by a decreasing hazard that eventually plateaus during an initial recovery period and is then followed by a gradually increasing hazard as host bone deteriorates with patient age and the prosthesis accumulates wear and tear. The resulting hazard curve forms a 'U' shape commonly termed a bathtub. We therefore also explored bathtub models.

The NJR observation period for both RS and THR patients extended to about 9 years. NICE requires a lifetime economic model to capture all benefits (and harms) of interventions; therefore, extrapolation of revision rates beyond the observed data was required. In most of the comparisons undertaken for this report the extrapolation of most models predicted a decreasing rate of revision (i.e. decreasing hazard); however, the bathtub models all described an increasing revision rate beyond the observed period. Increasing the hazard of revision appears reasonable for patients who are relatively young at the time of primary hip replacement and who might be expected to live with their prosthesis for ≥ 30 years. For older age groups it may be argued that a model predicting an increasing hazard for revision is unsuitable as, relative to younger generally more active patients, the prosthesis is subject to less wear and tear for a shorter time. The observed rate of revision during the observation period for NJR patients aged > 85 years was very low and minor relative to attrition because of death (see *Figure 26*). It is clear that for patients of advanced age there is a relative lack of clinical imperative to undertake revision and an extrapolation with an increasing hazard becomes less appropriate.

Published economic models of hip replacement have adopted various solutions for modelling THR revision rates. In common with several of these we modelled revision rates in the base case using a U-shaped (bathtub) hazard assumption.[38,44,273] This was supported by the goodness of fit to the observed data according to visual inspection, information criteria scores and plots of the log-Kaplan–Meier-estimated cumulative hazard compared with the log-modelled cumulative hazard.[359] Published analyses with long-term follow-up of patients also support increasing revision rates beyond 10 years from the primary intervention. Previous studies obtained an overall bathtub hazard by combining a Weibull fit for early failures with a Weibull fit for late failures.[38,44,273] We derived the bathtub hazard directly using the Stata package developed by Crowther and Lambert.[360] This had the advantages of parsimony and of not requiring arbitrary decisions about early and late failures. Higashi and Barendregt[273] used long-term follow-up studies for the second Weibull fit to obtain an increasing hazard in the long term; however, this suffers the disadvantage that very different populations were used for the early and late fits. Pennington *et al.*[44] employed a piece-wise procedure to generate the U-shaped hazard; however, after extrapolation this predicted that > 100% of patients sustained revision and at this point the rate required capping.

For revision rates the unit of analysis was the time to a patient's first revision. For patients who received THR for both hips simultaneously only the replacement that failed first was included as an event, and for those who received THR for both hips on separate occasions only the first primary intervention entered the analysis.

For RS a wide range of different femoral head sizes are used and revision rates have been reported to vary according to head size.[15] Only a narrow range of different head sizes are used for THR prostheses and expert clinical advice indicated that these are unrelated to RS head sizes so that comparisons between RS and THR according to head size were not undertaken.

Summary

The Kaplan–Meier-estimated rates of revision during approximately 9 years of follow-up of NJR patients indicated that the probability of revision differed between interventions. RS had a considerably higher frequency of revision than THR; this held across both sexes. The five categories of THR selected also differed in the observed revision rate, with CeCoP (category E) tending to have a lower rate of revision than other categories; again, this held generally across age groups and sex.

For all interventions several parametric models generated good fits to the observed data. The differences between models with a good fit over the observation period were minor relative to differences generated on extrapolation. Extrapolations generated from well-fitting models could be broadly divided into those predicting a gradual increase in rate of revision with time (usually, but not always, these were bathtub models) and those predicting a gradual decrease in rate of revision with time. Data summarised in *Appendix 24* from several sources (the Swedish registry,[95] the RCT of Kim *et al.*[129] and long-term follow-up observational studies[368–372]) tended to support the proposition of an increasing hazard, at least for the first decade or so beyond the 9 years of NJR data.

On the other hand it is clear that NJR patients who receive a THR in old age (e.g. > 85 years) have a low probability of surgery for THR revision. In general, it appears likely that revisions beyond the observed data first occur at an increasing rate and later at a decreasing rate. The parametric fits did not capture this putative pattern well and it is difficult to ascertain when rates might change from increasing to decreasing for different age groups. However, the lower rate of revision seen for CeCoP (category E) THR relative to other categories was maintained across models that differed in the direction of the hazard after extrapolation beyond the observed data.

The differences between models in the extrapolation of revision rates require about a decade beyond the observation period before becoming substantial. By that time discounting and higher mortality rates will tend to attenuate the influence of differing extrapolations on the results from an economic model. Therefore, it may be anticipated that, over a lifetime, different modelling approaches to extrapolation (increasing hazard for each intervention or alternatively decreasing hazard for each) might not have a large influence on the economic outcomes for the interventions relative to their observed differences.

Our assessment of THR and RS against the revision rate benchmark from TA2[46] and TA44[25] of 10% at 10 years suggests that a new benchmark of < 10% at 10 years would now appear to be appropriate for THR technologies, but that RS technologies may still require considerable improvement to meet the 10% benchmark.

Chapter 8 Warwick economic assessment

This chapter describes the structure of the economic model, the main assumptions of the model, the scenarios evaluated and the sensitivity analyses. The underlying model is based on that of Fitzpatrick *et al.*,[373] which has been adapted for our decision problem and updated with new data.

Methods

De novo analysis

Patients

We used NJR data to investigate revision rates. Detailed information on this is given in *Chapters 5* and *7*.

We used propensity matching to match by age and sex NJR THR category A–E patients with RS patients. These matched populations were used to generate modelled revision rates for our economic model for the base case for decision problem (1) (see *Chapter 2*). Furthermore, we performed subgroup analyses in which RS and THR matched populations were stratified by sex, and models of time to revision were controlled for age. For decision problem (2) (see *Chapter 2*), in the base case we compared THR categories A–E, irrespective of age and sex. In sensitivity analysis we controlled for age and sex. For subgroup analysis we stratified by age (< 65 years and > 65 years) and by sex, and the modelled time to revision was controlled for age. The selection of the subgroup aged > 65 years reflected a population unlikely to be considered suitable only for THR and not suitable for RS (see *Table 60* for population details).

Model structure

An economic model was developed based on a Markov multistate model, as shown in *Figure 58*.

In the model, each patient can enter one of four health states following primary surgery:

1. Successful primary (RS or THR) surgery (if initial surgery is successful, patients enter this health state).
2. Revision surgery arises at the second-year cycle (if initial surgery fails, patients may then require a revision). If necessary, patients can move into this state more than once. Patients stay in this health state for one cycle only.
3. Successful revision surgery (if revision surgery is successful, patients enter this health state).
4. Death (this is an absorbing health state and patients may enter this state because of operative mortality or because of death from other causes).

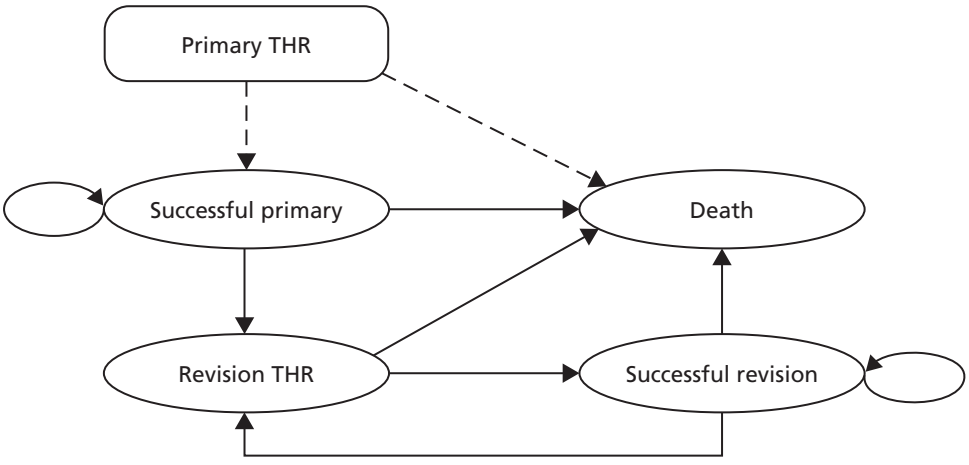


FIGURE 58 Markov model. Reproduced with permission from Fitzpatrick *et al.*[373]

For RS compared with THR and for different categories of THR compared with each other, similar models were built (see *Figure 58*), with different estimates of transition probabilities, utilities and costs.

The cycle length for each model was set at 1 year and transitions between each health state occur at the end of each cycle. Before submission of the final report, a third party who was not directly involved in the assessment cross-checked the inputs to the model and fully rebuilt the model as a structural cross-check. All discrepancies were discussed with the assessment team and the appropriate final set of model inputs and model structure were agreed on for the final report.

Based on the external assessment, it was assumed that all THR events occurred at the start of the annual cycle, with mortality from other causes (non-THR events) occurring at the end of each cycle. We also noticed that the estimates for the first-year revision rates were high over the first several months after implantation of a prosthesis but that for category E this was less pronounced than for other categories. Therefore, the transition from successful primary health state to revision THR was assumed to occur at any time and was not specified as occurring at the start of the second annual cycle.

For both review questions we adopted a 10-year and a lifetime horizon. The 10-year time horizon reflects observed IPD from the NJR, and the lifetime horizon follows the recommendation from NICE that the time horizon should be sufficiently extended to capture all benefits likely to accrue from an intervention.[374] The analysis was conducted from the perspective of the NHS and personal social services (PSS). All costs are in pounds sterling in 2011/12 prices. Health outcomes were measured in QALYs. Results are expressed as incremental cost per QALY gained. An annual discount rate of 3.5% was applied to both costs and outcomes.[374]

The key features of the analysis are listed in *Table 75.*

TABLE 75 Key features of the analysis

Element of HTA	Reference case	Section in *Guide to the Methods of Technology Appraisal*[374]
Defining the decision problem	Clinical effectiveness and cost-effectiveness analysis of different types of THR and RS for the treatment of pain and disability in people with end-stage arthritis of the hip (scope developed by NICE[375])	5.2.5 and 5.2.6
Comparator(s)	Different types of primary THR compared with surface replacement for people in whom both procedures are suitable; different types of primary THR compared with each other for people who are not suitable for hip RS	5.2.5 and 5.2.6
Perspective costs	NHS and PSS	5.2.7–5.2.10
Perspective benefits	All health effects on individuals	5.2.7–5.2.10
Type of economic evaluation	Cost-effectiveness analysis	5.2.11 and 5.2.12
Synthesis of evidence on outcomes	Based on NJR database	5.3
Measure of health effects	QALYs	5.4
Source of data for measurement of health-related quality of life	Based on PROMs database (reported directly by patients and carers)	5.4
Source of preference data for valuation of changes in health-related quality of life	Representative sample of the public	5.4
Discount rate	An annual rate of 3.5% on both costs and health effects	5.6
Equity weighting	An additional QALY has the same weight regardless of the other characteristics of the individuals receiving the health benefit	5.12

Base-case analysis

For the base-case analysis we estimated the cost-effectiveness of THR compared with RS for patients who were eligible for both procedures using revision rates modelled using a bathtub model. Utilities for successful implant health states were varied with patient age throughout the model. Costs were based on NHS Supply Chain costs (Dr Philip Lewis, NHS Supply Chain, 2013, personal communication).

Similarly, we estimated the cost-effectiveness of the different categories of THR prostheses using revision rates based on the bathtub model. Utilities for successful implant health states were varied with patient age throughout the model. Again, costs were based on NHS Supply Chain costs (Dr Philip Lewis, NHS Supply Chain, 2013, personal communication).

Structural model assumptions

Transition probabilities

Time to revision was described according to well-fitting parametric models (the base case for the comparison of THR with RS and for the comparison between different THR categories was based on the bathtub model; in sensitivity analysis of THR compared with THR a log-normal parametric model was used, adjusted for age and sex). The risk of rerevision was based on rerevision rates obtained from the manufacturer's submissions to NICE (sourced from the New Zealand joint registry[376] by the manufacturer).

Utilities

Utilities for both models for the base-case analysis were obtained from the PROMs database (see *Chapter 6*). The mean EQ-5D-3L scores for the successful primary health state and successful revision health state were reduced by the mean EQ-5D-3L scores for the respective age band and sex at the end of each 10-year cycle to represent the impact of ageing on general health-related quality of life. The age-related utilities were assumed to be the same for the comparison of RS with THR and for the comparison of different types of THR. We assumed that at 6 months patients would have fully recovered from the surgery and this assumption was supported by the EQ-5D-3L responses obtained from patients at baseline, 3 months, 6 months and 12 months from Edlin *et al.*[40]

Costs

For the comparison of THR with RS and the comparison between different types of THR, prices of the primary prostheses were based on the list prices obtained from the NHS Supply Chain. We assumed that, for the comparison between THR and RS, if initial RS surgery failed the patient would then be revised with a THR prosthesis and not a RS prosthesis. The prices of the revision prosthesis and the rerevision prosthesis were obtained from Vanhegan *et al.*[292] based on a weighted average of the mean costs of all revision procedures. For the comparison between different types of THR, we assumed that, if initial THR surgery failed, the same type of prosthesis was used for each category. Hence, we included the mean implant cost from Vanhegan *et al.*[292] based on a weighted average of the mean costs of all revision procedure.

For both sets of comparisons we included follow-up costs in the first year after surgery and the surgical cost of adverse event(s) resulting in revision surgery but because of a lack of reliable data we were not able to include the cost of other treatments for adverse events in the months following revision surgery. We have also not included end-of-life costs[19,373] (*Table 76*).

TABLE 76 Summary of assumptions

Parameter	Assumptions
Transition probabilities	Time to revision was assumed to be described according to well-fitting parametric models. The risk of rerevision was based on the rerevision rate obtained from the manufacturer's submissions to NICE
Utilities	Utilities for the base-case analysis were obtained from the PROMs database. The utilities were assumed to be the same for the comparison between RS with THR and the comparison between different types of THR
Costs	For the comparison between THR and RS and the comparison between different types of THR, the prices of the primary prostheses were based on the list prices obtained from the NHS Supply Chain. The price of the revision prosthesis and the rerevision prosthesis were obtained from Vanhegan *et al.*[292] based on a weighted average of the mean costs of all revision procedure

Estimation of model parameters

Resource use and cost inputs

Resource use and associated costs were required for the following health states:

1. successful primary procedure
2. revision procedure
3. successful revision procedure.

Health states 1 and 2 have two phases: a short-term phase with costs associated with surgery and the immediate aftermath of surgery, followed by a more prolonged phase including the costs of maintenance.

Rationale for the choice of parameter values

The process of identifying the relevant literature can be found in *Chapter 6*. Of the 11 core studies, three cost-effectiveness studies provided data for the economic model. These were the studies by Edlin *et al.*,[40] Vale *et al.*[19] and Vanhegan *et al.*[292]

Edlin *et al.*[40] reported a cost–utility analysis of RS compared with THR alongside a RCT using NHS and PSS perspectives and costs were reported as UK pounds in 2009/10 prices. The study used HRG4 reference costs combined with NHS trust finance department list prices for implants and IPD on LOS. Resource use data and personal costs were obtained from patient-reported data. The study reported costs after 12 months by type of hip replacement (THR vs. RS) including the costs of initial operation/care, subsequent inpatient, outpatient, primary and community care, aids and medications, as well as private and social costs.

Vale *et al.*[19] assessed the clinical effectiveness and cost-effectiveness of RS compared with watchful waiting (i.e. patient monitoring, drug-based treatment and supportive activities including physiotherapy), THR and other bone-conserving treatments.[19] Cost data were reported in UK pounds in 2000/1 prices; costs for THR and revision THR were taken from the literature and prostheses costs for RS were obtained from manufacturers. Cost components for surgical interventions including use of the operating theatre, staff, radiography, outpatient visits and first-year follow-up costs were reported.

Vanhegan *et al.*[292] investigated the costs of revision THR. Costs were reported in UK pounds in 2007/8 prices and were obtained from the finance department of the tertiary centre and included costs of the implant, materials and augmentation, use of the operating theatre and recovery room, the inpatient stay and laboratory tests, radiology, pharmacy, physiotherapy and occupational therapy. The study provided cost data on 13 different implants and data on resource use and costs by reason for revision (aseptic loosening, deep infection, periprosthetic fracture and dislocation).

All three core studies provided important and relevant costs for THR and RS patients for use in the economic model, with prices updated to 2011/12 prices by applying the projected Health Service Cost Index (HSCI).[377] It is also important to mention that none of the studies identified in the literature included costs per component of prosthesis as grouped in our analysis.

Base-case cost inputs: resurfacing arthroplasty compared with total hip replacement

The cost of the primary THR or RS includes the cost of the prosthesis, the initial operation and the inpatient hospital stay. The cost of the RS prosthesis was obtained from the NHS Supply Chain (Dr Philip Lewis, NHS Supply Chain, 2013, personal communication). Information provided detailed the full list price for three suppliers using their most common brands of implant. These data were anonymised by averaging the cost for each component (*Table 77*). In real life these prices are often discounted (using a discount de-escalator based on the volume of the purchase).

The costs of the THR prostheses were also obtained from the NHS Supply Chain. We obtained the full list price for the five most commonly used suppliers (details of suppliers were anonymised) using their most common brands of implant. We calculated a weighted mean THR cost based on the frequency of use of the different categories of THR (categories A–E) in the RS vs. THR comparison (*Table 78*).

The cost of the surgery itself was assumed to be the same for both THR and RS. The costs of theatre overheads, theatre staff and number of radiographs, etc. were taken from Vale *et al.*[19] and updated to current prices.[377] The total cost of surgery was estimated at £2805 (*Table 79*).

TABLE 77 Resurfacing arthroplasty prosthesis costs as reported by the NHS Supply Chain

		Supplier list price (£)		
Component	Average unit cost (£)	Supplier 1	Supplier 2	Supplier 3
Acetabular cup, HA coated	1583	1690	1535	1523
Resurfacing head, cemented	1031	1140	865	1089
Mixing bowl[a]	31	NA	NA	NA
Cement (one pack)[a]	27	NA	NA	NA
Total cost	2672			

HA, hydroxyapatite; NA, not applicable or not reported.
a The price for this item is sourced from one supplier.

TABLE 78 Total hip replacement prosthesis costs as reported by the NHS Supply Chain

Category	Number of male patients	Number of female patients	Total number of patients	Mean cost (£)	Weighted cost (£)
A	6080	3812	9892	1557	589
B	2177	741	2918	3016	336
C	5803	2414	8217	3869	1215
D	1104	477	1581	2650	160
E	2100	1459	3559	1996	271
Weighted cost of THR prosthesis					2571

TABLE 79 Total cost of surgery

	1996 prices		2011/12 prices
Resource use	Primary THR (units)	Total cost (£)	Total cost (£)
Theatre overheads	134 minutes	655	1799
Theatre staff	–	232	637
Number of radiographs	6	134	368
Total cost per patient			2805

The average LOS was based on point estimates as reported in Edlin *et al.*[40] The total cost of the inpatient stay for RS was estimated to be £1628. This was based on an average cost per day of a hospital stay of £296, multiplied by the average LOS of 5.5 days.[40] The average LOS for THR was 5.7 days and the total cost of the inpatient stay for THR was estimated to be £1687. RS was associated with a slightly shorter LOS (5.5 vs. 5.7 days); although this difference was not statistically significant, we assigned this slightly shorter LOS so as not to overestimate the cost of RS.

Cost of a revision procedure (total hip replacement or resurfacing arthroplasty)

The costs of revision were assumed to be the same for both THR and RS. The cost of a revision hip arthroplasty was obtained from Vanhegan *et al.*;[292] the data were based on 305 successive revisions following THR in 286 patients between 1999 and January 2008. In this study, patient-specific resource use data were reported for the implant, materials, use of the theatre, use of the recovery room, inpatient stay, physiotherapy, occupational therapy and pharmacy, radiology and laboratory, with costs based on NHS 2007/8 rates for payment by results (PbR).

Costs were inflated to 2011/12 prices by applying the projected HSCI.[378] Importantly, the study also reported mean costs for revision surgery for aseptic cases, deep infection, periprosthetic fracture and dislocation. Hence, the cost of revision was calculated based on a weighted average of the mean costs of all revision procedures (*Table 80*).

We used this cost because the frequency of aseptic loosening found in this study is comparable to that reported in the NJR in 2006.

TABLE 80 Costs of revision

Indication	Number of patients	Mean cost (£), 2007/8 prices	Mean cost (£), 2011/12 prices
Aseptic loosening	194	11,897	13,226
Deep infection	76	21,937	24,387
Periprosthetic fracture	24	18,185	20,216
Dislocation	11	10,893	12,109
Weighted average			16,517

Cost of a successful revision procedure (total hip replacement or resurfacing arthroplasty)

The cost of follow-up post primary THR or RS was obtained from the study by Edlin *et al.*,[40] which was based on resource use, using patient-reported data at 3, 6 and 12 months. Cost data on outpatient care, primary and community care, aids and adaptations provided by the NHS/social services, medication (pain relief and other NHS medication) and personal costs (out-of-pocket expenditure such as medicine usage and time off work for either the patient or a carer) were reported for both the THR arm and the RS arm. The NHS and social care costs of follow-up in 2011/12 prices were £394 for the THR arm and £501 for the RS arm at 12 months (*Table 81*).

We used cost data from this study because they were based on a RCT and the mean age of RS patients (56.3 years) in this study was comparable to that reported in the NJR database (55 years).

Base-case cost inputs: comparison of different types of hip replacement

Resource use and cost assumptions were mostly assumed to be the same as for the comparison between THR and RS. The cost of primary THR included the operation cost, prosthesis cost, hospital ward cost and follow-up cost. The cost of the operation were assumed to be the same for all types of prostheses.

The total cost of the inpatient stay was estimated to be £1687, based on the average cost per day of a hospital stay, multiplied by the average LOS (5.7 days), as reported in Edlin *et al.*[40] The total cost of surgery including radiography, theatre time, staff and overheads was estimated at £2805.[377] Outpatient costs and other follow-up costs were estimated to be £394 based on Edlin *et al.*[40] (see *Table 81*).

Prosthesis cost

We were not able to use published costs for the costs of the prostheses because prostheses were grouped as cemented, cementless or hybrid rather than being based on the separately identifiable prosthesis components as categorised in our analysis (categories A–E). Our base-case cost for each category of prosthesis was obtained from the NHS Supply Chain (Dr Philip Lewis, NHS Supply Chain, 2013, personal communication). Anonymised information was available detailing the list price per component for all five categories. The cost data from the five most commonly used suppliers using their most common brands of implant were available and an average cost was calculated. Again, this is subject to a volume de-escalator in price for the NHS (*Table 82*).

The pricing of a bone cement pack including bone cement, mixing devices and pressuriser was available from one supplier only. We have itemised the cost of a bone cement pack for a cemented stem and cup and a cemented stem only (*Table 83*).

TABLE 81 Cost of a successful revision procedure for THR and RS

	2009/10 prices (£)		2011/12 prices (£)	
Costs	Total cost RS	Total cost THR	Total cost RS	Total cost THR
Outpatient care	360	276	383	294
Primary/community care	63	49	67	52
Aids and adaptations	21	21	22	22
Medications	27	24	29	26
Total cost			501	394

TABLE 82 Prosthesis costs for the comparison between different types of THR

Component	Average unit cost (£)	Supplier 1 (£)	Supplier 2 (£)	Supplier 3 (£)	Supplier 4 (£)	Supplier 5 (£)
Category A – CeMoP						
Cemented stem	701.60	625	523	706	798	856
Metal head	297.20	204	231	272	375	404
Polyethylene cup – cemented	249.60	164	227	311	332	214
Cemented stem centraliser	47.50	NA	19	76	NA	NA
Bone cement plug	58.38	44.5	49	NA	81	59
Cemented stem and cup extras	203.10					
Total	1557.38					
Category B – CeLMoP						
Cementless HAC stem	1342.20	1370	1129	1110	1816	1286
Metal stem	292.20	204	231	226	396	404
Metal cup – cementless HA	883.40	910	759	892	941	915
Liner – polyethylene	412.20	190	447	435	547	442
Fixation screw	85.60	82	96	73	74	103
Total	3015.60					
Category C – CeLCoC						
Cementless HAC stem	1342.20	1370	1129	1110	1816	1286
Ceramic head	735.80	620	764	738	857	700
Metal cup – cementless HA	883.40	910	759	892	941	915
Liner ceramic	821.80	815	759	789	1,046	700
Fixation screw	85.60	82	96	73	74	103
Total	3868.80					
Category D – HyMoP						
Cemented stem	701.60	625	523	706	798	856
Metal head	297.20	204	231	272	375	404
Metal cup – cementless HA	883.40	910	759	892	941	915
Liner polyethylene	412.20	190	447	435	547	442
Cemented stem centraliser	47.50	NA	19	76	NA	NA
Bone cement plug	58.38	44.5	49	NA	81	59
Fixation screw	85.60	82	96	73	74	103
Cemented stem extras	163.90					
Total	2649.78					
Category E – CeCoP						
Cemented stem	701.60	625	523	706	798	856
Ceramic head	735.80	620	764	738	857	700
Polyethylene cup – cemented	249.60	164	227	311	332	214
Cemented stem centraliser	47.50	NA	19	76	NA	NA
Bone cement plug	58.38	44.5	49	NA	81	59
Cemented stem and cup extras	203.10					
Total	1995.98					

HA, hydroxyapatite coated; NA, not applicable or not reported.

TABLE 83 Cost of a bone cement pack

Pack	Component	Total cost (£)
Cemented stem and cup	Cement 40-g and 80-g pack	203.10
	Cement syringe	
	Femoral pressuriser	
	Cement mixing pot	
	Acetabular pressuriser	
Cemented stem	Cement 80-g pack	163.90
	Cement syringe	
	Femoral pressuriser	
	Cement mixing pot	

A summary of the transition probabilities, utilities and cost inputs for the cost–utility model

The justification for the transition probabilities between health states based on parametric models of time to revision consisted of model diagnostic plots, visual goodness of fit and information criteria. Prostheses costs were obtained from the NHS Supply Chain as alternative sources of information were lacking.

Utilities were calculated from information in the PROMs database. This was justified because it represented patient-centred EQ-5D-3L data in a population appropriate to the decision problem and the NJR database.

Costs used for the elements of the interventions were justified on the basis of our literature search for relevant information. Mortality associated with surgery was adapted from the value common to all other hip replacement models.

The bathtub parameters used to calculate the transition probabilities between health states employed for the base case are summarised in *Table 84*.

Table 85 provides a summary of the inputs (transition probabilities, utilities and costs) used in the base-case analysis.

TABLE 84 Bathtub parameters for the comparison between RS and THR and the comparison between different types of THR

Comparison	Prosthesis	Bathtub alpha	Bathtub beta	Bathtub gamma
RS vs. THR (matched)				
Base case	RS	0.0030976	0.0358272	3.971709
Base case	THR	0.0005699	0.0123899	1.918951
THR vs. THR				
Base case	CeMoP (category A)	0.0003396	0.0083374	2.163733
Base case	CeLMoP (category B)	0.0004045	0.0337383	6.832735
Base case	CeLCoC (category C)	0.0005333	0.0236369	4.051712
Base case	HyMoP (category D)	0.0003642	0.0158328	4.68618
Base case	CeCoP (category E)	0.0001935	0.0039017	0.6967542

TABLE 85 Summary of the transition probabilities, utilities and cost inputs used in the base-case analysis

Inputs	Mean value	SE	Distribution		Source
Transition probabilities					
Surgical mortality[a]	0.0050	0.001			NJR[48]
Risk of rerevision	0.0326	NA			DePuy submission
			Beta distribution, parameter alpha	**Beta distribution, parameter beta**	
Utilities					
Age 50–60 years	0.7529	0.004	1296	488	PROMs[97]
Age 60–70 years	0.7789	0.002	7397	2427	PROMs[97]
Age 70–80 years	0.7637	0.002	22,244	6315	PROMs[97]
Age 80+ years	0.7210	0.003	28,054	8681	PROMs[97]
Revision surgery	0.5624	0.340	9092	3518	PROMs[97]
			Gamma distribution, parameter alpha	**Gamma distribution, parameter beta**	
Costs (£) *RS vs. THR*					
RS					
Prosthesis cost	2778	NA	NA	NA	NHS Supply Chain
Surgery costs (excluding prosthesis)	1485	NA	NA	NA	Vale *et al.*[19]
Hospital inpatient stay	1628	NA	NA	NA	Edlin *et al.*[40]
Successful primary RS	501	44	130	4	Edlin *et al.*[40]
Revision surgery	16,517	456	1314	13	Vanhegan *et al.*[292]
Successful revision surgery	394	30	169	2	Edlin *et al.* 2012[40]
THR					
Prosthesis cost	2571	NA	NA	NA	NHS Supply Chain
Surgery costs (excluding prosthesis)	1485	NA	NA	NA	Vale *et al.*[19]
Hospital inpatient stay	1687	NA	NA	NA	Edlin *et al.*[40]
Successful primary THR	394	30	169	2	Edlin *et al.*[40]
Revision surgery	16,517	456	1314	13	Vanhegan *et al.*[292]
Successful revision surgery	394	30	169	2	Edlin *et al.* 2012[40]
Different types of THR					
Category A – CeMoP	1557	NA	NA	NA	NHS Supply Chain
Category B – CeLMoP	3017	NA	NA	NA	NHS Supply Chain
Category C – CeLCoC	3869	NA	NA	NA	NHS Supply Chain
Category D – HyMoP	2650	NA	NA	NA	NHS Supply Chain
Category E – CeCoP	1996	NA	NA	NA	NHS Supply Chain
Other costs (£)					
Surgery costs (excluding prosthesis)	1485	NA	NA	NA	Vale *et al.*[19]
Hospital inpatient stay	1687	NA	NA	NA	Edlin *et al.*[40]
Successful primary THR	394	30	169	2	Edlin *et al.*[40]
Revision surgery	16,517	456	1314	13	Vanhegan *et al.*[292]
Successful revision surgery	394	30	169	2	Edlin *et al.* 2012[40]

NA, not applicable or not reported.
a Surgical mortality was the same for THR, RS and revision.

Cost-effectiveness analysis

The base-case analysis is based on costs and outcomes for all THR and RS patients over two time horizons: 10 years and lifetime.

For the RS compared with THR base-case analysis, the male and female patients who received RS were successfully propensity matched by age with THR patients from THR categories A–E, and transition probabilities were calculated using bathtub model fits (predicting an increasing hazard beyond the 10-year observation period). The bathtub model defines a decreasing followed by an increasing hazard with time according to the equation:

$$\text{Bathtub:} \qquad h = at + \frac{b}{(1+gt)}, \tag{4}$$

where *a, b* and *g* are constant parameters and *t* is time.

For the comparison of different types of THR base-case analysis, transition probabilities were calculated using bathtub model fits for categories A–E.

We report total mean costs and total mean QALYs related to THR and RS, and incremental costs per QALY (ICERs) gained. The cost-effectiveness model for all THR categories had more than two mutually exclusive comparisons; we report total mean costs and total mean QALYs. The categories were ranked in order of increasing cost. We eliminated categories for which another category was cheaper and more effective (simple dominance). If there was a linear combination of two other categories that were more costly and less effective, these were eliminated (extended dominance). With the remaining options, we calculated incremental costs per QALY gained.

We present first the deterministic results, followed by the probabilistic results. To represent the uncertainty in the parameters used in the model and to illustrate sampling uncertainty, we undertook probabilistic analyses using 1000 simulations. The results from these simulations were plotted on a cost-effectiveness plane with 95% CIs. Each point is a simulation from the probabilistic analysis. The plot illustrates the uncertainty surrounding the incremental costs and QALYs for the two groups being compared. We also produced CEACs to illustrate the effect of sampling uncertainty, in which individual model parameters were sampled from the appropriate probability distribution. CEACs were reported for a WTP threshold from £0 to £50,000. The perspective taken is from the UK NHS and PSS. Discounting of costs and benefits at 3.5% was undertaken according to UK guidelines.[374]

Sensitivity analyses

Sensitivity analyses were conducted by altering base-case inputs to the model. Several types of subgroup and scenario analysis were explored, encompassing changes to the RS/THR comparison and the THR/THR comparison.

Subgroup analysis for the comparison between resurfacing arthroplasty and total hip replacement and the comparison between different types of total hip replacement

(a) Revision rates were much higher for women receiving RS than for men and, because the revision rate varies according to the age of the patient, subgroup analyses focused on comparing populations stratified by sex and controlled for age. Therefore, in the sensitivity analysis we separately compared the cost-effectiveness of RS and THR for men and women aged 40, 50, and 60 years at the time of the primary implant, using age-matched populations and a bathtub model stratified by sex and controlled for age.

(b) For THR compared with THR, the modelled time to revision was stratified by age (< 65 years and > 65 years) and sex and models were controlled for age. We undertook these subgroup analyses because the use of different categories of THR prosthesis differed by age and sex and because recipients of hip replacement interventions aged > 65 years approximate a population unlikely to be considered candidates for RS. We compared the cost-effectiveness of different types of THR for patients aged < 65 years (40, 50 and 60 years) using a bathtub model and for patients aged > 65 years (70 and 80 years) using a log-normal model (*Table 86*).

For subgroup analyses, mean EQ-5D index scores were split by sex and age band (*Table 87*).

TABLE 86 Subgroup analysis: time to revision for RS vs. THR and THR vs. THR

Sex	Prosthesis	Bathtub alpha	Bathtub beta	Bathtub gamma	Bathtub age coefficient
RS vs. THR (matched): bathtub parameters					
Male	RS	0.0020179	0.0370237	4.443342	–0.0380901
Male	THR	0.0006006	0.0135972	2.384484	–0.0258836
Female	RS	0.0044984	0.0280047	2.558539	–0.0118076
Female	THR	0.0005964	0.0099966	1.314233	–0.016463
THR vs. THR: bathtub parameters					
Male < 65 years	CeMoP (category A)	0.0003869	0.008084	0.7177154	–0.0207576
Male < 65 years	CeLMoP (category B)	0.0010417	0.0245433	4.822729	–0.0024683
Male < 65 years	CeLCoC (category C)	0.0006243	0.0212657	3.032461	–0.0110798
Male < 65 years	HyMoP (category D)	0.0005998	0.0237569	3.576745	–0.0172004
Male < 65 years	CeCoP (category E)	0.0004695	0.0033726	1.782609	–0.0327686
Female < 65 years	CeMoP (category A)	0.0006692	0.0132853	3.675229	–0.0293667
Female < 65 years	CeLMoP (category B)	Not resolved			
Female < 65 years	CeLCoC (category C)	0.0006154	0.0215004	3.952961	–0.0088734
Female < 65 years	HyMoP (category D)	0.00076	0.0077105	3.21092	0.0048101
Female < 65 years	CeCoP (category E)	0.0004703	0.0071811	3.211915	–0.0078225
		Log-normal mu	**Log-normal sigma**	**Log-normal age coefficient**	
THR vs. THR: log-normal parameters					
Male > 65 years	CeMoP (category A)	10.37363	4.075863	0.0020929	
Male > 65 years	CeLMoP (category B)	10.52551	4.554688	–0.0483328	
Male > 65 years	CeLCoC (category C)	9.611438	4.12394	–0.0448092	
Male > 65 years	HyMoP (category D)	10.31021	4.093764	0.0126215	
Male > 65 years	CeCoP (category E)	10.54446	3.971899	–0.0407056	
Female > 65 years	CeMoP (category A)	9.815575	3.636813	0.033098	
Female > 65 years	CeLMoP (category B)	12.10535	5.138115	–0.0241371	
Female > 65 years	CeLCoC (category C)	11.471	4.744101	–0.0287428	
Female > 65 years	HyMoP (category D)	12.18021	4.757849	0.0504173	
Female > 65 years	CeCoP (category E)	10.13035	3.562737	0.0631827	

TABLE 87 Summary of utility inputs for subgroup analysis

Age group (years)	Mean value	SE	Beta distribution, parameter alpha	Beta distribution, parameter beta	Source
Men					
40–50	0.736	0.0179	443	159	PROMs[97]
50–60	0.767	0.0066	3133	952	PROMs[97]
60–70	0.762	0.0038	9112	2393	PROMs[97]
70–80	0.790	0.0034	11,488	3054	PROMs[97]
80+		0.0071	2816	964	PROMs[97]
Women					
40–50	0.720	0.0129	872	339	PROMs[97]
50–60	0.742	0.0058	4287	1491	PROMs[97]
60–70	0.769	0.0032	13,128	3944	PROMs[97]
70–80	0.747	0.0029	16,732	5667	PROMs[97]
80+	0.710	0.0048	6305	2575	PROMs[97]
Revision surgery					
Males	0.575	0.009	1496	1106	PROMs[97]
Females	0.553	0.007	2201	1779	PROMs[97]

Sensitivity analysis around the base-case time to revision for the comparison between resurfacing arthroplasty and total hip replacement

(a) The bathtub model was controlled for age and sex because the age distributions of the matched populations were somewhat removed from the normal distribution (see *Chapter 7*). Transition probabilities were then calculated for the average population (35% female, age 55.8 years) (*Table 88*).

Sensitivity analyses around the base-case time to revision for the comparison between different types of total hip replacement

(a) The bathtub model was controlled for age and sex. This was carried out because both age and sex differed between categories and both variables influenced the time to revision (see *Chapter 9*). Transition probabilities were then calculated for the age and sex mix across all five categories (63.5% female, age 71.6 years).
(b) A log-normal model was used because the information criteria scores and the visual plot for this model showed it to be the next best fit after the bathtub model, while providing a decreasing hazard on extrapolation that may be more suitable for older populations.
(c) A log-normal model controlled for age and sex was used because both age and sex differed between categories and both variables were associated with time to revision. Transition probabilities were then calculated for the age and sex mix across all five categories (63.5% female, age 71.6 years) (*Table 89*).

TABLE 88 Sensitivity analysis: time to revision for RS vs. THR

Comparison	Prosthesis	Bathtub alpha	Bathtub beta	Bathtub gamma	Bathtub age coefficient	Bathtub sex coefficient
RS vs. THR (matched)						
Sensitivity analysis	RS	0.00373026	0.04400835	3.8505838	–0.02491814	–0.4098118
Sensitivity analysis	THR	0.00058692	0.01189397	1.989425	–0.02238228	0.05307551

TABLE 89 Sensitivity analysis: time to revision for THR vs. THR

Comparison	Prosthesis	Bathtub alpha	Bathtub beta	Bathtub gamma	Bathtub age coefficient	Bathtub sex coefficient
THR vs. THR – bathtub model controlled for age and sex						
Sensitivity analysis	CeMoP (category A)	0.0003132	0.008041	2.081738	–0.0236324	0.2120103
Sensitivity analysis	CeLMoP (category B)	0.0003712	0.030807	6.827069	0.0014804	0.2144175
Sensitivity analysis	CeLCoC (category C)	0.0004542	0.0203098	4.028858	–0.0070475	0.1657326
Sensitivity analysis	HyMoP (category D)	0.000317	0.0145044	4.595129	–0.019714	0.2461955
Sensitivity analysis	CeCoP (category E)	0.0001675	0.0034053	0.680878	–0.0149548	0.1011695
		Log-normal mu	**Log-normal sigma**			
THR vs. THR – log-normal model						
Sensitivity analysis	CeMoP (category A)	9.738756	3.716562			
Sensitivity analysis	CeLMoP (category B)	10.71464	4.573634			
Sensitivity analysis	CeLCoC (category C)	9.526446	4.034555			
Sensitivity analysis	HyMoP (category D)	10.66382	4.215337			
Sensitivity analysis	CeCoP (category E)	9.574467	3.481879			
		Log-normal mu	**Log-normal sigma**	**Log-normal age coefficient**	**Log-normal sex coefficient**	
THR vs. THR – log-normal model controlled for age and sex						
Sensitivity analysis	CeMoP (category A)	9.825973	3.730391	0.03258	–0.3417841	
Sensitivity analysis	CeLMoP (category B)	10.84608	4.563342	–0.0077298	–0.3729022	
Sensitivity analysis	CeLCoC (category C)	9.747396	4.036228	0.0093327	–0.2627816	
Sensitivity analysis	HyMoP (category D)	10.85018	4.238437	0.0314349	–0.3886501	
Sensitivity analysis	CeCoP (category E)	9.729236	3.482196	0.01658	–0.1431533	

Sensitivity analyses for cost inputs

For these sensitivity analyses we varied the prosthesis cost using the highest and lowest cost estimates from the list prices supplied by the NHS Supply Chain:

i. RS vs. THR comparison:

(a) highest list price for both RS and THR prostheses
(b) lowest list price for both RS and THR prostheses (*Table 90*).

ii. THR vs. THR comparison:

(a) highest list price for all THR prostheses
(b) lowest list price for all THR prostheses (*Table 91*).

iii. We assumed a 20% price de-escalator to reflect what NHS trusts would pay for implants in reality (this is usually at a discounted rate based on the volume of purchase):

(a) RS vs. THR comparison: the impact of this assumption was not tested
(b) THR vs. THR comparison: a 20% reduction in the cost of each category of prosthesis (see *Table 91*).

TABLE 90 Prosthesis costs for sensitivity analysis for the comparison between RS and THR: highest and lowest list prices for both RS and THR (weighted average of all categories)

Prosthesis	Base-case list price (£)	Highest list price (£)	Lowest list price (£)
THR	2571	3073	2180
RS	2778	2994	2487

TABLE 91 Prosthesis costs for sensitivity analysis for the comparison between different types of THR: highest and lowest unit costs for each category of prosthesis and a 20% reduction in prosthesis list price for each category

Prosthesis	Component	Highest average unit cost (£)	Lowest average unit cost (£)	20% reduction in prosthesis list price: average unit cost (£)
Category A – CeMoP	Cemented stem	1789	1241	1246
	Metal head			
	Polyethylene cup – cemented			
	Cemented stem centraliser			
	Bone cement plug			
	Cemented stem and cup extras			
Category B – CeLMoP	Cementless HAC stem	3774	2662	2413
	Metal stem			
	Metal cup – cementless HA			
	Liner – polyethylene			
	Fixation screw			
Category C – CeLCoC	Cementless HAC stem	4734	3507	3095
	Ceramic head			
	Metal cup – cementless HA			
	Liner – ceramic			
	Fixation screw			
Category D – HyMoP	Cemented stem	2980	2219	2120
	Metal head			
	Metal cup – cementless HA			
	Liner – polyethylene			
	Cemented stem centraliser			
	Bone cement plug			
	Fixation screw			
	Cemented stem extras			
E – CeCoP	Cemented stem	2271	1657	1597
	Ceramic head			
	Polyethylene cup – cemented			
	Cemented stem centraliser			
	Bone cement plug			
	Cemented stem and cup extras			

HA, hydroxyapatite; HAC, hydroxyapatite coated.

Sensitivity analyses for utility inputs

In the base case, utility values were obtained from the PROMs data set.[97] For the sensitivity analysis, utility values were taken from Rolfson *et al.*[298] This study reported 1-year postoperative utility values for 32,396 patients from the SHAR using a UK EQ-5D tariff. Utility values from the PROMs data set were applied to re-revision health as in the base case. The impact of this assumption was tested only for the comparison between different types of THR and not for the comparison between RS and THR (*Table 92*).

One-way sensitivity analysis for category E compared with category A total hip replacement (tornado diagram)

One-way sensitivity analysis was conducted to examine the individual impact of the net monetary benefit of category E (CeCoP) compared with category A (CeMoP) THR. All parameters were varied around the base-case values within the plausible ranges as specified.

Scenario analysis around revision rates using values obtained from clinical trials/registries

We did not feel that it would be appropriate to use data from other clinical trials/registries to check our findings because the clinical effectiveness studies of revision rates that we identified were based on low counts and/or on small trials with a great deal of uncertainty. Overall, across the THR/THR and THR/RS comparisons, trials were often based on selective populations or interventions. Data that could be obtained from studies examining revision rates were inconclusive and often the results had wide CIs.

Results of the cost-effectiveness analysis

We present here the deterministic and probabilistic cost-effectiveness results for the comparison between RS and THR and the comparison between different types of THR.

Base-case results

Resurfacing arthroplasty compared with total hip replacement

In the base-case analysis we compared the cost-effectiveness of different types of primary THR compared with RS for people in whom both procedures are suitable.

TABLE 92 Summary of utility inputs for sensitivity analysis

Age group (years)	Mean value	SE	Beta distribution, parameter alpha	Beta distribution, parameter beta	Source
50–60	0.77	0.0036	10,006	2989	Rolfson *et al.*[298]
60–70	0.80	0.0021	28,270	7067	Rolfson *et al.*[298]
70–80	0.78	0.0021	30,273	8538	Rolfson *et al.*[298]
80+	0.73	0.0035	11,350	4197	Rolfson *et al.*[298]

Table 93 shows the deterministic and probabilistic results for the 10-year and lifetime horizons. For all scenarios the mean cost of RS was higher than that of THR and the mean QALYs were lower. For all scenarios the ICER for RS was dominated by THR, that is, THR was cheaper and more effective than RS.

Figure 59a and *b* shows the cost-effectiveness planes for THR compared with RS for the 10-year and lifetime horizons. The graph clearly shows that THR dominates RS, as the iterations fall in the north-west quadrant of the plane, that is, RS is clearly more costly and less effective than THR. *Figure 59c* and *d* shows the CEACs for the two time horizons. For a WTP threshold from £0 to £50,000 per QALY, THR is the more cost-effective option.

TABLE 93 Base-case deterministic and probabilistic results for all patients using the bathtub model

Analysis	RS	THR
Deterministic *10-year time horizon*		
Total mean cost (£)	22,519	11,879
Total mean QALYs	7.2830	7.4147
Incremental cost (£)	10,641	
Incremental QALYs	–0.1317	
ICER (£)	Dominated	
Lifetime horizon		
Total mean cost (£)	29,603	18,113
Total mean QALYs	14.6968	14.7846
Incremental cost (£)	11,490	
Incremental QALYs	–0.0879	
ICER (£)	Dominated	
Probabilistic *10-year time horizon*		
Total mean cost (£)	22,615	11,887
Total mean QALYs	7.2823	7.4150
Incremental cost (£)	10,729	
Incremental QALYs	–0.1327	
ICER (£)	Dominated	
Lifetime horizon		
Total mean cost (£)	29,770	18,120
Total mean QALYs	14.6963	14.7848
Incremental cost (£)	11,650	
Incremental QALYs	–0.0885	
ICER (£)	Dominated	

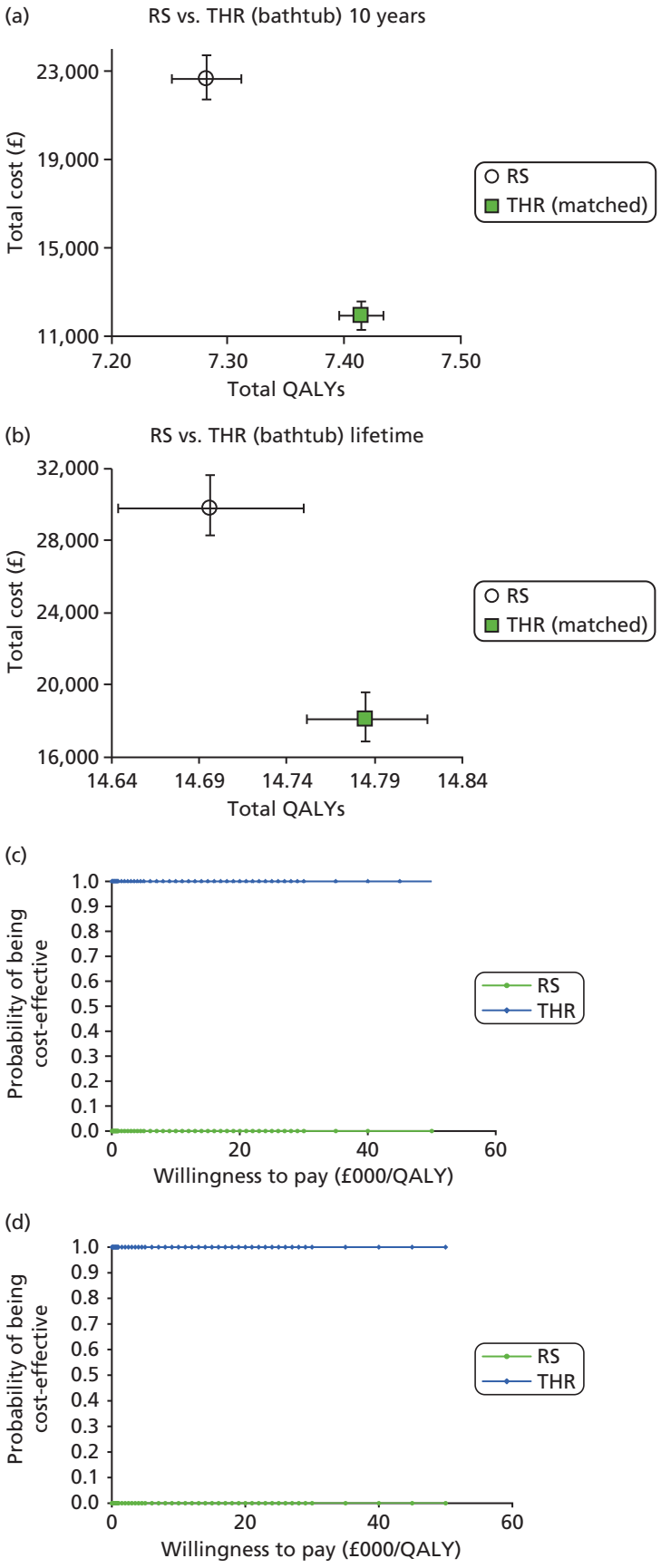


FIGURE 59 Cost-effectiveness planes and CEACs for THR vs. RS. (a) Cost-effectiveness plane for 10-year time horizon; (b) cost-effectiveness plane for lifetime horizon; (c) CEAC for 10-year time horizon; and (d) CEAC for lifetime horizon.

Comparison of different categories of total hip replacement

In the base-case analysis, using a bathtub model, we compared the cost-effectiveness of different categories of primary THR with each other for patients who were not suitable for RS. *Table 94* shows the deterministic and probabilistic results for the 10-year and lifetime horizons; results were ranked by the least costly option. For the 10-year time horizon (both deterministic and probabilistic), category A was cheaper than all of the other categories; however, the QALYs were slightly higher for category E than for the other categories. The ICER for category A compared with category E was £166,217 per QALY gained for the deterministic analysis and £225,225 per QALY gained for the probabilistic analysis. However, when looking at the lifetime scenario (both deterministic and probabilistic), the mean cost for category E was slightly lower and the mean QALYs for category E were slightly higher than the corresponding values for the other categories. Hence, category E dominated the other four categories.

Figure 60a and *b* shows the cost-effectiveness planes with 95% CIs for the comparison between different types of THR. For the 10-year time horizon, although category A is cheaper, category E generates more QALYs. For the lifetime horizon, category E is more cost-effective (i.e. cheaper and more effective) than the other four categories. *Figure 60c* and *d* shows the CEACs for the comparison between different types of THR using a bathtub model. For the 10-year time horizon, if the decision-maker is willing to pay £20,000 per QALY, category A is 95% more cost-effective than the other four categories. For the lifetime horizon, if a decision-maker is willing to pay anything from £0 to £50,000 per QALY, category E is > 90% cost-effective.

TABLE 94 Base-case deterministic and probabilistic results for all THR patients using the bathtub model

Analysis	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic *10-year time horizon*						
A	9444	7.4189	–	–	–	–
E	9743	7.4207	E vs. A	299	0.0018	166,217
D	10,588	7.4182	D vs. E	845	–0.0025	Dominated
B	11,155	7.4156	B vs. D	567	–0.0026	Dominated
C	12,112	7.4143	C vs. B	957	–0.0013	Dominated
Lifetime horizon						
E	14,522	14.7909	–	–	–	–
A	14,801	14.7887	A vs. E	278	–0.0022	Dominated
D	16,040	14.7881	D vs. A	1240	–0.0006	Dominated
B	16,804	14.7861	B vs. D	764	–0.0020	Dominated
C	18,226	14.7845	C vs. B	1422	–0.0016	Dominated
Probabilistic *10-year time horizon*						
A	9449	7.4199	–	–	–	–
E	9775	7.4213	E vs. A	326	0.0014	225,225
D	10,594	7.4192	D vs. E	820	–0.0021	Dominated
B	11,160	7.4165	B vs. D	566	–0.0026	Dominated
C	12,121	7.4152	C vs. B	961	–0.0014	Dominated
Lifetime horizon						
E	14,456	14.7914	–	–	–	–
A	14,740	14.7892	A vs. E	284	–0.0022	Dominated
D	15,975	14.7885	D vs. A	1234	–0.0006	Dominated
B	16,730	14.7866	B vs. D	755	–0.0019	Dominated
C	18,163	14.7850	C vs. B	1432	–0.0016	Dominated

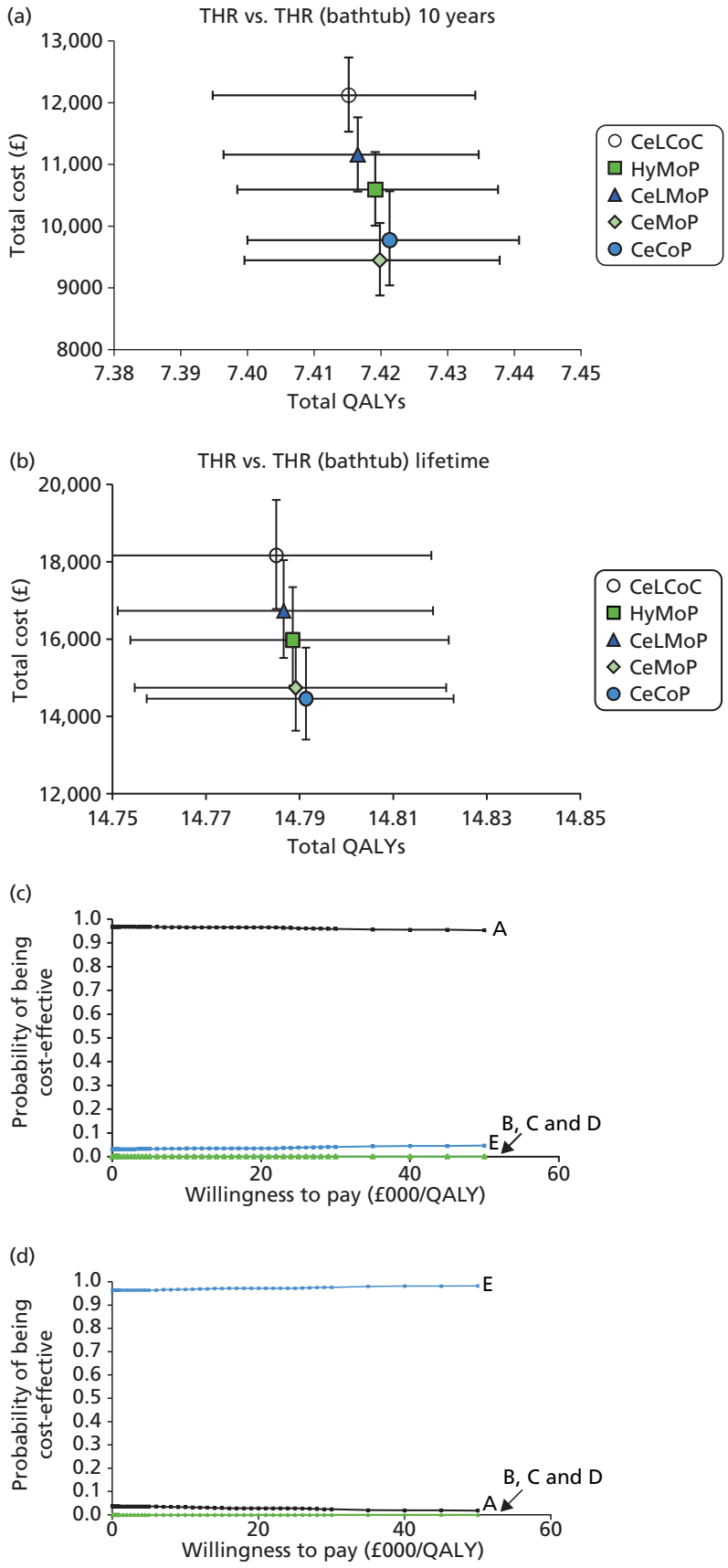


FIGURE 60 Cost-effectiveness planes and CEACs for the comparison between different types of THR. (a) Cost-effectiveness plane for 10-year time horizon; (b) cost-effectiveness plane for lifetime horizon; (c) CEAC for 10-year time horizon; and (d) CEAC for lifetime horizon.

Sensitivity analysis results

This section presents the results from the deterministic and probabilistic sensitivity analyses.

Subgroup analysis: resurfacing arthroplasty compared with total hip replacement

Tables 95 and *96* shows the deterministic and probabilistic results, respectively, for RS compared with THR, presented separately for men and women by age group (40, 50 and 60 years). The incremental cost difference and the incremental QALY difference between THR and RS were higher for women than for men for all age groups. Following the base-case results, RS is clearly dominated by THR, that is, THR is cheaper and more effective than RS.

The results from *Tables 95* and *96* are reflected in the cost-effectiveness planes and CEACs (*Figures 61* and *62*).

TABLE 95 Deterministic results for 40-, 50- and 60-year-old male and female patients

	Age 40 years		Age 50 years		Age 60 years	
Analysis	RS	THR	RS	THR	RS	THR
Women *10-year time horizon*						
Total mean cost (£)	23,230	11,877	23,142	11,665	22,967	11,427
Total mean QALYs	7.0604	7.1891	7.1940	7.3373	7.2501	7.4072
Incremental cost (£)	11,353		11,476		11,541	
Incremental QALYs	–0.1287		–0.1432		–0.1571	
ICER (£)	Dominated		Dominated		Dominated	
Lifetime horizon						
Total mean cost (£)	33,272	21,637	31,248	18,790	28,677	15,904
Total mean QALYs	16.7060	16.8272	14.9977	15.1024	12.6013	12.6785
Incremental cost (£)	11,635		12,458		12,773	
Incremental QALYs	–0.1212		–0.1047		–0.0772	
ICER (£)	Dominated		Dominated		Dominated	
Men *10-year time horizon*						
Total mean cost (£)	22,100	12,022	22,019	11,671	21,820	11,307
Total mean QALYs	7.2311	7.3407	7.4061	7.5345	7.3816	7.5205
Incremental cost (£)	10,078		10,348		10,513	
Incremental QALYs	–0.1096		–0.1284		–0.1389	
ICER (£)	Dominated		Dominated		Dominated	
Lifetime horizon						
Total mean cost (£)	30,805	21,523	28,798	18,126	26,313	15,003
Total mean QALYs	16.5899	16.6779	14.7441	14.8238	12.1711	12.2304
Incremental cost (£)	9283		10,672		11,310	
Incremental QALYs	–0.0879		–0.0797		–0.0593	
ICER (£)	Dominated		Dominated		Dominated	

TABLE 96 Probabilistic results for 40-, 50- and 60-year-old male and female patients

	Age 40 years		Age 50 years		Age 60 years	
Analysis	RS	THR	RS	THR	RS	THR
Women *10-year time horizon*						
Total mean cost (£)	23,233	11,883	23,125	11,672	22,962	11,414
Total mean QALYs	7.0599	7.1886	7.1937	7.3370	7.2495	7.4069
Incremental cost (£)	11,349		11,453		11,549	
Incremental QALYs	–0.1287		–0.1433		–0.1574	
ICER (£)	Dominated		Dominated		Dominated	
Lifetime horizon						
Total mean cost (£)	33,291	21,720	31,247	18,802	28,669	15,883
Total mean QALYs	16.7033	16.8251	14.9976	15.1024	12.6010	12.6783
Incremental cost (£)	11,570		12,445		12,785	
Incremental QALYs	–0.1218		–0.1047		–0.0773	
ICER (£)	Dominated		Dominated		Dominated	
Men *10-year time horizon*						
Total mean cost (£)	22,106	12,027	22,015	11,659	21,828	11,307
Total mean QALYs	7.2313	7.3408	7.4061	7.5334	7.3814	7.5204
Incremental cost (£)	10,080		10,357		10,521	
Incremental QALYs	–0.1095		–0.1284		–0.1389	
ICER (£)	Dominated		Dominated		Dominated	
Lifetime horizon						
Total mean cost (£)	30,765	21,533	28,778	18,143	26,314	15,022
Total mean QALYs	16.5895	16.6775	14.7433	14.8232	12.1706	12.2301
Incremental cost (£)	9231		10,635		11,292	
Incremental QALYs	–0.0880		–0.0799		–0.0595	
ICER (£)	Dominated		Dominated		Dominated	

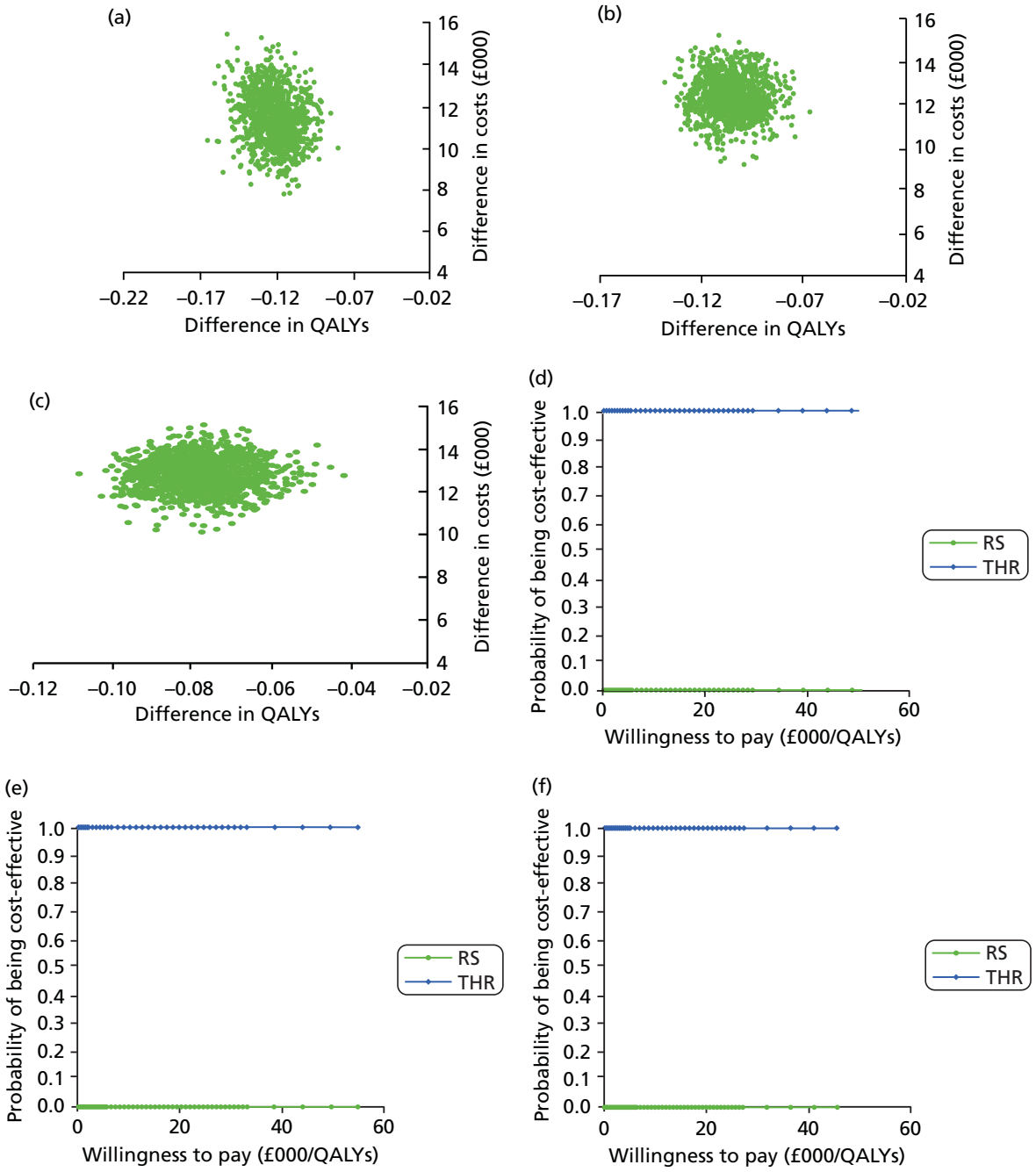


FIGURE 61 Cost-effectiveness planes and CEACs for the comparison between THR and RS for female patients by age group (lifetime horizon). (a) Cost-effectiveness plane – 40 years; (b) cost-effectiveness plane – 50 years; (c) cost-effectiveness plane – 60 years; (d) CEAC – 40 years; (e) CEAC – 50 years; and (f) CEAC – 60 years.

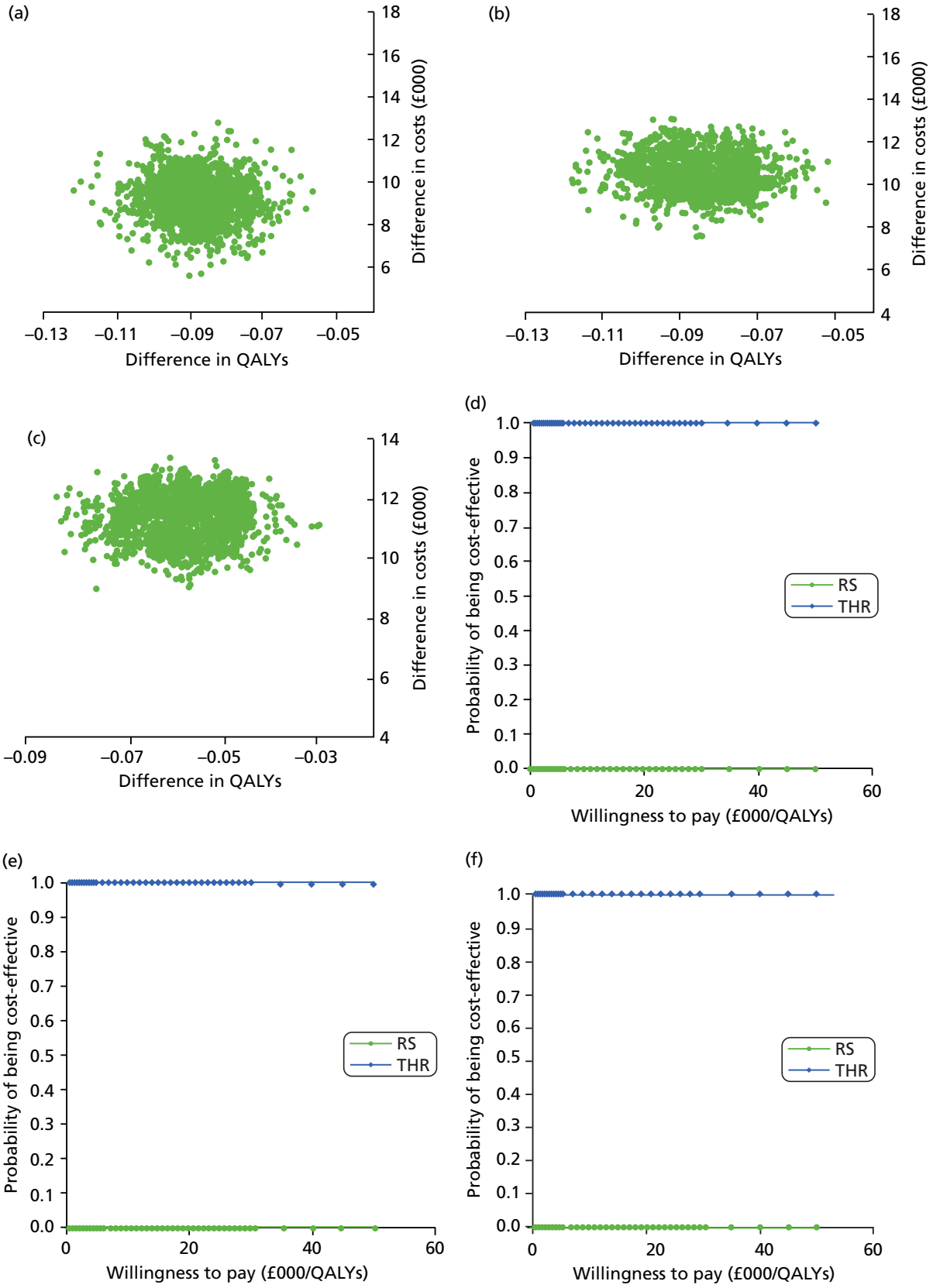


FIGURE 62 Cost-effectiveness planes and CEACs for the comparison between THR and RS for male patients by age group (lifetime horizon). (a) Cost-effectiveness plane – 40 years; (b) cost-effectiveness plane – 50 years; (c) cost-effectiveness plane – 60 years; (d) CEAC – 40 years; (e) CEAC – 50 years; and (f) CEAC – 60 years.

Subgroup analyses: comparison of different types of total hip replacement (patients aged > 65 years)

The deterministic and probabilistic results for the different THR categories over a 10-year time horizon, split by age and sex, are shown in *Tables 97* and *98*, respectively, along with the corresponding ICERs (when appropriate). For both men and women aged both 70 years and 80 years, although category A was cheaper, category E was more effective.

TABLE 97 Deterministic results for men and women aged > 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 70 years *Women*						
A	9047	6.8159	–	–	–	–
E	9364	6.8173	E vs. A	317	0.0014	231,970
D	10,134	6.8160	D vs. E	770	–0.0013	Dominated
B	10,586	6.8150	B vs. D	452	–0.0010	Dominated
C	11,427	6.8151	C vs. B	841	0.0001	5,773,991
A	9047	6.8159	–	–	–	–
E	9364	6.8173	E vs. A	317	0.0014	231,970
C	11,427	6.8151	C vs. E	2,063	–0.0022	Dominated
Men						
A	8900	6.8903	–	–	–	–
E	9238	6.8915	E vs. A	338	0.0012	281,096
D	10,028	6.8898	D vs. E	790	–0.0016	Dominated
B	10,506	6.8885	B vs. D	478	–0.0013	Dominated
C	11,451	6.8874	C vs. B	944	–0.0011	Dominated
Age 80 years *Women*						
A	8175	5.1980	–	–	–	–
E	8495	5.1984	E vs. A	320	0.0004	803,012
D	9263	5.1981	D vs. E	768	–0.0003	Dominated
B	9829	5.1975	B vs. D	566	–0.0006	Dominated
C	10,681	5.1975	C vs. B	851	–0.0000	Dominated
Men						
A	8035	5.0689	–	–	–	–
E	8464	5.0690	E vs. A	429	0.0000	12,763,540
D	9138	5.0689	D vs. E	673	–0.0001	Dominated
B	9752	5.0679	B vs. D	615	–0.0010	Dominated
C	10,695	5.0675	C vs. B	942	–0.0004	Dominated

TABLE 98 Probabilistic results for men and women aged > 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 70 years *Women*						
A	9046	6.8161	–	–	–	–
E	9362	6.8174	E vs. A	316	0.0014	229,667
D	10,139	6.8160	D vs. E	777	–0.0014	Dominated
B	10,591	6.8151	B vs. D	452	–0.0010	Dominated
C	11,425	6.8153	C vs. B	834	0.0002	3,786,953
A	9046	6.8161	–	–	–	–
E	9362	6.8174	E vs. A	316	0.0014	229,667
C	11,425	6.8153	C vs. E	2,063	–0.0022	Dominated
Men						
A	8891	6.8905	–	–	–	–
E	9268	6.8912	E vs. A	377	0.0007	512,560
D	10,027	6.8900	D vs. E	759	–0.0013	Dominated
B	10,503	6.8886	B vs. D	476	–0.0013	Dominated
C	11,508	6.8868	C vs. B	1,005	–0.0018	Dominated
Age 80 years *Women*						
A	8170	5.1985	–	–	–	–
E	8490	5.1989	E vs. A	320	0.0004	804,850
D	9260	5.1985	D vs. E	770	–0.0003	Dominated
B	9828	5.1979	B vs. D	568	–0.0006	Dominated
C	10,675	5.1979	C vs. B	846	0.0000	1,573,299,053
A	8170	5.1985	–	–	–	–
E	8490	5.1989	E vs. A	320	0.0004	804,850
C	10,675	5.1979	C vs. E	2184	–0.0009	Dominated
Men						
A	8029	5.0687	–	–	–	–
E	8501	5.0686	E vs. A	472	–0.0002	Dominated
D	9140	5.0687	D vs. E	639	0.0001	8,491,620
B	9753	5.0676	B vs. D	614	–0.0010	Dominated
C	10,768	5.0669	C vs. B	1015	–0.0007	Dominated
A	8029	5.0687	–	–	–	–
D	9140	5.0687	D vs. A	1110	–0.0001	Dominated

Tables 99 and *100* show the deterministic and probabilistic results, respectively, for men and women aged > 65 years (70 and 80 years) for the lifetime horizon, along with the corresponding ICERs (when appropriate). For both men and women aged 70 years, although category A was cheaper, category E was more effective. For women aged 80 years, category A was cheaper and category D generated more QALYs; for men aged 80 years, category A was cheaper but category E generated more QALYs for the deterministic analysis and category D generated more QALYs for the probabilistic analysis. The corresponding CEACs are shown in *Figure 63*.

TABLE 99 Deterministic results for men and women aged > 65 years for a lifetime horizon.

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 70 years *Women*						
A	10,635	9.4317	–	–	–	–
E	10,916	9.4318	E vs. A	281	0.0001	3,208,305
D	11,694	9.4316	D vs. E	778	–0.0001	Dominated
B	12,160	9.4315	B vs. D	466	–0.0001	Dominated
C	13,005	9.4316	C vs. B	845	0.0000	23,645,296
A	10,635	9.4317	–	–	–	–
E	10,916	9.4318	E vs. A	281	0.0001	3,208,305
C	13,005	9.4316	C vs. E	2,090	–0.0002	Dominated
Men						
A	10,111	8.9914	–	–	–	–
E	10,428	8.9916	E vs. A	317	0.0002	1,424,339
D	11,247	8.9913	D vs. E	819	–0.0003	Dominated
B	11,738	8.9911	B vs. D	492	–0.0003	Dominated
C	12,712	8.9909	C vs. B	973	–0.0002	Dominated
Age 80 years *Women*						
A	8688	6.0572	–	–	–	–
E	8993	6.0573	E vs. A	305	0.0002	1,911,863
D	9768	6.0574	D vs. E	774	0.0000	15,988,179
B	10,350	6.0573	B vs. D	583	0.0000	Dominated
C	11,204	6.0573	C vs. B	854	–0.0001	Dominated
Men						
A	8391	5.6873	–	–	–	–
E	8820	5.6873	E vs. A	429	0.0000	118,964,663
D	9494	5.6873	D vs. E	674	0.0000	Dominated
B	10,123	5.6868	B vs. D	629	–0.0005	Dominated
C	11,075	5.6866	C vs. B	952	–0.0003	Dominated

TABLE 100 Probabilistic results for men and women aged > 65 years for a lifetime horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 70 years *Women*						
A	10,636	9.4314	–	–	–	–
E	10,919	9.4315	E vs. A	282	0.0001	3,168,484
D	11,708	9.4314	D vs. E	789	–0.0001	Dominated
B	12,168	9.4313	B vs. D	460	–0.0001	Dominated
C	13,006	9.4313	C vs. B	838	0.0000	20,570,154
A	10,636	9.4314	–	–	–	–
E	10,919	9.4315	E vs. A	282	0.0001	3,168,484
C	13,006	9.4313	C vs. E	2088	–0.0002	Dominated
Men						
A	10,099	8.9914	–	–	–	–
E	10,458	8.9915	E vs. A	359	0.0002	2,342,245
D	11,243	8.9913	D vs. E	786	–0.0002	Dominated
B	11,732	8.9910	B vs. D	489	–0.0003	Dominated
C	12,778	8.9907	C vs. B	1,046	–0.0003	Dominated
Age 80 years *Women*						
A	8690	6.0579	–	–	–	–
E	8995	6.0581	E vs. A	305	0.0002	1,964,904
D	9774	6.0582	D vs. E	779	0.0001	15,297,263
B	10,356	6.0581	B vs. D	582	0.0000	Dominated
C	11,205	6.0580	C vs. B	850	–0.0001	Dominated
Men						
A	8395	5.6873	–	–	–	–
E	8866	5.6872	E vs. A	471	–0.0001	Dominated
D	9508	5.6873	D vs. E	643	0.0001	12,759,024
B	10,133	5.6868	B vs. D	625	–0.0004	Dominated
C	11,164	5.6864	C vs. B	1031	–0.0004	Dominated
A	8395	5.6873	–	–	–	–
D	9508	5.6873	D vs. A	1114	–0.0001	Dominated

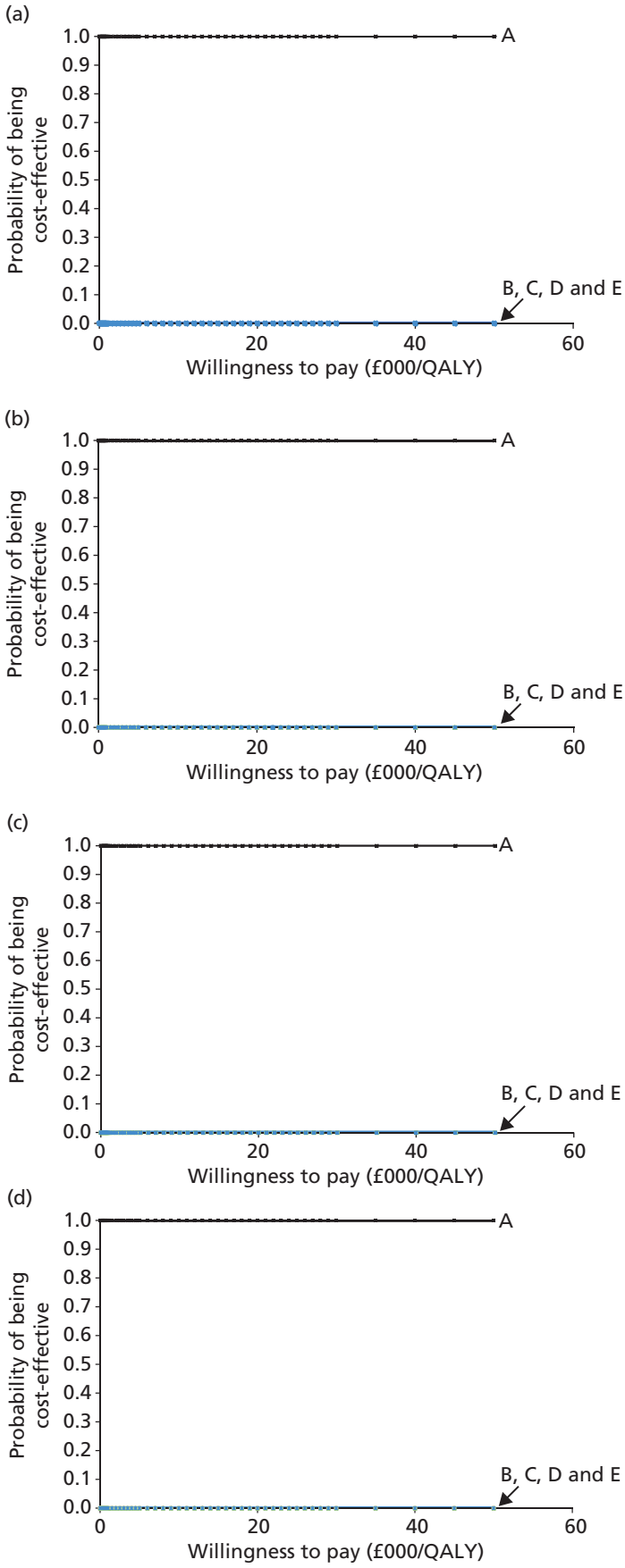


FIGURE 63 Cost-effectiveness acceptability curves for the comparison between different types of THR (patients aged > 65 years). (a) Lifetime horizon, women aged 70 years; (b) lifetime horizon, women aged 80 years; (c) lifetime horizon, men aged 70 years; and (d) lifetime horizon, men aged 80 years.

Subgroup analyses: comparison of different types of total hip replacement (patients aged < 65 years)

Deterministic and probabilistic results over the 10-year time horizon are shown in *Tables 101* and *102*, respectively, for the different THR categories split by age and sex, along with the corresponding ICERs (when appropriate). For men in the age groups 40, 50 and 60 years, although category A was cheaper, category E was more effective.

TABLE 101 Deterministic results for men aged < 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
A	10,097	7.3299	–	–	–	–
E	10,289	7.3330	E vs. A	192	0.0031	62,892
D	11,398	7.3274	D vs. E	1109	–0.0056	Dominated
B	11,742	7.3277	B vs. D	344	0.0004	947,877
C	12,452	7.3294	C vs. B	711	0.0016	434,139
A	10,097	7.3299	–	–	–	–
E	10,289	7.3330	E vs. A	192	0.0031	62,892
B	11,742	7.3277	B vs. E	1452	–0.0052	Dominated
C	12,452	7.3294	C vs. B	710	0.0016	434,139
A	10,097	7.3299	–	–	–	–
E	10,289	7.3330	E vs. A	192	0.0031	62,892
C	12,452	7.3294	C vs. E	2163	–0.0036	Dominated
Age 50 years						
A	9833	7.5230	–	–	–	–
E	9991	7.5270	E vs. A	157	0.0039	40,250
D	11,133	7.5202	D vs. E	1143	–0.0068	Dominated
B	11,647	7.5182	B vs. D	514	–0.0020	Dominated
C	12,274	7.5213	C vs. B	627	0.0030	205,546
A	9833	7.5230	–	–	–	–
E	9991	7.5270	E vs. A	157	0.0039	40,250
C	12,274	7.5213	C vs. E	2283	–0.0057	Dominated
Age 60 years						
A	9529	7.5085	–	–	–	–
E	9685	7.5126	E vs. A	156	0.0042	37,466
D	10,819	7.5056	D vs. E	1134	–0.0071	Dominated
B	11,460	7.5016	B vs. D	642	–0.0040	Dominated
C	12,025	7.5057	C vs. B	565	0.0042	135,491
A	9529	7.5085	–	–	–	–
E	9685	7.5126	E vs. A	156	0.0042	37,466
C	12,025	7.5057	C vs. E	2340	–0.0069	Dominated

TABLE 102 Probabilistic results for men aged < 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
A	10,178	7.3290	–	–	–	–
E	10,390	7.3318	E vs. A	212	0.0028	74,551
D	11,623	7.3247	D vs. E	1233	–0.0071	Dominated
B	11,837	7.3266	B vs. D	214	0.0019	112,217
C	12,474	7.3292	C vs. B	637	0.0025	253,807
A	10,178	7.3290	–	–	–	–
E	10,390	7.3318	E vs. A	212	0.0028	74,551
B	11,837	7.3266	B vs. E	1447	–0.0052	Dominated
C	12,474	7.3292	C vs. B	637	0.0025	253,807
A	10,178	7.3290	–	–	–	–
E	10,390	7.3318	E vs. A	212	0.0028	74,551
C	12,474	7.3292	C vs. E	2084	–0.0027	Dominated
Age 50 years						
A	9835	7.5227	–	–	–	–
E	10,021	7.5262	E vs. A	187	0.0035	52,927
D	11,172	7.5193	D vs. E	1151	–0.0069	Dominated
B	11,662	7.5177	B vs. D	490	–0.0016	Dominated
C	12,284	7.5208	C vs. B	622	0.0031	199,704
A	9835	7.5227	–	–	–	–
E	10,021	7.5262	E vs. A	187	0.0035	52,927
C	12,284	7.5208	C vs. E	2263	–0.0054	Dominated
Age 60 years						
A	9529	7.5091	–	–	–	–
E	9685	7.5132	E vs. A	157	0.0041	37,843
D	10,815	7.5062	D vs. E	1130	–0.0070	Dominated
B	11,465	7.5021	B vs. D	650	–0.0041	Dominated
C	12,028	7.5063	C vs. B	564	0.0042	134,913
A	9529	7.5091	–	–	–	–
E	9685	7.5132	E vs. A	157	0.0041	37,843
C	12,028	7.5063	C vs. E	2343	–0.0069	Dominated

For men, deterministic and probabilistic results over a lifetime horizon are shown in *Tables 103* and *104*, respectively, along with the corresponding ICERs (when appropriate). In the age group 40 years, category A dominated the other four categories; for those aged 50 years and 60 years, category A was cheaper but category E was more effective. *Figure 64* shows the corresponding CEACs.

For women, deterministic and probabilistic results over a 10-year time horizon are shown in *Tables 105* and *106*, respectively, along with the corresponding ICERs (when appropriate). For women aged 40 and 50 years, category E dominated the other four categories; for women aged 60 years, although category A was cheaper, category E was more effective.

For women, deterministic and probabilistic results over a lifetime horizon are shown in *Tables 107* and *108*, respectively, along with the corresponding ICERs (when appropriate). In the age groups 40, 50 and 60 years, category E dominated the other four categories, that is, category E was cheaper and more effective. The corresponding CEACs are shown in *Figure 64*.

TABLE 103 Deterministic results for men aged < 65 years for a lifetime horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
A	18,350	16.6684	–	–	–	–
E	19,351	16.6677	E vs. A	1000	–0.0008	Dominated
D	20,572	16.6625	D vs. E	1222	–0.0052	Dominated
C	21,270	16.6656	C vs. D	697	0.0032	219,152
B	21,712	16.6593	B vs. C	442	–0.0063	Dominated
A	18,350	16.6684	–	–	–	–
C	21,270	16.6656	C vs. A	2919	–0.0028	Dominated
Age 50 years						
A	15,579	14.8116	–	–	–	–
E	15,998	14.8132	E vs. A	419	0.0016	257,281
D	17,560	14.8081	D vs. E	1561	–0.0052	Dominated
C	18,579	14.8090	C vs. D	1020	0.0010	1,059,918
B	19,016	14.8047	B vs. C	437	–0.0044	Dominated
A	15,579	14.8116	–	–	–	–
E	15,998	14.8132	E vs. A	419	0.0016	257,281
C	18,579	14.8090	C vs. E	2581	–0.0042	Dominated
Age 60 years						
A	12,929	12.2177	–	–	–	–
E	13,082	12.2192	E vs. A	153	0.0014	105,773
D	14,606	12.2158	D vs. E	1524	–0.0034	Dominated
C	15,819	12.2158	C vs. D	1213	0.0001	14,646,830
B	16,011	12.2132	B vs. C	192	–0.0026	Dominated
A	12,929	12.2177	–	–	–	–
E	13,082	12.2192	E vs. A	153	0.0014	105,773
C	15,819	12.2158	C vs. E	2737	–0.0033	Dominated

TABLE 104 Probabilistic results for men aged < 65 years for a lifetime horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
A	18,556	16.6662	–	–	–	–
E	19,587	16.6651	E vs. A	1031	–0.0011	Dominated
D	21,069	16.6577	D vs. E	1481	–0.0073	Dominated
C	21,304	16.6646	C vs. D	235	0.0068	34,383
B	21,877	16.6570	B vs. C	573	–0.0076	Dominated
A	18,556	16.6662	–	–	–	–
C	21,304	16.6646	C vs. A	2748	–0.0016	Dominated
Age 50 years						
A	15,626	14.8108	–	–	–	–
E	16,071	14.8124	E vs. A	444	0.0016	279,122
D	17,608	14.8074	D vs. E	1538	–0.0051	Dominated
C	18,581	14.8085	C vs. D	973	0.0012	843,588
B	19,032	14.8041	B vs. C	451	–0.0044	Dominated
A	15,626	14.8108	–	–	–	–
E	16,071	14.8124	E vs. A	444	0.0016	279,122
C	18,581	14.8085	C vs. E	2511	–0.0039	Dominated
Age 60 years						
A	12,957	12.2174	–	–	–	–
E	13,113	12.2188	E vs. A	156	0.0014	109,045
D	14,617	12.2155	D vs. E	1503	–0.0033	Dominated
C	15,831	12.2155	C vs. D	1215	0.0001	15,339,725
B	16,029	12.2128	B vs. C	198	–0.0027	Dominated
A	12,957	12.2174	–	–	–	–
E	13,113	12.2188	E vs. A	156	0.0014	109,045
C	15,831	12.2155	C vs. E	2718	–0.0033	Dominated

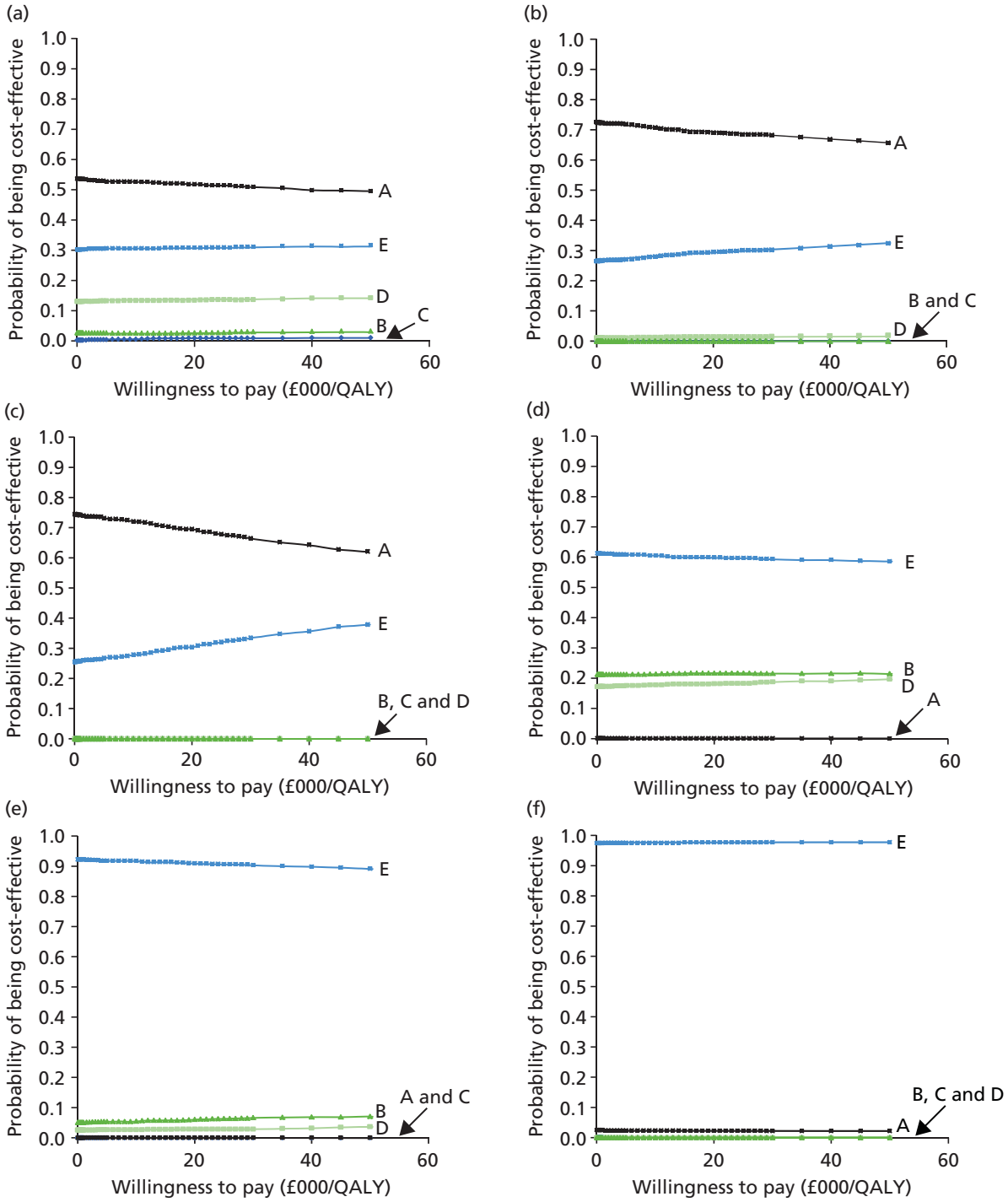


FIGURE 64 Cost-effectiveness acceptability curves for the comparison between different types of THR for women and men by age group (lifetime horizon). (a) Men aged 40 years; (b) men aged 50 years; (c) men aged 60 years; (d) women aged 40 years; (e) women aged 50 years; and (f) women aged 60 years.

TABLE 105 Deterministic results for women aged < 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
E	10,064	7.1950	–	–	–	–
A	10,437	7.1856	A vs. E	374	–0.0094	Dominated
D	10,805	7.1940	D vs. A	368	0.0084	43,732
B	11,540	7.1897	B vs. D	735	–0.0043	Dominated
C	12,381	7.1898	C vs. B	841	0.0001	7,772,228
E	10,064	7.1950	–	–	–	–
D	10,805	7.1940	D vs. E	742	–0.0010	Dominated
C	12,381	7.1898	C vs. D	1575	–0.0042	Dominated
Age 50 years						
E	9978	7.3423	–	–	–	–
A	10,035	7.3359	A vs. E	57	–0.0064	Dominated
D	10,802	7.3401	D vs. A	766	0.0042	181,105
B	11,355	7.3376	B vs. D	553	–0.0025	Dominated
C	12,251	7.3371	C vs. B	896	–0.0006	Dominated
E	9978	7.3423	–	–	–	–
D	10,802	7.3401	D vs. E	823	–0.0022	Dominated
Age 60 years						
A	9670	7.4075	–	–	–	–
E	9846	7.4112	E vs. A	176	0.0037	48,110
D	10,743	7.4078	D vs. E	897	–0.0033	Dominated
B	11,137	7.4073	B vs. D	394	–0.0005	Dominated
C	12,074	7.4061	C vs. B	937	–0.0012	Dominated

TABLE 106 Probabilistic results for women aged < 65 years for a 10-year time horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
E	9983	7.1954	–	–	–	–
A	10,502	7.1843	A vs. E	520	–0.0011	Dominated
D	10,967	7.1916	D vs. A	464	0.0073	63,507
B	11,630	7.1881	B vs. D	663	–0.0035	Dominated
C	12,405	7.1890	C vs. B	775	0.0009	889,457
E	9983	7.1954	–	–	–	–
D	10,967	7.1916	D vs. E	984	–0.0038	Dominated
C	12,405	7.1890	C vs. D	1438	–0.0027	Dominated
Age 50 years						
E	9936	7.3426	–	–	–	–
A	10,049	7.3355	A vs. E	113	–0.0071	Dominated
D	10,849	7.3393	D vs. A	800	0.0038	209,865
B	11,384	7.3371	B vs. D	535	–0.0022	Dominated
C	12,253	7.3368	C vs. B	869	–0.0002	Dominated
E	9936	7.3426	–	–	–	–
D	10,849	7.3393	D vs. E	913	–0.0033	Dominated
Age 60 years						
A	9673	7.4075	–	–	–	–
E	9849	7.4111	E vs. A	176	0.0037	48,113
D	10,749	7.4077	D vs. E	900	–0.0034	Dominated
B	11,147	7.4072	B vs. D	398	–0.0006	Dominated
C	12,075	7.4061	C vs. B	928	–0.0010	Dominated

TABLE 107 Deterministic results for women aged < 65 years for a lifetime horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
E	18,647	16.8374	–	–	–	–
B	18,814	16.8361	B vs. E	167	–0.0013	Dominated
D	20,033	16.8340	D vs. B	1218	–0.0020	Dominated
A	21,595	16.8180	A vs. D	1562	–0.0160	Dominated
C	21,886	16.8289	C vs. A	291	0.0109	26,657
E	18,647	16.8374	–	–	–	–
C	21,886	16.8289	C vs. E	3238	–0.0085	Dominated
Age 50 years						
E	16,426	15.1069	–	–	–	–
B	16,923	15.1053	B vs. E	497	–0.0016	Dominated
A	17,854	15.1003	A vs. B	931	–0.0050	Dominated
D	18,024	15.1042	D vs. A	170	0.0039	43,755
C	19,366	15.1022	C vs. D	1342	–0.0020	Dominated
E	16,426	15.1069	–	–	–	–
D	18,024	15.1042	D vs. E	1598	–0.0027	Dominated
Age 60 years						
E	14,026	12.6801	–	–	–	–
A	14,343	12.6785	A vs. E	317	–0.0016	Dominated
B	14,844	12.6798	B vs. A	501	0.0013	398,183
D	15,599	12.6787	D vs. B	755	–0.0011	Dominated
C	16,655	12.6779	C vs. D	1056	–0.0008	Dominated
E	14,026	12.6801	–	–	–	–
B	14,844	12.6798	B vs. E	818	–0.0004	Dominated

TABLE 108 Probabilistic results for women aged < 65 years for a lifetime horizon

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Age 40 years						
E	18,179	16.8404	–	–	–	–
B	19,050	16.8351	B vs. E	871	–0.0053	Dominated
D	20,368	16.8317	D vs. B	1318	–0.0034	Dominated
A	21,704	16.8178	A vs. D	1335	–0.0139	Dominated
C	21,959	16.8291	C vs. A	255	0.0113	22,538
E	18,179	16.8404	–	–	–	–
C	21,959	16.8291	C vs. E	3780	–0.0113	Dominated
Age 50 years						
E	16,425	15.1072	–	–	–	–
B	16,980	15.1048	B vs. E	735	–0.0024	Dominated
A	17,875	15.0999	A vs. B	895	–0.0049	Dominated
D	18,135	15.1035	D vs. A	259	0.0036	71,800
C	19,379	15.1018	C vs. D	1245	–0.0017	Dominated
E	16,425	15.1072	–	–	–	–
D	18,135	15.1035	D vs. E	1889	–0.0037	Dominated
Age 60 years						
E	14,031	12.6798	–	–	–	–
A	14,359	12.6781	A vs. E	328	–0.0017	Dominated
B	14,873	12.6793	B vs. A	514	0.0012	414,092
D	15,624	12.6782	D vs. B	751	–0.0011	Dominated
C	16,673	12.6774	C vs. D	1048	–0.0008	Dominated
E	14,031	12.6798	–	–	–	–
B	14,873	12.6793	B vs. E	842	–0.0004	Dominated

Sensitivity analysis: time to revision (bathtub model adjusted for age and sex)

Table 109 shows the deterministic and probabilistic results for all patients using a bathtub model adjusted for age and sex. In line with the base-case analysis, RS was dominated by THR for all time horizons, that is, THR was cheaper and more effective than RS). The corresponding cost-effectiveness planes and CEACs are shown in *Figure 65*.

Table 110 shows the deterministic and probabilistic results for all THR patients using the bathtub model adjusted for age and sex. As in the base-case analysis, for the 10-year time horizon (both deterministic and probabilistic) category A was cheaper than all of the other categories; however, category E conferred slightly more QALYs than the other four categories. The ICER for category A compared with category E was £127,420 per QALY gained for the deterministic analysis and £176,776 per QALY gained for the probabilistic analysis.

When looking at the lifetime scenarios (both deterministic and probabilistic), the mean cost for category E was slightly lower and the mean QALYs for category E were slightly higher than the corresponding values for the other four THR categories. Hence, category E dominated the other four categories. The corresponding cost-effectiveness planes and CEACs are shown in *Figure 66*.

TABLE 109 Deterministic and probabilistic results for all patients using the bathtub model adjusted for age and sex

Analysis	RS	THR
Deterministic *10-year time horizon*		
Total mean cost (£)	22,560	11,899
Total mean QALYs	7.2824	7.4144
Incremental cost (£)	10,661	
Incremental QALYs	–0.1320	
ICERs (£)	Dominated	
Lifetime horizon		
Total mean cost (£)	29,664	18,254
Total mean QALYs	14.6964	14.7843
Incremental cost (£)	11,410	
Incremental QALYs	–0.0879	
ICERs (£)	Dominated	
Probabilistic *10-year time horizon*		
Total mean cost (£)	22,729	11,912
Total mean QALYs	7.2804	7.4141
Incremental cost (£)	10,817	
Incremental QALYs	–0.1337	
ICERs (£)	Dominated	
Lifetime horizon		
Total mean cost (£)	29,836	18,268
Total mean QALYs	14.6958	14.7845
Incremental cost (£)	11,568	
Incremental QALYs	–0.0887	
ICERs (£)	Dominated	

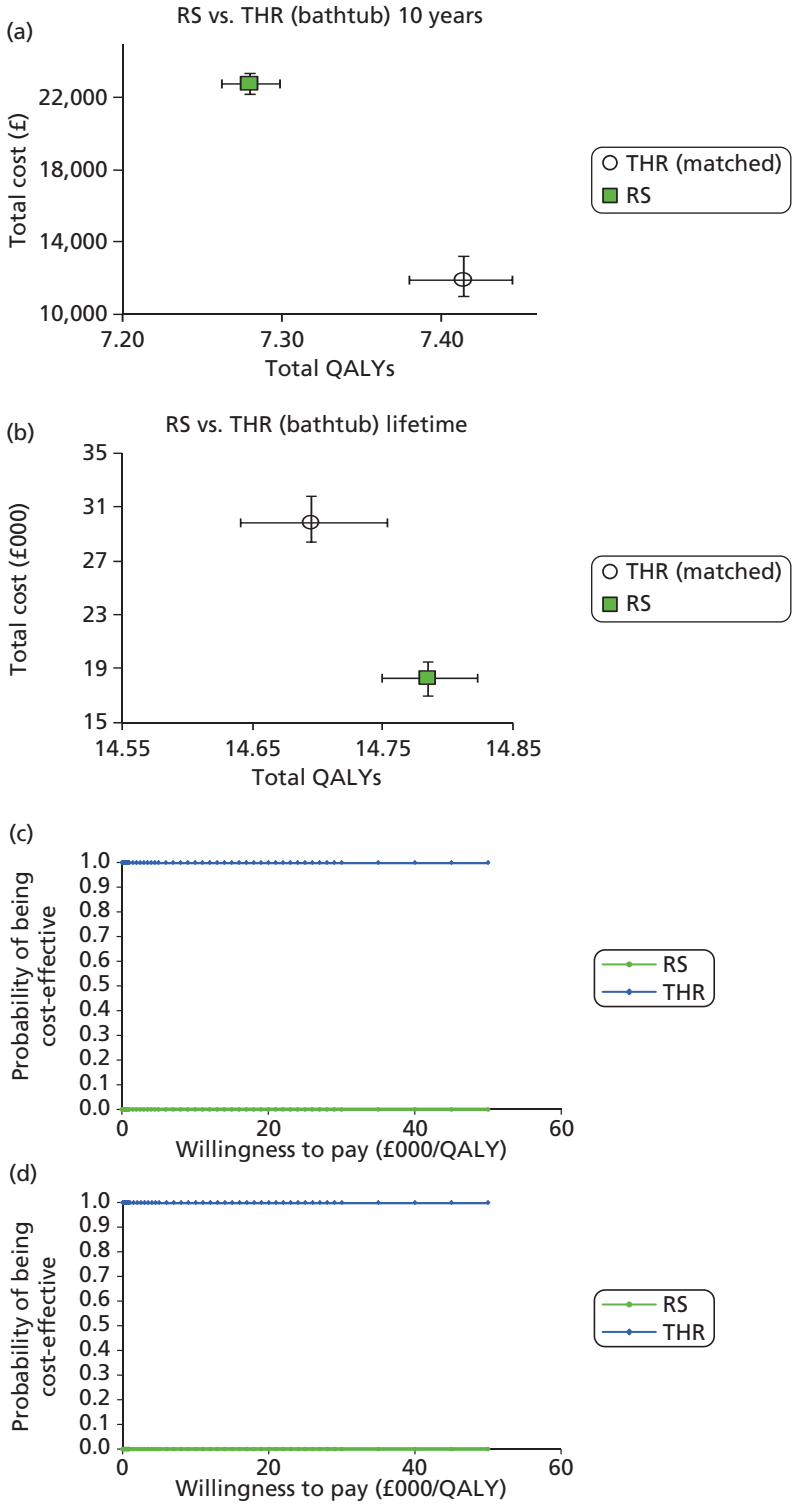


FIGURE 65 Cost-effectiveness planes and CEACs for the comparison between RS and THR using the bathtub model adjusted for age and sex. (a) Cost-effectiveness plane for a 10-year time horizon; (b) cost-effectiveness plane for a lifetime horizon; (c) CEAC for a 10-year time horizon; and (d) CEAC for a lifetime horizon.

TABLE 110 Deterministic and probabilistic results for all THR patients using the bathtub model adjusted for age and sex

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic						
10-year time horizon						
A	9458	7.4187	–	–	–	–
E	9731	7.4208	E vs. A	273	0.0021	127,420
D	10,578	7.4183	D vs. E	846	–0.0025	Dominated
B	11,147	7.4157	B vs. D	569	–0.0027	Dominated
C	12,035	7.4152	C vs. B	888	–0.0004	Dominated
Lifetime horizon						
E	14,533	14.7909	–	–	–	–
A	14,817	14.7886	A vs. E	283	–0.0023	Dominated
D	15,965	14.7883	D vs. A	1148	–0.0003	Dominated
B	16,784	14.7862	B vs. D	819	–0.0021	Dominated
C	17,963	14.7854	C vs. B	1180	–0.0007	Dominated
Probabilistic						
10-year time horizon						
A	9449	7.4190	–	–	–	–
E	9754	7.4207	E vs. A	304	0.0017	176,776
D	10,572	7.4186	D vs. E	818	–0.0021	Dominated
B	11,135	7.4160	B vs. D	563	–0.0026	Dominated
C	12,027	7.4155	C vs. B	891	–0.0005	Dominated
Lifetime horizon						
E	13,954	14.7935	–	–	–	–
A	14,834	14.7881	A vs. E	881	–0.0055	Dominated
D	15,976	14.7878	D vs. A	1142	–0.0003	Dominated
B	16,801	14.7856	B vs. D	825	–0.0021	Dominated
C	17,972	14.7849	C vs. B	1171	–0.0007	Dominated

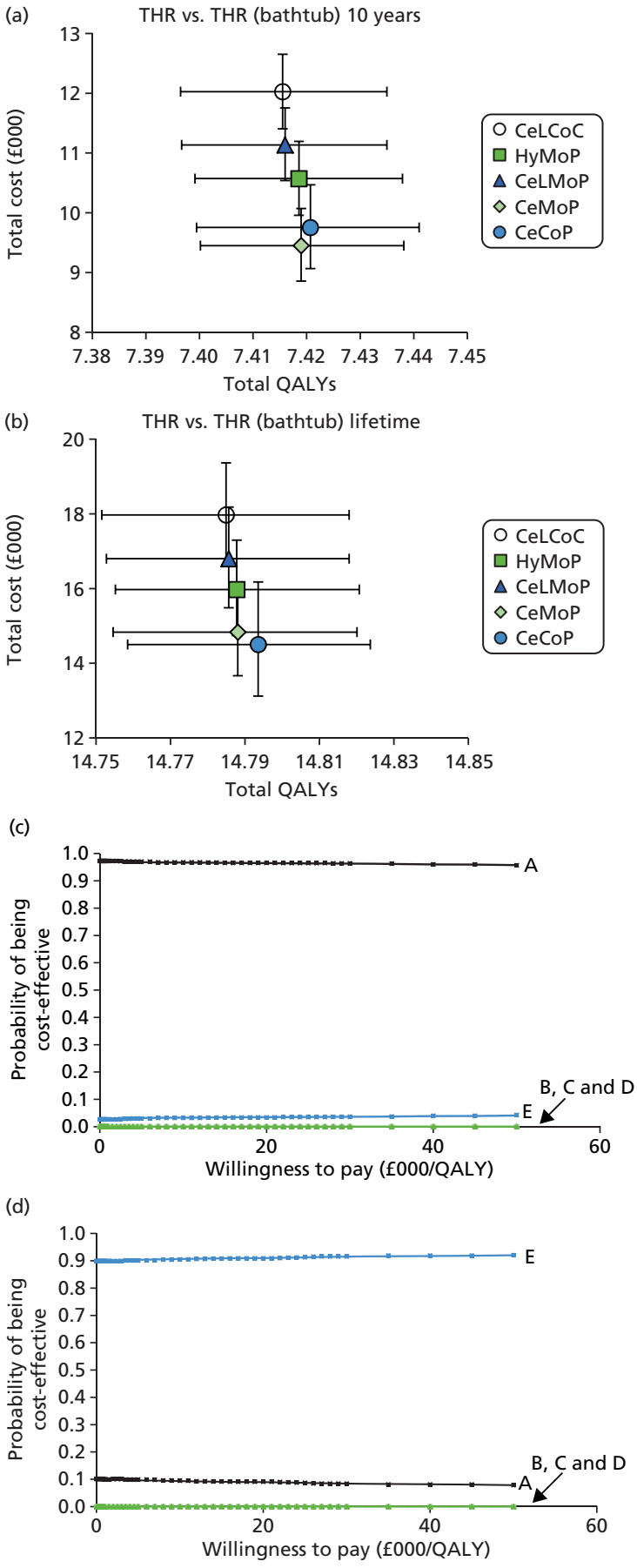


FIGURE 66 Cost-effectiveness planes and CEACs for the comparison between different types of THR using the bathtub model adjusted for age and sex. (a) Cost-effectiveness plane for 10-year time horizon; (b) cost-effectiveness plane for lifetime horizon; (c) CEAC for 10-year time horizon; and (d) CEAC for lifetime horizon.

Sensitivity analysis: time to revision (log-normal model)

For this sensitivity analysis we used a log-normal model of time to revision to compare the cost-effectiveness of the different categories of THR. *Table 111* shows that, for both the deterministic analysis and the probabilistic analysis for both time horizons, category A was cheaper and category E was more effective than the other categories. The corresponding ICERs are also reported in *Table 111*.

Figure 67a and *b* shows the cost-effectiveness planes with the 95% CIs for the comparison between different types of THR. For both the 10-year horizon and the lifetime horizon, although category A is cheaper, category E generates more QALYs. *Figure 67c* and *d* shows the CEACs for the two time horizons using a log-normal model. For both the 10-year time horizon and the lifetime horizon, if a decision-maker is willing to pay anything from £0 to £50,000, category A is nearly 100% cost-effective.

TABLE 111 Deterministic and probabilistic results for all THR patients using the log-normal model

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic						
10-year time horizon						
A	9331	7.4203	–	–	–	–
E	9690	7.4214	E vs. A	359	0.0010	342,781
D	10,446	7.4200	D vs. E	756	–0.0013	Dominated
B	10,986	7.4177	B vs. D	541	–0.0023	Dominated
C	11,901	7.4169	C vs. B	915	–0.0008	Dominated
Lifetime horizon						
A	13,476	14.7919	–	–	–	–
E	13,794	14.7926	E vs. A	318	0.0007	442,830
D	14,568	14.7917	D vs. E	773	–0.0009	Dominated
B	15,192	14.7901	B vs. D	624	–0.0016	Dominated
C	16,190	14.7895	C vs. B	998	–0.0006	Dominated
Probabilistic						
10-year time horizon						
A	9334	7.4200	–	–	–	–
E	9700	7.4210	E vs. A	366	0.0010	384,106
D	10,452	7.4197	D vs. E	752	–0.0013	Dominated
B	10,991	7.4174	B vs. D	539	–0.0023	Dominated
C	11,907	7.4166	C vs. B	916	–0.0008	Dominated
Lifetime horizon						
A	13,464	14.7918	–	–	–	–
E	13,799	14.7924	E vs. A	335	0.0006	522,741
D	14,562	14.7916	D vs. E	762	–0.0008	Dominated
B	15,183	14.7900	B vs. D	621	–0.0016	Dominated
C	16,179	14.7894	C vs. B	997	–0.0006	Dominated

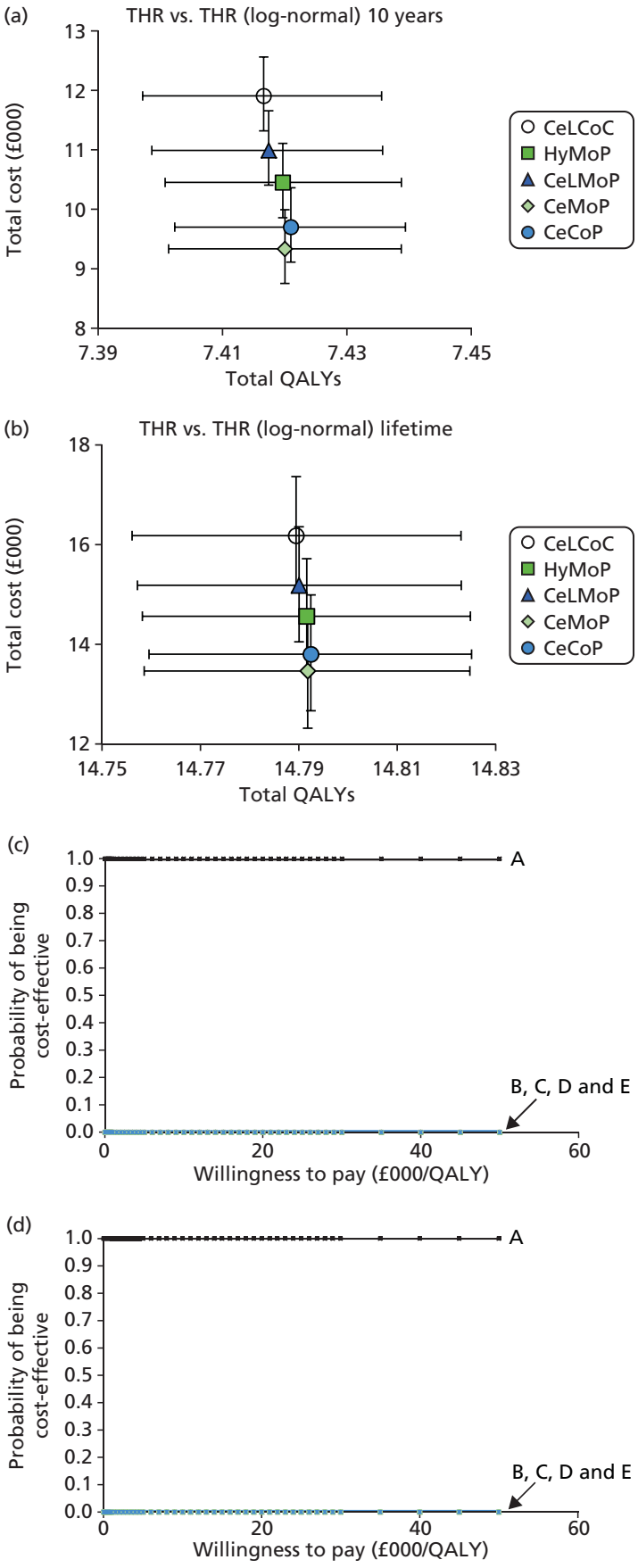


FIGURE 67 Cost-effectiveness planes and CEACs for the comparison between different types of THR using a log-normal model. (a) Cost-effectiveness plane for 10-year time horizon; (b) cost-effectiveness plane for lifetime horizon; (c) CEAC for 10-year time horizon; and (d) CEAC for lifetime horizon.

Sensitivity analysis: time to revision (log-normal model adjusted for age and sex)

For this sensitivity analysis we used a log-normal model for time to revision adjusted for age and sex to compare the cost-effectiveness of the different categories of THR. *Table 112* shows that, for both the deterministic analysis and the probabilistic analysis for both time horizons, category A was cheaper; however, category E was clearly more effective than the other four categories. The corresponding ICERs are also reported in *Table 112*.

The corresponding cost-effectiveness planes and CEACs are shown in *Figure 68*. For both the 10-year time horizon and the lifetime horizon, if the decision-maker is willing to pay £20,000 per QALY, category A is nearly 100% cost-effective.

TABLE 112 Deterministic and probabilistic results for all THR patients using the log-normal model adjusted for age and sex

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic						
10-year time horizon						
A	9349	7.4201	–	–	–	–
E	9667	7.4217	E vs. A	318	0.0016	202,741
D	10,446	7.4200	D vs. E	779	–0.0017	Dominated
B	10,982	7.4178	B vs. D	536	–0.0022	Dominated
C	11,858	7.4175	C vs. B	876	–0.0003	Dominated
Lifetime horizon						
A	13,505	14.7917	–	–	–	–
E	13,753	14.7928	E vs. A	248	0.0011	227,031
D	14,567	14.7917	D vs. E	814	–0.0011	Dominated
B	15,185	14.7902	B vs. D	618	–0.0015	Dominated
C	16,119	14.7899	C vs. B	934	–0.0002	Dominated
Probabilistic						
10-year time horizon						
A	9339	7.4202	–	–	–	–
E	9665	7.4216	E vs. A	327	0.0015	223,741
D	10,438	7.4201	D vs. E	773	–0.0016	Dominated
B	10,973	7.4179	B vs. D	534	–0.0022	Dominated
C	11,849	7.4176	C vs. B	877	–0.0003	Dominated
Lifetime horizon						
A	13,493	14.7912	–	–	–	–
E	13,755	14.7923	E vs. A	263	0.0010	255,638
D	14,559	14.7912	D vs. E	804	–0.0011	Dominated
B	15,175	14.7897	B vs. D	616	–0.0015	Dominated
C	16,112	14.7894	C vs. B	937	–0.0003	Dominated

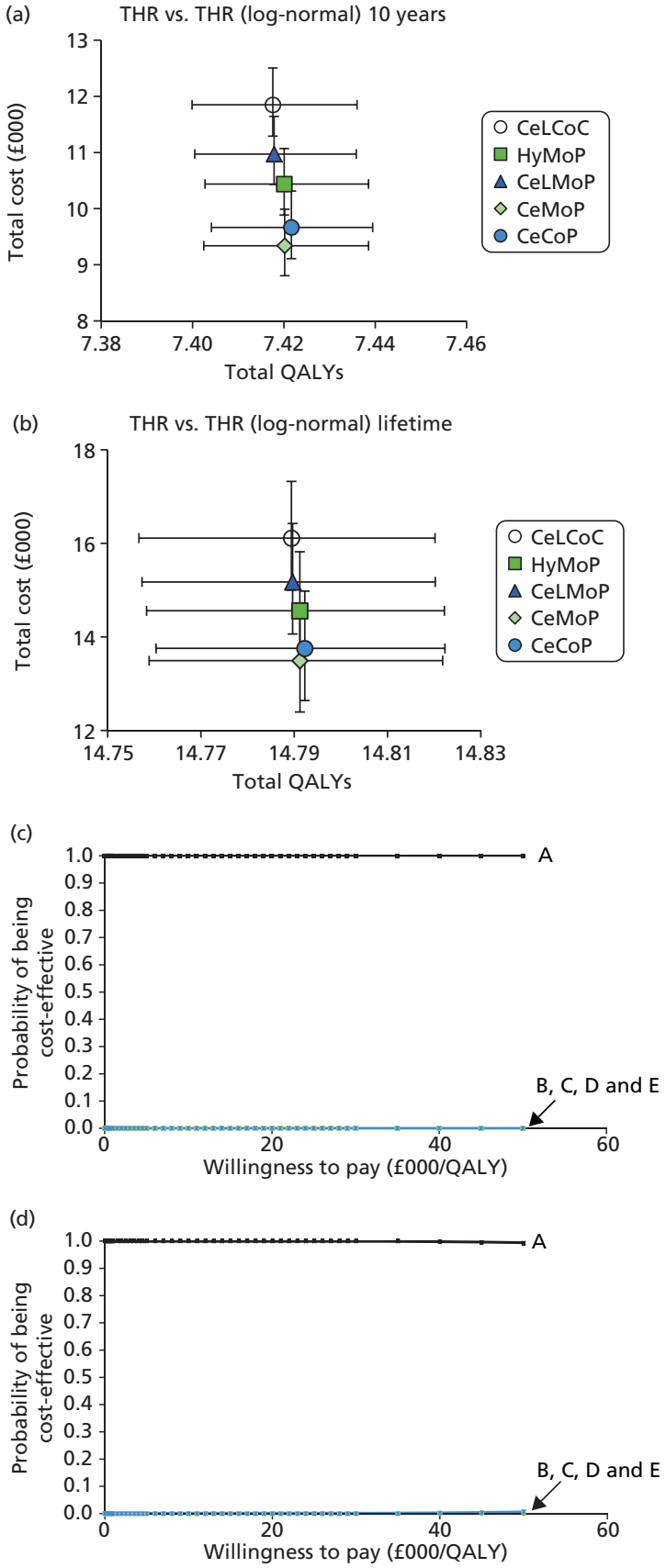


FIGURE 68 Cost-effectiveness planes and CEACs for the comparison between different types of THR using the log-normal model adjusted for age and sex. (a) Cost-effectiveness plane for 10-year time horizon; (b) cost-effectiveness plane for lifetime horizon; (c) CEAC for 10-year time horizon; and (d) CEAC for lifetime horizon.

Sensitivity analysis: costs

For this sensitivity analysis we compared the cost-effectiveness of RS and THR using the highest and lowest cost estimates for prostheses from the list prices supplied by the NHS Supply Chain (*Table 113*). For both time horizons (10 years and lifetime) for both low and high costs, RS was dominated by THR, that is, RS was more expensive and less effective than THR. The corresponding CEACs are shown is *Figure 69*.

TABLE 113 Deterministic and probabilistic results for the lowest and highest cost estimates for prostheses for THR vs. RS patients

	Lowest cost		Highest cost	
Category	RS	THR	RS	THR
Deterministic: 10-year time horizon				
Total mean cost (£)	22,228	11,487	22,735	12,380
Total mean QALYs	7.2830	7.4147	7.2830	7.4147
Incremental cost (£)	10,741		10,355	
Incremental QALYs	–0.1317		–0.1317	
ICER (£)	Dominated		Dominated	
Deterministic: lifetime horizon				
Total mean cost (£)	29,312	17,722	29,819	18,614
Total mean QALYs	14.6968	14.7846	14.6968	14.7846
Incremental cost (£)	11,590	11,205		
Incremental QALYs	–0.0879	–0.0879		
ICER (£)	Dominated		Dominated	
Probabilistic: 10-year time horizon				
Total mean cost (£)	22,318	11,516	22,816	12,392
Total mean QALYs	7.2818	7.4146	7.2811	7.4141
Incremental cost (£)	10,803		10,425	
Incremental QALYs	–0.1328		–0.1330	
ICER (£)	Dominated		Dominated	
Probabilistic: lifetime horizon				
Total mean cost (£)	29,459	17,754	29,991	18,652
Total mean QALYs	14.6976	14.7857	14.6948	14.7839
Incremental cost (£)	11,705		11,339	
Incremental QALYs	–0.0880		–0.0890	
ICER (£)	Dominated		Dominated	

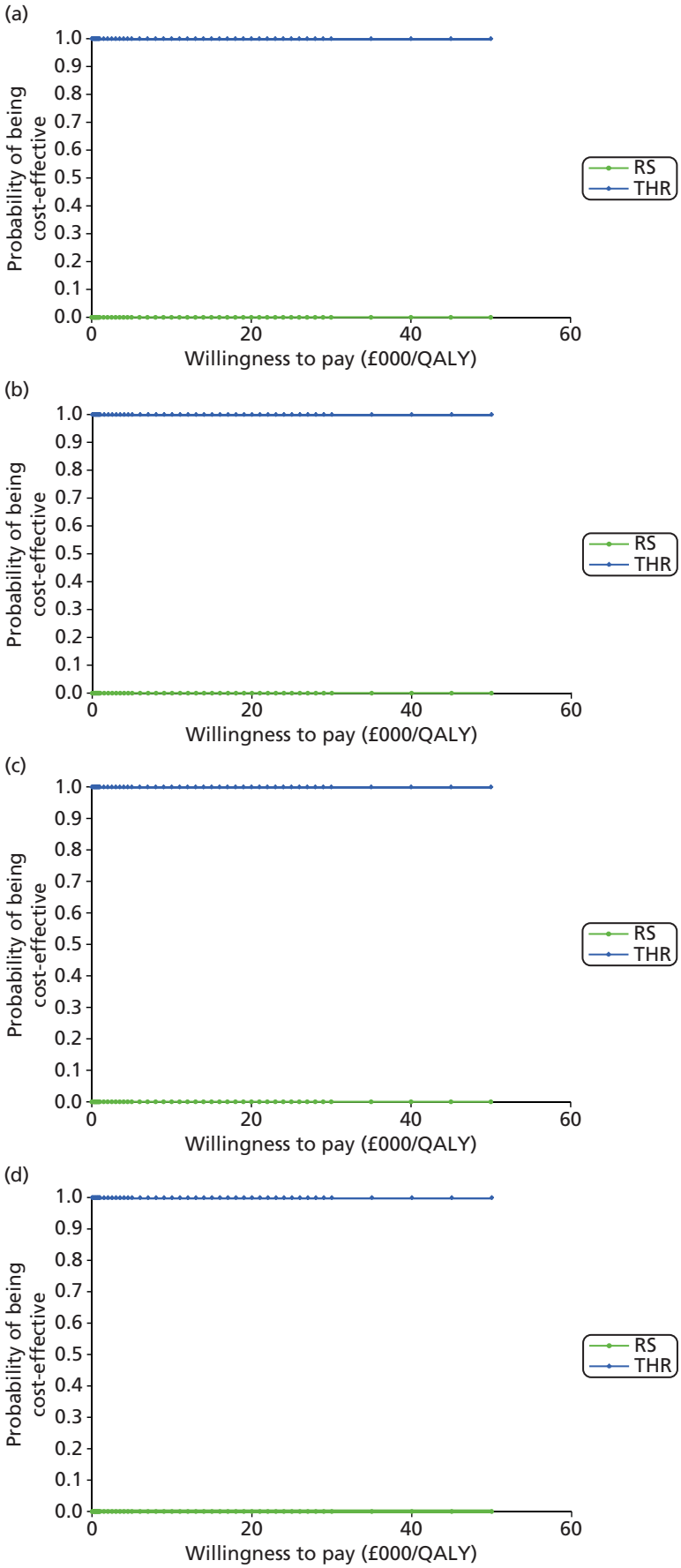


FIGURE 69 Cost-effectiveness acceptability curves for RS vs. THR using the lowest and highest cost estimates for prostheses. (a) 10-year time horizon, lowest cost estimate; (b) lifetime horizon, lowest cost estimate; (c) 10-year time horizon, highest cost estimate; and (d) lifetime horizon, highest cost estimate.

For the comparison between different types of THR, we compared the cost-effectiveness for all THR patients using the highest cost estimates for prostheses from the list prices supplied by the NHS Supply Chain. *Table 114* shows that, for the 10-year time horizon, although category A was cheaper, category E was more effective. The ICER for the deterministic analysis was £190,326 per QALY gained and for the probabilistic analysis was £297,098 per QALY gained. For the lifetime horizon, category E dominated the other four categories. The corresponding CEACs are shown in *Figure 70*.

TABLE 114 Deterministic and probabilistic results for all THR patients using the highest cost estimates for prostheses using the bathtub model

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic						
10-year time horizon						
A	9675	7.4189	–	–	–	–
E	10,018	7.4207	E vs. A	343	0.0018	190,326
D	10,918	7.4182	D vs. E	900	–0.0025	Dominated
B	11,913	7.4156	B vs. D	995	–0.0026	Dominated
C	12,977	7.4143	C vs. B	1064	–0.0013	Dominated
Lifetime horizon						
E	14,798	14.7909	–	–	–	–
A	15,032	14.7887	A vs. E	235	–0.0022	Dominated
D	16,371	14.7881	D vs. A	1338	–0.0006	Dominated
B	17,562	14.7861	B vs. D	1192	–0.0020	Dominated
C	19,091	14.7845	C vs. B	1529	–0.0016	Dominated
Probabilistic						
10-year time horizon						
A	9672	7.4191	–	–	–	–
E	10,055	7.4204	E vs. A	383	0.0013	297,098
D	10,917	7.4184	D vs. E	862	–0.0020	Dominated
B	11,909	7.4158	B vs. D	992	–0.0026	Dominated
C	12,973	7.4145	C vs. B	1063	–0.0013	Dominated
Lifetime horizon						
E	14,814	14.7909	–	–	–	–
A	15,030	14.7889	A vs. E	217	–0.0020	Dominated
D	16,378	14.7883	D vs. A	1347	–0.0007	Dominated
B	17,570	14.7863	B vs. D	1193	–0.0020	Dominated
C	19,076	14.7848	C vs. B	1506	–0.0015	Dominated

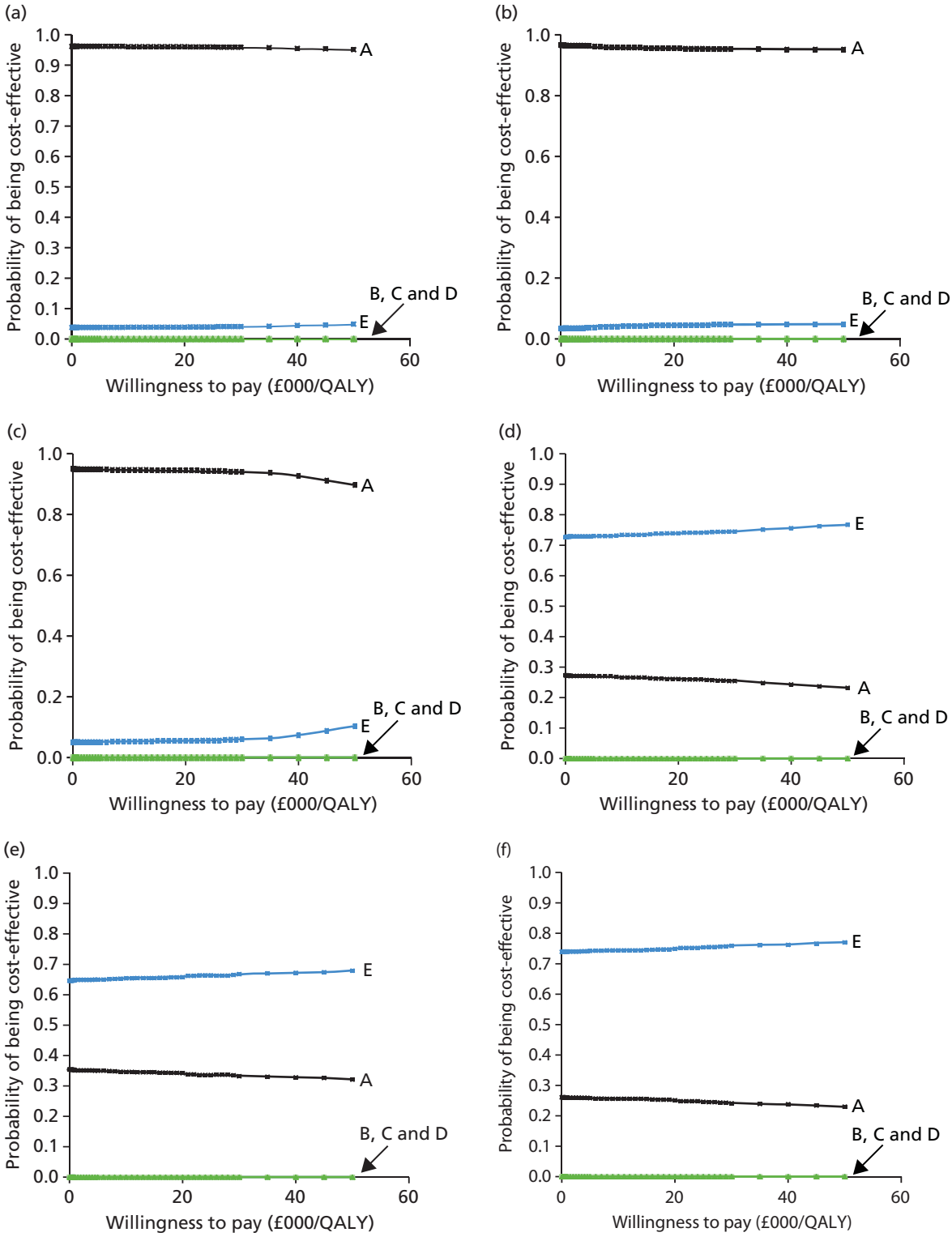


FIGURE 70 Cost-effectiveness acceptability curves for all THR patients using the lowest and highest costs for prostheses and assuming a 20% price de-escalator. (a) 10-year time horizon, lowest cost estimate; (b) 10-year time horizon, highest cost estimate; (c) 10-year time horizon, 20% de-escalator; (d) lifetime horizon, lowest cost estimate; (e) lifetime horizon, highest cost estimate; and (f) lifetime horizon, 20% de-escalator.

Using the lowest cost estimates for prostheses from the list prices supplied by the NHS Supply Chain, *Table 115* shows that, for the 10-year time horizon, although category A was cheaper, category E was more effective. For the lifetime horizon, category E dominated the other four categories. The corresponding CEACs are shown in *Figure 70*.

For this sensitivity analysis we also compared the cost-effectiveness of the different THR categories using a 20% price de-escalator to reflect in reality what NHS trusts would pay for the implants.

Table 116 shows that, for the 10-year time horizon, although category A was cheaper, category E was more effective. For the lifetime horizon, category E dominated the other four categories. The corresponding CEACs are shown in *Figure 70*.

TABLE 115 Deterministic and probabilistic results for all THR patients using the lowest cost estimates for prostheses using the bathtub model

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic *10-year time horizon*						
A	9046	7.4189	–	–	–	–
E	9322	7.4207	E vs. A	277	0.0018	153,663
D	10,080	7.4182	D vs. E	758	–0.0025	Dominated
B	10,801	7.4156	B vs. D	721	–0.0026	Dominated
C	11,750	7.4143	C vs. B	949	–0.0013	Dominated
Lifetime horizon						
E	14,102	14.7909	–	–	–	–
A	14,402	14.7887	A vs. E	301	–0.0022	Dominated
D	15,533	14.7881	D vs. A	1130	–0.0006	Dominated
B	16,450	14.7861	B vs. D	918	–0.0020	Dominated
C	17,864	14.7845	C vs. B	1414	–0.0016	Dominated
Probabilistic *10-year time horizon*						
A	9042	7.4187	–	–	–	–
E	9326	7.4204	E vs. A	283	0.0017	165,912
D	10,081	7.4180	D vs. E	755	–0.0024	Dominated
B	10,799	7.4154	B vs. D	719	–0.0026	Dominated
C	11,750	7.4140	C vs. B	950	–0.0013	Dominated
Lifetime horizon						
E	13,618	14.7917	–	–	–	–
A	14,391	14.7887	A vs. E	773	–0.0040	Dominated
D	15,534	14.7870	D vs. A	1143	–0.0007	Dominated
B	16,437	14.7851	B vs. D	903	–0.0020	Dominated
C	17,840	14.7835	C vs. B	1403	–0.0016	Dominated

TABLE 116 Deterministic and probabilistic results for all THR patients assuming a price de-escalator of 20% using the bathtub model

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic *10-year time horizon*						
A	9132	7.4189	–	–	–	–
E	9344	7.4207	E vs. A	212	0.0018	117,489
D	10,058	7.4182	D vs. E	714	–0.0025	Dominated
B	10,552	7.4156	B vs. D	494	–0.0026	Dominated
C	11,338	7.4143	C vs. B	786	–0.0013	Dominated
Lifetime horizon						
E	14,123	14.7909	–	–	–	–
A	14,489	14.7887	A vs. E	366	–0.0022	Dominated
D	15,510	14.7881	D vs. A	1,021	–0.0006	Dominated
B	16,201	14.7861	B vs. D	690	–0.0020	Dominated
C	17,452	14.7845	C vs. B	1,252	–0.0016	Dominated
Probabilistic *10-year time horizon*						
A	9138	7.4184	–	–	–	–
E	9296	7.4209	E vs. A	158	0.0025	62,906
D	10,066	7.4177	D vs. E	770	–0.0032	Dominated
B	10,558	7.4155	B vs. D	492	–0.0026	Dominated
C	11,342	7.4138	C vs. B	784	–0.0013	Dominated
Lifetime horizon						
E	14,012	14.7910	–	–	–	–
A	14,484	14.7883	A vs. E	472	–0.0026	Dominated
D	15,504	14.7877	D vs. A	1020	–0.0006	Dominated
B	16,193	14.7857	B vs. D	689	–0.0020	Dominated
C	17,450	14.7841	C vs. B	1257	–0.0016	Dominated

Sensitivity analysis: utilities

For this sensitivity analysis, utility values from Rolfson *et al.*[298] were used in the Markov model. *Table 117* shows the deterministic and probabilistic results for the 10-year and lifetime horizons. For the 10-year time horizon (both deterministic and probabilistic), category A was cheaper than the other four categories; however, slightly more QALYs were generated for category E than for the other four categories. The ICER for category A compared with category E was £153,067 per QALY gained for the deterministic analysis and £150,644 per QALY gained for the probabilistic analysis. However, when looking at the lifetime scenarios (both deterministic and probabilistic), category E dominated the other four categories. The corresponding CEACs are shown in *Figure 71*.

TABLE 117 Deterministic and probabilistic results using utility values from Rolfson *et al.*[298]

Category	Total mean cost (£)	Total mean QALYs	Comparison	Incremental cost (£)	Incremental QALYs	ICER (£)
Deterministic *10-year time horizon*						
A	9444	7.5764	–	–	–	–
E	9743	7.5783	E vs. A	299	0.0020	153,067
D	10,588	7.5757	D vs. E	845	–0.0027	Dominated
B	11,155	7.5728	B vs. D	567	–0.0029	Dominated
C	12,112	7.5714	C vs. B	957	–0.0014	Dominated
Lifetime horizon						
E	14,522	15.1174	–	–	–	–
A	14,801	15.1146	A vs. E	278	–0.0028	Dominated
D	16,040	15.1139	D vs. A	1240	–0.0007	Dominated
B	16,804	15.1115	B vs. D	764	–0.0024	Dominated
C	18,226	15.1094	C vs. B	1422	–0.0021	Dominated
Probabilistic *10-year time horizon*						
A	9443	7.5760	–	–	–	–
E	9741	7.5780	E vs. A	298	0.0020	150,644
D	10,590	7.5752	D vs. E	848	–0.0027	Dominated
B	11,153	7.5724	B vs. D	564	–0.0028	Dominated
C	12,114	7.5709	C vs. B	960	–0.0015	Dominated
Lifetime horizon						
E	14,504	15.1178	–	–	–	–
A	14,795	15.1149	A vs. E	291	–0.0029	Dominated
D	16,023	15.1142	D vs. A	1228	–0.0007	Dominated
B	16,807	15.1118	B vs. D	784	–0.0024	Dominated
C	18,208	15.1098	C vs. B	1402	–0.0020	Dominated

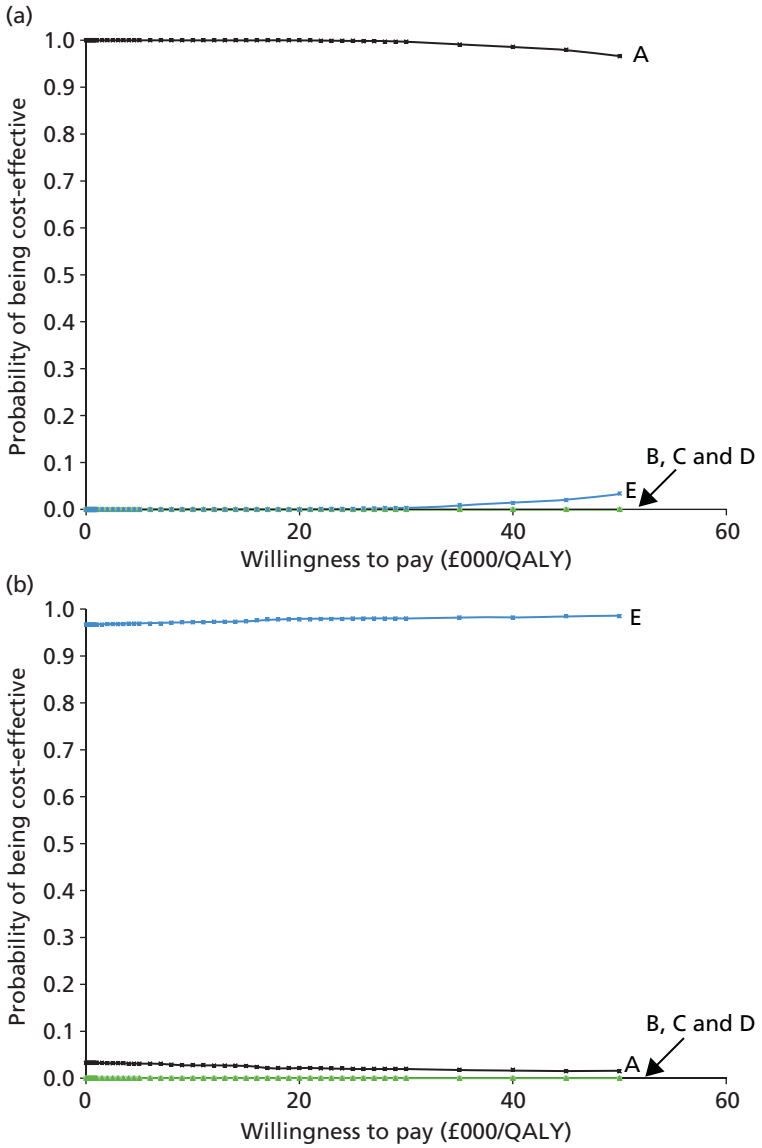


FIGURE 71 Cost-effectiveness acceptability curves for the comparison of different types of THR using utility values from Rolfson *et al.*[298] (a) 10-year time horizon; and (b) lifetime horizon.

One-way sensitivity analysis (tornado diagram)

We undertook sensitivity analysis in which we varied a number of important variables while holding others constant to compare the relative importance of particular variables in driving our estimates of the lifetime net monetary benefit of CeCoP (category E) compared with CeMoP (category A) at a WTP threshold of £20,000. The tornado diagram (*Figure 72*) illustrates our findings. For each variable the diagram indicates the changes to the inputs.

The diagram is centred around the net monetary benefit of CeCoP (category E) compared with CeMoP (category A) at a WTP threshold of £20,000 (£321). We can see that the cost of the prosthesis is the most important factor and that for each of CeCoP and CeMoP a variation of 30% in cost has a dramatic effect on our calculation of net monetary benefit. The discount rate for costs and the costs of revision are also important, as is the CeMoP alpha parameter, that is, the revision rate setting for CeMoP within the model.

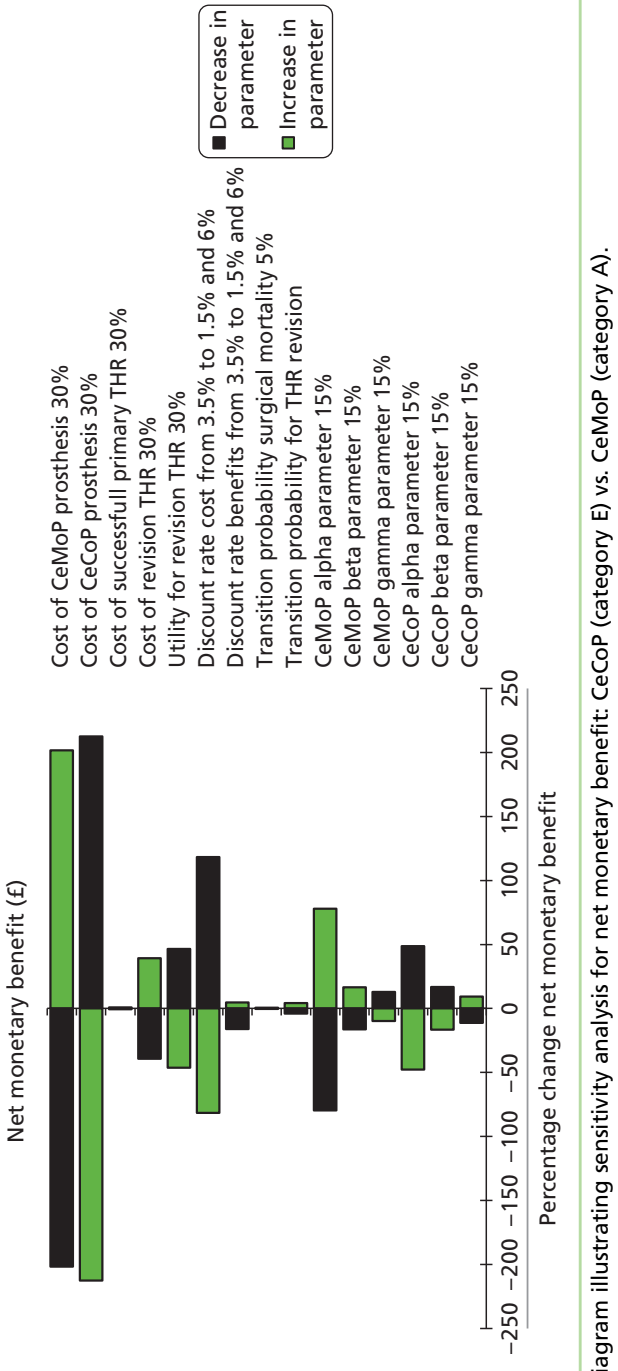


FIGURE 72 Tornado diagram illustrating sensitivity analysis for net monetary benefit: CeCoP (category E) vs. CeMoP (category A).

Discussion of the economic assessment

We built a Markov, multistate model to investigate both RS and THR. Health states included successful primary surgery, revision surgery, successful revision surgery and death. The cycle length was 1 year.

We adopted a 10-year and a lifetime horizon. The analysis was conducted from the perspective of the NHS and PSS. All costs are in UK pounds in 2011/12 prices. Health outcomes were measured in QALYs. Results are expressed as incremental cost per QALY gained. An annual discount rate of 3.5% was applied to both costs and outcomes. We ran the model deterministically and probabilistically with 1000 iterations. We calculated CEACs and undertook sensitivity analyses.

We used NHS Supply Chain costs for both RS and THR for follow-up and revision. We used age- and sex-adjusted utility values from the PROMs data set for both THR and RS. For the comparison of RS with THR we undertook sensitivity analyses stratified by sex and controlled for age. We assessed cost-effectiveness for men and women aged 40, 50 and 60 years using lifetime revision rates. We constructed CEACs comparing RS with THR overall and in separate age groups at different levels of WTP.

We compared the five categories of THR with each other, investigating patients eligible for THR (all patients) and those less eligible for RS (aged > 65 years) in sensitivity analyses. For the base case we used costs supplied by the NHS Supply Chain for each of the components of THR (cup, liner, head, stem and coating), including both cemented and cementless options when appropriate. We used the highest and lowest list prices supplied by the NHS Supply Chain in sensitivity analyses. We used age- and sex-adjusted utility values from the PROMs data set for before and after hip replacement and for revision.

We undertook sensitivity analyses and analysis of cost drivers including investigating age and sex categories, stratifying by age (< 65 years and > 65 years), different methods of extrapolation of revision rates (using a log-normal model) and varying prosthesis costs (using NHS list prices) and discount rates. We constructed CEACs comparing different types of THR overall and in separate age groups at different levels of WTP.

Summary of results

We found that the revision rates for RS were always higher than those for THR (all THR, all of our identified categories of THR combined and each of our THR categories separately).

The weighted mean cost of the THR prosthesis obtained from the NHS Supply Chain was £2571. The prosthesis cost for RS was sourced from the NHS Supply Chain and was reported as £2672, which is £101 more expensive than the cost of THR. This corresponds with the literature, in which the cost of RS has been reported as being more expensive than the cost of THR.[40] For all analyses the mean cost for RS was higher than that for THR and the mean QALYs gained were lower. The ICER for RS was dominated by THR, that is, THR was cheaper and more effective than RS (for a lifetime horizon in the base-case analysis, the total incremental cost of RS was £11,490 and the total incremental QALYs were –0.0879).

Very similar results were obtained for the deterministic and probabilistic results for RS compared with THR and when they were analysed separately in sensitivity analyses for men and women by age group (40, 50 and 60 years). For all age and sex groups RS remained clearly dominated by THR. The CEACs showed that, for all patients, THR was almost 100% cost-effective at any WTP level.

For different types of THR, given the lack of high-quality RCT evidence we used the NJR as our major source of information. We identified five categories of commonly used types of THR: category A: CeMoP (cemented–cemented with a polyethylene–metal articulation; 125,285 patients); category B: CeLMoP (cementless–cementless with a polyethylene–metal articulation; 37,874 patients); category C: CeLCoC (cementless–cementless with a ceramic–ceramic articulation; 34,754 patients); category D: HyMoP [hybrid (cementless–cemented) with a polyethylene–metal articulation; 28,471 patients] and category E: CeCoP (cemented–cemented with a polyethylene–ceramic articulation; 12,075 patients).

There were age and sex differences in the populations undergoing different types of THR and variations in revision rates. For all interventions, the revision rate at 9 years was substantially less than the benchmark rate of 10% (category A: 2.5%; category B: 3.2%; category C: 3.5%; category D: 2.5%; and category E: 1.6).

The costs of the different prostheses were as follows: category A (CeMoP): £1557.38; category B (CeLMoP): £3015.60; category C (CeLCoC): £3868.80; category D (HyMoP): £2649.78; and category E (CeCoP): £1995.98.

In the base-case analysis (both deterministic and probabilistic analysis), for all age and sex groups combined and using a bathtub model (indicating an increasing likelihood of need for revision with time) and a lifetime horizon, the mean cost for category E (CeCoP) was slightly lower and the mean QALYs for category E were slightly higher than the corresponding values for all other THR categories. Hence, category E dominated the other four categories.

For example, in the deterministic analysis, compared with category E, category A (CeMoP) cost £278 more (£14,801 vs. £14,523) and generated 0.0022 fewer QALYs (14.7887 vs. 14.7909). The probabilistic results were very similar. The CEACs demonstrated that, over a lifetime horizon, category E was 97.2% likely to be cost-effective compared with 2.8% for category A at a WTP threshold of £20,000 per QALY. For patients aged > 65 years, at a WTP threshold of £20,000 per QALY, category A was 100% cost-effective.

Sensitivity analyses using a log-normal model (indicating a decreasing risk of revision over time) for extrapolation beyond the observed data for revision rates found category A to be cheaper over a lifetime horizon for all age–sex groups combined. Although category E was more effective than the other four categories, category A was 100% cost-effective at a WTP threshold of £20,000 per QALY. Further sensitivity analysis using an age- and sex-adjusted log-normal model demonstrated the same finding: that over a lifetime horizon and at a WTP threshold of £20,000 per QALY, category A was 100% cost-effective.

Using one-way sensitivity analysis and varying the main inputs in the base-case analysis (e.g. varying costs by 30%) for all age–sex groups, when comparing category A with category E, the main drivers of difference were the costs of the components, the discount rate and the modelled revision rates.

Strengths and limitations

Although we undertook a rigorous systematic search for cost-effectiveness studies, we identified only one RCT of RS compared with THR.[40] This study reported NHS and PSS costs for the 12 months post hip replacement. The costs for a successful primary procedure were taken from the literature. Although these figures included all costs relevant to the in-hospital stay, they do not include the cost of long-term follow-up post discharge (after 12 months). Therefore, the cost of follow-up was taken from the study by Edlin *et al.*[40] We assumed that the cost of follow-up was the same in the first year and for all consecutive years across the lifetime of the model. This may have overestimated the cost of follow-up; however, little information is available in the literature to estimate the costs of, and resource use involved in, adverse events other than those requiring revision.

The difference in QALYs is negligible between the different categories of THR. On the basis of a negligible difference in QALYs it is therefore difficult to make a fair comparison between the categories in terms of outcomes. However, the costs of the prostheses vary. Category A was less expensive than category E and in the base case category E generated more QALYs than category A over a lifetime horizon. The prices for the different prostheses were obtained from the NHS Supply Chain and reflect list prices in line with the NICE reference case.[374] We therefore tested whether our results were robust to alternative costs. Here, we undertook a sensitivity analysis based on the highest and lowest list prices as reported from the NHS Supply Chain. We assumed a 20% price de-escalator to reflect what the NHS trusts would pay in reality for implants. Over a lifetime horizon, category E was less costly and more effective. This sensitivity analysis found that category E remained cost-effective even with changes to the prosthesis cost.

The cost of the prosthesis varied depending on which category was used for primary hip replacement. However, we assumed that the cost of the revision prosthesis was the same for all categories in our model. This may have either under- or overestimated the actual cost of the revision prosthesis but reflected a fair comparison across groups.

We tested whether or not our results were robust to alternative time to revision models. In the base-case analysis the revision rates were modelled using a bathtub model in which a high hazard for failure associated with surgery is followed by a decreasing hazard that plateaus during the initial recovery period and is then followed by a gradually increasing hazard with time. This time-to-revision model may disadvantage elderly patients who experience a lower revision rate. Therefore, in sensitivity analysis, revision rates were modelled using a log-normal model, which is a decreasing hazard model. Using this scenario, category A was less costly and less effective and category E was more costly and more effective using both a 10-year time horizon and a lifetime horizon. The decreasing hazard model is unlikely to capture the increasing likelihood of revision from wear and tear in the younger age group. Hence, we undertook another sensitivity analysis in which we modelled revision rates based on both bathtub and log-normal fits but adjusted for age and sex.

The utilities for the revision health state were based on PROMs data; however, PROMs data do not discriminate between different types of further surgery and so some utilities reported might reflect interventions other than revision. However, because in our model revision rate differences affect utility for 1 year only, the impact of revision rates on the overall QALYs is minimal. We were unable to incorporate adverse events that were not severe enough to lead to revision, although we were able to weight revision costs by different reasons for revision.

Ideally, outcomes, including adverse events, costs and quality-of-life data, would be collected for each patient in a single audit database. This was not the case and we had to use separate databases for outcomes and quality of life without the possibility of linking these. However, we carried out sensitivity analyses to take account of possible cost and modelled revision rate differences. We based our economic model on previous research but a strength is that we obtained an independent critique and assessment of our model and altered its structure in relation to these external comments.

Conclusion of the cost-effectiveness analysis

Compared with THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower. RS was therefore dominated by THR. Very similar results were obtained in the deterministic and probabilistic analyses and for all age and sex groups and THR was almost 100% cost-effective at any WTP level.

Revision rates for all types of THR were low. The costs of the different prostheses varied, depending partly on complexity (e.g. presence or absence of a liner). There were small but clear differences between categories in both costs and effectiveness as measured by QALYs and when age and sex were factored in. The mean total cost for category A was slightly lower and the mean QALY gain for category E was slightly higher for older age groups, in whom revision rates are lower. However, across all age–sex groups combined, in the base-case analysis, the mean cost for category E (CeCoP) was slightly lower and the mean QALYs gained for category E were slightly higher than the corresponding values for all other THR categories, for both deterministic and probabilistic analyses; therefore, category E dominated the other four categories.

Probabilistic analyses of costs and effectiveness of all categories of THR overlapped markedly, confirming that differences are relatively small. However, at the population level, although differences in costs and effectiveness are small, they are important when spread across thousands of iterations.

Comparison of the results with technology appraisal guidance 2, technology appraisal guidance 44, the manufacturer's submission and international registries

This section aims to compare the results of the Warwick economic model with TA2,[46] TA44[25] and the manufacturer's model. However, it must be noted that as we do not cover the same comparators we cannot directly compare models and findings.

National Institute for Health and Care Excellence guidance TA2,[46] issued in April 2000, suggests a benchmark revision rate of $\leq 10\%$ at 10 years. Similarly, TA44,[25] issued in June 2002, also suggests this benchmark, or a 3-year equivalent for RS. The available evidence underpinning the benchmark is old and incomplete relative to that currently available in the UK NJR and other registries. Although the THR prostheses examined in this report conformed to the $\leq 10\%$ revision rate at 10 years benchmark, the requirement for revision after RS did not (see *Chapter 6*).

One manufacturer, DePuy, submitted a review and economic analysis of THR and RS. Analyses of the following interventions were presented: cemented THR, cementless THR, hybrid THR, reverse hybrid THR and RS. Except for RS, these prosthesis types lack identity with those investigated here. The manufacturer used NJR IPD to determine revision rates and therefore, even though different prosthesis types were considered, the observed requirements for revision were broadly similar to those reported in *Chapter 6*. To extrapolate beyond the observed data the manufacturer fitted monotonic Weibull models to the observed data for all prostheses; the models were controlled for age and sex and generated a monotonically decreasing hazard with time. Decreasing was selected by the manufacturer as they suggested that other economic evaluations of this type (parametric distributions) had used Weibull distributions.

This statement is misleading as each of the economic evaluations referenced in fact employed two rather than one Weibull model, one for early and one for late revisions, so that the resulting hazard followed a U-shaped bathtub function and not a monotonic function with decreasing hazard as used by the manufacturer. The manufacturer's models predicted a decreasing hazard on extrapolation beyond the observed data but the requirements for revision beyond 10 years were not tabulated. Therefore, because of this lack of accessible data and because different prostheses were analysed, any comparison with the present results is problematic and unlikely to be informative.

Two major registries, the Swedish[96] and Australian[95] registries, provide longer follow-up of patients than the NJR, from which reliable data are available for about 9 years only.

These registries consider smaller numbers of patients but the Swedish registry provides relevant information for 19 years' follow-up. The bathtub model of hazard for revision implies that revision rates will gradually increase at some point after plateauing and this is supported by data in both of these registries. *Figure 73* shows time to revision for different age groups reported from the Swedish registry. This shows increasing rates of revision from between about 5 and 15 years of follow-up for most age groups; for these age groups the data are consistent with a bathtub hazard. For the oldest age group revision rates are relatively low and are probably not consistent with the bathtub model. Similar results are found from the Australian registry.

It should be borne in mind that long follow-up times (e.g. up to 20 years) necessitate looking at devices and practices that may no longer be widely used. The NJR data provided observed revision rates up to between 9 and 10 years only but these data may better reflect modern practice.

Further support for a bathtub model comes from the RCT by Kim *et al.*,[129] who reported extended follow-up to about 20 years; the reported revision rates were higher between 15 and 20 years than between 10 and 15 years. Several long-term follow-up observational studies provide similar evidence, as illustrated in *Figure 74*.

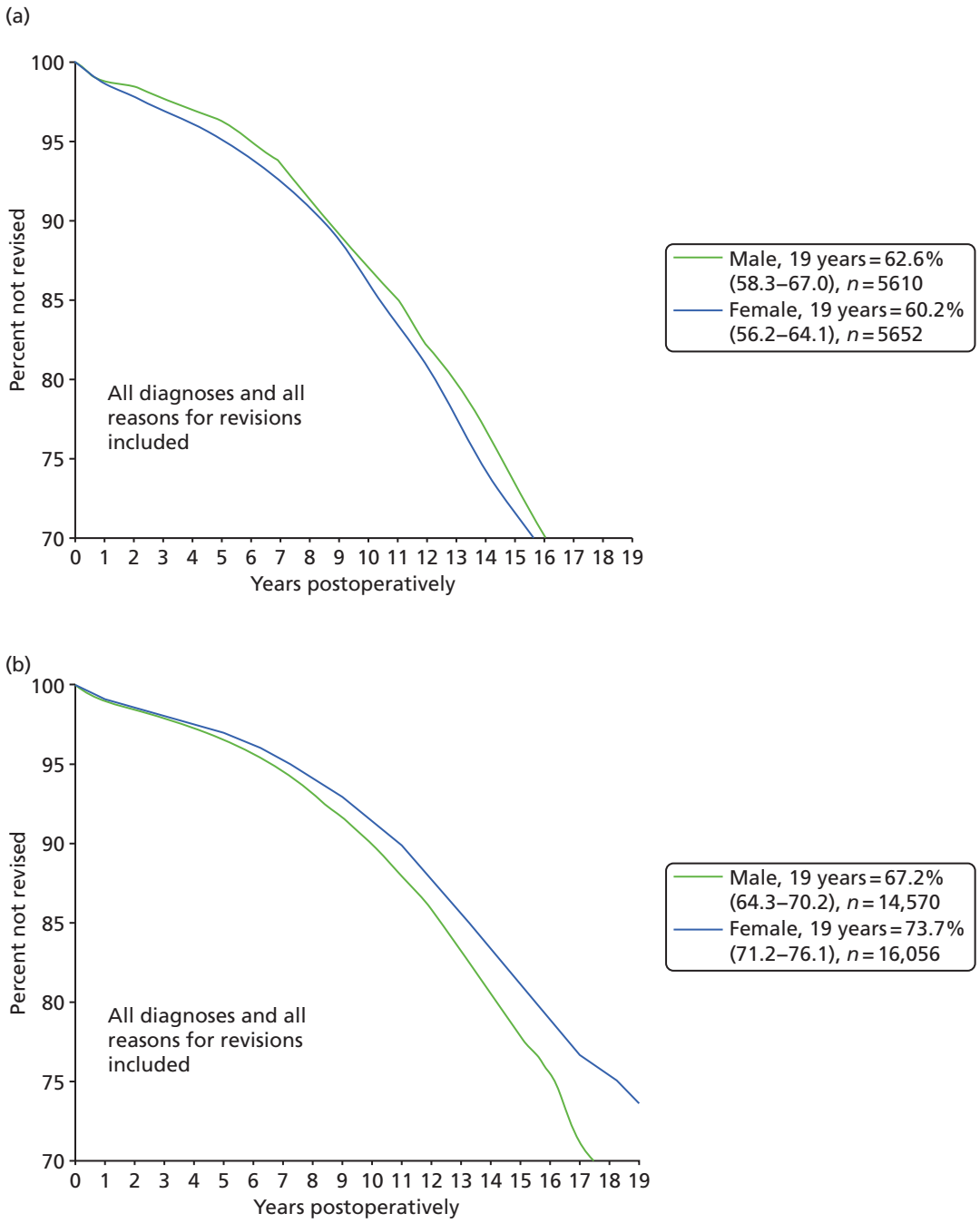


FIGURE 73 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010). (a) Patients aged < 50 years; (b) patients aged between 50 and 59 years; (c) patients aged between 60 and 75 years; and (d) patients aged > 75 years. Reproduced from Garellick *et al.*[96] and permission granted by the Swedish Hip Arthroplasty Register. (*continued*)

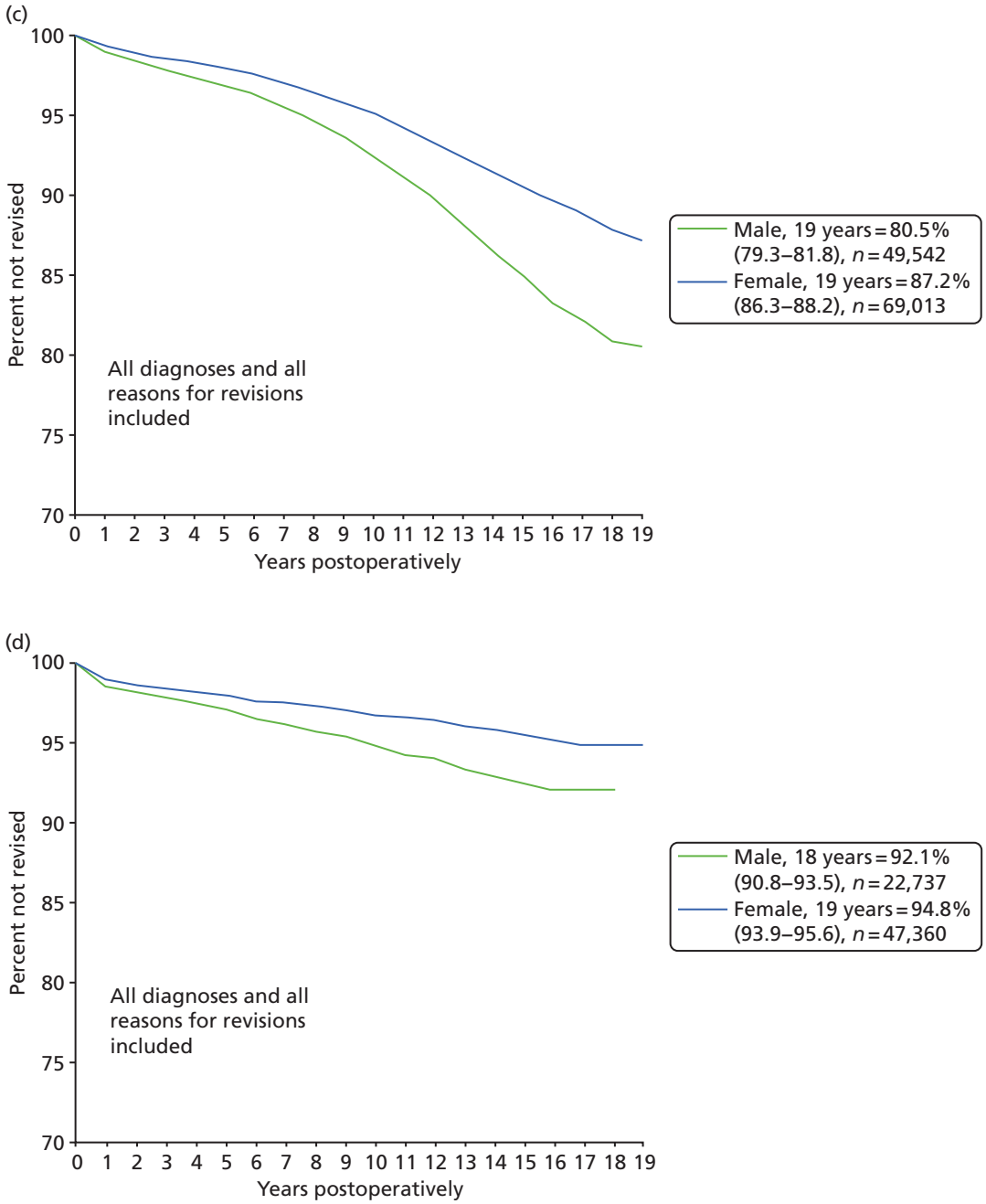


FIGURE 73 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010). (a) Patients aged < 50 years; (b) patients aged between 50 and 59 years; (c) patients aged between 60 and 75 years; and (d) patients aged > 75 years. Reproduced from Garellick *et al.*[96] and permission granted by the Swedish Hip Arthroplasty Register.

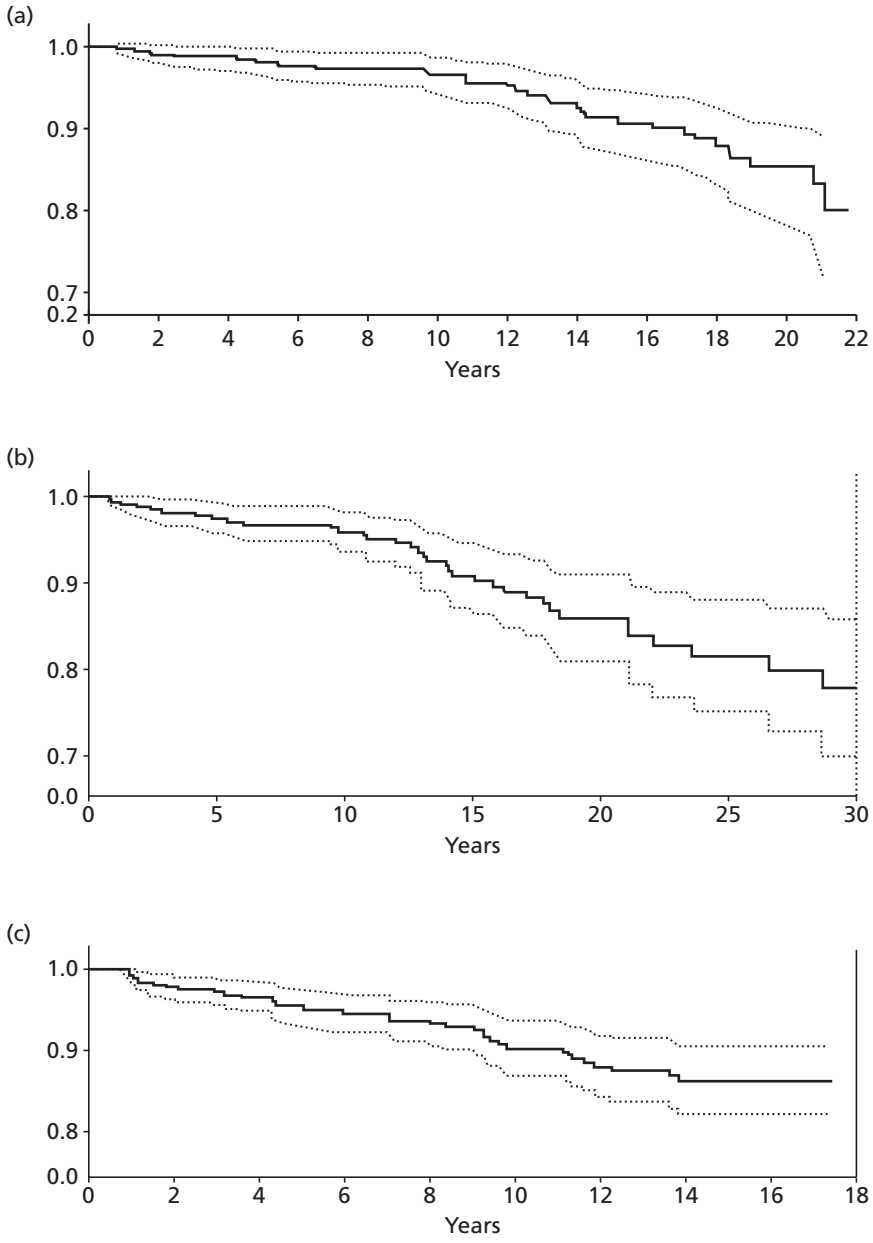


FIGURE 74 Time to revision results reproduced with permission from (a) Schulte *et al.*;[368] (b) Madey *et al.*;[369] and (c) Callaghan *et al.*[370]

Summary and critique of the manufacturers' submissions

Four manufacturer submissions were received (from DePuy International Ltd, Smith & Nephew, Inc., Stryker and JRI Orthopaedics Ltd). The following sections provide (1) a description of the submissions, (2) an evaluation of the literature searches, (3) the limitations and strengths of the clinical effectiveness reviews, (4) the overall quality considerations for the cost-effectiveness reviews, (5) a critique of the model structures (if possible) and (6) the main conclusions identified by the review team for each submission.

DePuy International Ltd

Contents of the submission

DePuy provided an economic model in Microsoft Excel (2010) and a 244-page technology assessment of the clinical effectiveness and cost-effectiveness of THR and RS for the treatment of pain or disability in adult patients with end-stage arthritis of the hip.

DePuy investigated the following comparators:

- different types of primary THR and hip RS compared with each other for people in whom both procedures are suitable
- different types of primary THR compared with each other for people in whom hip RS is not suitable.

The assessment included comprehensive systematic reviews of the clinical effectiveness and cost-effectiveness of the comparisons under review and a cost–utility analysis using a Markov model with probabilistic sensitivity analysis. The report provided details on methodology including inclusion criteria, details of the searches and databases searched for the reviews, and the structure, assumptions and sources of data for the Markov model. The model considered the following hip replacement procedures:

- cemented THR
- cementless THR
- hybrid THR
- reverse hybrid THR
- hip RS.

Data for the model were generally derived from the NJR (revision rates), the literature (utility data) and a microcosting analysis (costs). The PROMs database[97] and the New Zealand Joint Registry[376] were further data sources.

The overall conclusions were that THR dominated hip RS in patients suitable for both procedures and DePuy cemented THR was the optimal treatment strategy for both patients suitable for hip RS and those unsuitable for hip RS. Between different classes of THR, costs and QALYs overlapped considerably in sensitivity analyses for both patient populations.

DePuy recommended that the choice of prosthesis should not only be based on the results of cost–utility analyses but should also take into consideration the operational issues associated with the provision of hip replacement, the impact of training, the variability of costs and results between centres and the preference of different centres for the use of particular implants on the basis of effectiveness, efficiency and costs at a local level.

Literature search considerations

The searches reported in the DePuy submission are thorough and accurate. However, there are several concerns:

1. The MEDLINE In-Process & Other Non-Indexed Citations database was searched in the normal MEDLINE database with a strategy that ends by using limits assigned by NLM indexers. This means that all of the In-Process articles that the search initially found would not have been retrieved in the final set.
2. Most of the searches were limited by age group, which is not good practice because not all articles are age specific and NLM's indexing by age can be unreliable. For example, the systematic review by Ethgen *et al.*[192] included in the current report would not have been retrieved because it has not been indexed for age.
3. A grey literature search was not undertaken.

Strengths and limitations of the clinical effectiveness review

Strengths

The manufacturer's description of the underlying health problem and the overview of current service provision appear to be appropriate and relevant to the decision problem under consideration. The clinical evidence submitted by the manufacturer appears to reflect the characteristics of the patient population in England and Wales eligible for treatment. The interventions, comparators and outcomes described by the manufacturer match those described in the final scope. The review answers a clearly formulated research question, includes a comprehensive search and prespecified the inclusion/exclusion criteria. The screening of identified evidence and data extraction of eligible studies were carried out independently and the study and baseline population characteristics are well presented in tables.

Limitations

The clinical effectiveness review lacks a standardised quality assessment of the included studies and risk of bias assessment and the review does not report a list of excluded studies. It is unclear whether the extracted data were cross-checked by another reviewer and tables with study results are not presented. Furthermore, there is no narrative synthesis of study and baseline population characteristics (only in tables) and the results were not synthesised (i.e. given separately for each study). There is no discussion section in the report; instead, a short concluding paragraph is presented. However, the conclusions are vague for both comparisons, with no clear take-home message on what the overall findings are and whether they are conclusive. If findings were inconclusive, for instance because of clinical heterogeneity or inconsistent results, a statement acknowledging that fact should have been given. No information on the validity of the findings, implications, knowledge gaps, future research needs and limitations/advantages of the review is presented. Finally, the manufacturer's submission does not include a section on equity considerations.

Cost-effectiveness review: overall quality considerations

The reviews undertaken to identify health state utilities and costs for use in the economic analysis are comprehensive and accurate, using comprehensive searches and inclusion/exclusion criteria that are in line with the research question. A small number of relevant papers were not retrieved by the searches. The cost review was limited to studies reporting cost–utility analyses and cost per QALY outcomes; this might have restricted the review, resulting in the exclusion of studies reporting basic costs and/or resource use for patients undergoing THR or RS. Study selection is transparent; however, no table of excluded studies with reasons is given. The review did not provide a standardised quality assessment of the included studies nor of the key studies that provided data for the economic model. The data extraction tables are detailed but there is no indication whether data extraction was cross-checked by another reviewer. The review is lacking a narrative description of the included studies.

Even though the reviews identified a number of relevant studies, only two key studies were selected to provide data for the model, one for utilities and one for costs. These two studies,[38,298] investigated THR only and did not provide sufficiently up-to-date utility and cost data or revision rates or include long follow-up times.

Model structure

A Markov model using a state transition approach was developed in Excel. The structure of the model is consistent with previous cost-effectiveness models of THR for the HTA programme.[19,38] The manufacturer considered two cohorts of patients with pain and disability resulting from arthritis of the hip, one for whom hip RS or THR is suitable and one not suitable for RS who received THR. The population selected and the interventions and comparators are appropriate, as outlined in the NICE scope.[375] The model assumes a quarter-year cycle and a lifetime horizon is adopted. The perspective adopted for the analysis is that of the NHS and PSS. Both costs and benefits were discounted at 3.5%.

Categories of total hip replacement

The categories of THR included in the model comprised cemented, cementless, hybrid, reverse hybrid DePuy cementless (Corail®/Pinnacle®) and DePuy cemented. No clear justification was given for the choice of categories.

Methods defining the population for whom resurfacing arthroplasty is suitable/unsuitable

The definition of the two populations of patients, that for whom both THR and hip RS are suitable and that for whom only THR is suitable, was based on IPD provided by the NJR. The manufacturer used population characteristics data from the NJR to create their base case for both hip RS and TMR; this was applied to patients who were suitable or unsuitable for hip RS. The population characteristics of the two groups are given in *Table 118*. The impact of this assumption was tested in subsequent sensitivity analyses.

Resource costs and utility values used in the model

Tables 119 and *120* list the utilities and costs, respectively, used in the manufacturer's model, the sources of the values and the manufacturer's justification for using the values.

TABLE 118 Patient populations considered in the manufacturer's model based on NJR data

Population	Mean age (years)	Male (%)	Assumption
RS	55.3	70.9	Suitable for THR and RS
THR	70.4	37.5	Not suitable for RS
All patients	69.2	40.5	

TABLE 119 Utility values used in the manufacturer's model

Utility	Value	Source	Justification
Preoperative utility	0.41	Rolfson *et al.*[298]	This study had a very large sample size (32,396 patients from the SHAR) and reported preoperative and 1-year postoperative utility values. Reported EQ-5D scores using the UK EQ-5D tariff (p. 103)
Postoperative utility	0.78	Rolfson *et al.*[298]	
Post-revision disutility	0.145	Dawson *et al.*[295]	To reflect the lower quality of life associated with a subsequent surgical intervention, which was considered appropriate by clinical experts (p. 115)

TABLE 120 Resource use costs used in the manufacturer's model

Resource	Cost (£)	Source	Justification
Prosthesis (p. 118)			
Cemented	1029.00	Unit costs: DePuy list prices. Number of units: assumption based on the 9th Annual Report of the NJR[48]	List price prosthesis costs in line with the NICE reference case (p. 116)
Cementless	2550.50		
Hybrid	2011.50		
Reverse hybrid	1568.00		
All THR	1811.32	Weighted average of all THR prostheses	NR
RS	1029.00	Same as for cemented THR	Cemented prostheses are the least costly and therefore lifetime costs are less likely to be overestimated. Expert clinical opinion suggested that approximately 90% of RS procedures are performed with cement on the femoral side and therefore cost is similar to that for cemented THR
Surgical resource use: anaesthetic costs (p. 119)			
Cemented	Commercial-in-confidence data removed	Based on data obtained from a leading NHS orthopaedic hospital in England and validated by expert clinical opinion	NR
Cementless	Commercial-in-confidence data removed		
Hybrid	Commercial-in-confidence data removed		
Reverse hybrid	Commercial-in-confidence data removed		
All THR	Commercial-in-confidence data removed	Weighted average based on the number of primary THRs reported in the 9th Annual Report of the NJR[48]	NR
RS	Commercial-in-confidence data removed	See cemented THR	See cemented THR
Surgical resource use: surgical consumables (p. 119)			
Cemented	Commercial-in-confidence data removed	Based on data obtained from a leading NHS orthopaedic hospital in England and validated by expert clinical opinion	NR
Cementless	Commercial-in-confidence data removed		
Hybrid	Commercial-in-confidence data removed		
Reverse hybrid	Commercial-in-confidence data removed		
All THR	Commercial-in-confidence data removed	Weighted average based on the number of primary THRs reported in the 9th Annual Report of the NJR[48]	NR
RS	Commercial-in-confidence data removed	See cemented THR	See cemented THR

TABLE 120 Resource use costs used in the manufacturer's model (*continued*)

Resource	Cost (£)		Source	Justification
Staff and theatre time (p. 119)				
	Number of minutes:	Total cost (£):		
Cemented	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Microcosting analysis	To provide an accurate assessment of the cost differences between different prosthesis classes as NHS reference costs do not disaggregate costs for procedures with cement and those without cement (p. 115)
Cementless	Commercial-in-confidence data removed	Commercial-in-confidence data removed		
Hybrid	Commercial-in-confidence data removed	Commercial-in-confidence data removed		
Reverse hybrid	Commercial-in-confidence data removed	Commercial-in-confidence data removed		
All THR	Commercial-in-confidence data removed	Commercial-in-confidence data removed		
RS	Commercial-in-confidence data removed	Commercial-in-confidence data removed		
Cost of the primary procedure (p. 121)				
RS	Commercial-in-confidence data removed		See cemented THR	See cemented THR
Cemented THR	Commercial-in-confidence data removed		Microcosting analysis	To provide an accurate assessment of the cost differences between different prosthesis classes as NHS reference costs do not disaggregate costs for procedures with cement and those without cement (p. 115)
Cementless THR	Commercial-in-confidence data removed			
Hybrid THR	Commercial-in-confidence data removed			
Reverse hybrid THR	Commercial-in-confidence data removed			
All THR	Commercial-in-confidence data removed			
DePuy cementless (Corail/ Pinnacle)	Commercial-in-confidence data removed			
DePuy cemented	Commercial-in-confidence data removed			
Follow-up costs	467.00		Department of Health PbR 2012–13 tariff	Assumed to cover rehabilitation costs during the first 3 months post surgery
Cost of revision	13,399.42		Assumption that revision cost is double the mean cost of the primary procedure irrespective of class	Based on expert opinion, the cost of revision surgery is considerably greater than the cost of the primary procedure

NR, not reported.
Commercial-in-confidence data removed.

The base-case results

Total hip replacement compared with resurfacing arthroplasty in patients suitable for both procedures

The base-case results reported by DePuy for the comparison between THR and RS (p. 124) are shown in *Tables 121* and *122*.

Comparison of different categories of total hip replacement in patients not suitable for resurfacing arthroplasty

The base-case results for the comparison of different THR categories in patients not suitable for hip RS showed that DePuy cemented THR was the most cost-effective intervention, dominating cemented THR, reverse hybrid THR, all THR, DePuy cementless THR, cementless THR and hybrid THR (p. 125) (*Table 123*). Hybrid THR had an ICER of (commercial-in-confidence information has been removed) compared with DePuy cemented THR.

Results of the sensitivity analyses undertaken by the manufacturer

DePuy undertook five scenario analyses to 'investigate the impact on the results of key methodological assumptions, including those relating to procedure costs, HRQoL [health-related quality of life], and the extrapolation of the NJR data set' (p. 128).

1. NHS reference costs[379] for hip replacement procedures were used instead of costs from the micro study. The analysis identified hybrid THR as the optimal strategy at a WTP threshold of £20,000 per QALY for patients suitable for THR and RS. Comparison of different categories of THR showed only small differences in total costs and QALYs gained.
2. EQ-5D utilities from PROMs were used to investigate the impact of health-related quality of life on the ICERs. In this analysis, DePuy cemented THR was the optimal strategy at a WTP threshold of £20,000 per QALY for both patient populations.
3. An exponential model for the risk of revision was used to investigate the impact of transition probabilities that were independent of time. In this analysis the cost of hip RS was substantially greater than that of any class of THR, and DePuy cemented THR was the optimal strategy at a WTP threshold of £20,000 per QALY for both patient populations.
4. The impact of alternative Weibull models of revision stratified by age at primary procedure (< 70 years) was investigated. DePuy cemented THR and cemented THR accrued the lowest costs and DePuy cemented THR was the optimal treatment strategy at a WTP threshold of £20,000 per QALY in both patient populations.
5. The impact of alternative Weibull models of revision stratified by age at primary procedure (< 55 years) was investigated. In this analysis DePuy cemented THR was the most expensive class of THR and hybrid THR was the optimal strategy at a WTP threshold of £20,000 per QALY in both patient populations.

The results of the scenario analyses are reported in *Tables 124* and *125*.

TABLE 121 Base-case results for the comparison between THR and RS in patients suitable for both procedures

Procedure	Cost (£)	Life-years	QALYs
All THR	8894	14.391	11.115
RS	11,399	14.387	11.009
Difference	–2504.31	0.004	0.106
ICER (£)	All THR dominates		

TABLE 122 Base-case results for the comparison between different categories of THR and RS in patients suitable for both procedures

			Commercial-in-confidence information removed		Commercial-in-confidence information removed		
Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed					
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed

Commercial-in-confidence data removed.

TABLE 123 Base-case results for the comparison of different THR categories in patients not suitable for RS

			Commercial-in-confidence information removed		Commercial-in-confidence information removed		
Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed					
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed

Commercial-in-confidence data removed.

TABLE 124 Results reported in the DePuy submission for the scenario analyses for patients suitable for THR and hip RS

	Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed	
Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed	
Commercial-in-confidence data removed	Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed	

Commercial-in-confidence data removed.

TABLE 125 Results reported in the DePuy submission for the scenario analyses for patients not suitable for hip RS

	Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed		Commercial-in-confidence information removed	
Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed	
Commercial-in-confidence data removed	Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed		Commercial-in-confidence data removed	

Commercial-in-confidence data removed.

Univariate sensitivity analyses were carried out for patients suitable for THR and RS. From this a Tornado diagram was produced, with the key parameters varied from the base-case inputs across a plausible range of values. This generally showed that all THR was cost-effective (dominant in most cases) in every univariate sensitivity analysis.

Probabilistic sensitivity analyses were carried out in the form of 10,000 Monte Carlo simulations for patients who were suitable for both THR and hip RS and for patients who were unsuitable for RS. For patients unsuitable for RS, probabilistic sensitivity analysis showed that there is substantial overlap between each of the technologies in terms of costs and QALYs and that the incremental differences are negligible. DePuy concluded that all classes of THR may be considered equivalent. For patients suitable for THR and hip RS, RS was associated with substantially higher costs and fewer QALYs than all classes of THR.

The results of the probabilistic sensitivity analysis are reported in *Tables 126* and *127*.

Strengths and weaknesses of the model

Strengths

The model by DePuy has several strengths. These are:

1. The model is a de novo cohort model with transition probabilities (NJR data base), utilities (literature) and resource use (microcosting analysis). By rerunning the model, the review team could replicate the base-case deterministic and probabilistic results of the model.
2. Resource use was based on a detailed bottom-up costing method (i.e. time and motion study).
3. Prostheses costs were based on the manufacturer's list prices rather than the average selling price available to the NHS, which is conservative from the NHS perspective.
4. Costs for the model were reported separately as surgical, in-hospital stay and implant costs. This is in contrast to models in the literature, which tend to use NHS reference costs, which include all three cost components in a single value. NHS reference costs were subsequently investigated in sensitivity analyses.

TABLE 126 Results reported in the DePuy submission for the probabilistic sensitivity analysis for patients suitable for THR and hip RS

Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed

Commercial-in-confidence data removed.

TABLE 127 Results reported in the DePuy submission for the probabilistic sensitivity analysis for patients not suitable for hip RS

Commercial-in-confidence information removed	Commercial-in-confidence information removed	Commercial-in-confidence information removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed
Commercial-in-confidence data removed	Commercial-in-confidence data removed	Commercial-in-confidence data removed

Commercial-in-confidence data removed.

Limitations

The limitations of the model related to the following areas.

Revision rates Revision rates were modelled using a single Weibull fit that predicted a monotonic decreasing hazard through time. A bathtub hazard was briefly considered following Briggs *et al.*[38] The graphs of observed revision rates that were included in the submission indicate that for most an increasing rate of revision occurred from about 4 years after primary hip replacement and therefore it is likely that a bathtub model could have been used. The submission acknowledges that this is a limitation of the modelling. The manufacturer's probabilistic sensitivity analyses were described as 'including the use of multivariate distribution for revision model regression parameters'; however, this was difficult to confirm with the model version received.

The submission claims that the Weibull parametric distribution was 'chosen because all previous economic evaluations which assumed parametric distributions assumed Weibull distributions', naming the models of Briggs *et al.*[38] and Higashi and Barendregt.[273] This statement is misleading because the two models mentioned used two Weibull fits (one to early and one to late failures) to generate a U-shaped hazard, whereas in direct contrast the manufacturer's single Weibull generates a monotonic decreasing hazard.

Health-related quality of life The manufacturer has applied a disutility score of 0.145 following revision and referenced it to Briggs *et al.*[367] It should be noted that the figure for disutility was originally from a regression model output. Dawson *et al.*[295] reported the mean EQ-5D scores of 601 revision patients in the UK; following revision surgery the mean EQ-5D score at 1 year was 0.62. However, applying disutility (0.145) to the postoperative utility score does not reflect the lower quality of life as reported in the original study (0.62 vs. 0.635).

Resources and costs The cost-effectiveness of the DePuy cementless prosthesis (i.e. Corail/Pinnacle) and the DePuy cemented prosthesis was compared with that of the different types of THR and RS. In the base-case analysis the costs were based on a microcosting analysis and in a scenario analysis NHS reference costs were used.[379] It was assumed that all patients who received a primary THR received a metal-on-polyethylene articulation (regardless of whether they received a cemented, cementless or hybrid prosthesis). We agree with the manufacturer that the list prices for DePuy products do not reflect the prices available to the NHS, which results in uncertainty around the manufacturer's ICERs.

The variability in resource use observed across the sample population used to estimate the costs from the NHS hospital in the time and motion study has not been specified in the manufacturer's report, which further increases the uncertainty around the cost data inputs. The cost data for surgical resource use, anaesthetic costs and theatre time reported in Appendices E and H in the DePuy submission are all based on this microcosting study undertaken by DePuy. Because the observational methods and the variance in resource use across the sample population were not reported in the submission, the review team was unable to verify the data. Although undertaking a time and motion study to determine cost data inputs is desirable, to report a base-case economic analysis using costs that cannot be verified is questionable.

DePuy assumed a unit cost of an inpatient stay of £295.29 basing the calculations on LOS data detailed in the NHS reference cost database[379] (*Table 128*). However, individual costs for the respective HRG codes were not reported and DePuy did not detail how it derived the costs for the weighted average LOS, which meant that the review team was unable to replicate the value used.

The review team's clinical expert opinion suggests that the cost of revision surgery is greater than the cost of primary THR/RS, but revisions are carried out for a variety of reasons and to assume that the cost of all revision procedures is the same is not reasonable. In light of this, the manufacturer should have presented a sensitivity analysis around the costs associated with different indications for revision surgery.

Overall results The manufacturer has presented base-case deterministic and probabilistic results. All THR dominates RS in the comparison of patients suitable for THR and RS. In the patient population in whom RS is not suitable, DePuy cemented THR was reported as the most cost-effective intervention. However, this result is dependent on the allocation of the relatively high cost of surgery with a hybrid prosthesis. However, no methodology was reported detailing how the model controlled for age and sex differences, even though differences in both age and sex distributions were reported by DePuy (p. 77) (*Table 129*). No attempts were made to identify the cost-effectiveness of the different types of prosthesis based on age and sex. Subgroup analysis of patients based on age and sex is desirable when comparing THR and RS because of the dissimilarities among the different patient populations.

The base-case probabilistic results are similar to the deterministic results. Although the model was probabilistic, the parameters in the model were assumed to be independent and no attempt has been made to check for correlation between the parameters.

For the base-case analysis the manufacturer's submission was largely in line with the NICE reference case.[374] However, costs in the base-case analysis were not based on NHS reference costs but on a microcosting study. As mentioned earlier the microcosting study could not be verified; however, the NHS reference cost estimates were based on a large sample size for both primary and revision surgery

TABLE 128 Mean LOS for patients receiving primary THR or hip RS

HRG name (currency code)	FCEs	LOS (days)	Unit cost per inpatient stay per day (£)	Source
Major Hip Procedures for non Trauma Category 1 with Major CC (HB12A)	2573	9.92		*NHS reference costs*[379]
Major Hip Procedures for non Trauma Category 1 with CC (HB12B)	6433	5.53		
Major Hip Procedures for non Trauma Category 1 without CC (HB12C)	34,414	4.45		
Weighted average	43,420	4.93	295.29	

CC, currency code; FCEs, finished consultant episodes.

TABLE 129 Age and sex of patients receiving primary hip replacements in 2011[a]

	Cemented THR	Cementless THR	Other THR (e.g. hybrid)	RS	Total
Total procedures, *n* (%)	25,789 (36)	31,307 (44)	12,794 (18)	1782 (2)	71,672
Total procedures with patient data, *n* (%)	24,739 (96)	29,751 (95)	12,241 (96)	1600 (90)	68,331 (95)
Female, *n* (% of class)	16,112 (65)	16,731 (56)	7743 (63)	241 (15)	40,827 (60)
Male, *n* (% of class)	8627 (35)	13,020 (44)	4498 (37)	1359 (85)	27,504 (40)
Average age (years) (SD), IQR	72.8 (9.7), 67.2–79.5	65.4 (11.3), 58.8–73.3	69.6 (10.9), 63.5–77.3	54.2 (9.5), 48.6–60.7	67.2 (13.4), 62.0–76.7

IQR, interquartile range.
a Adapted from Table 2.5 from the *National Joint Registry for England and Wales. 9th Annual Report*.[48]

($n = 43,420$ for primary surgery and $n = 26,797$ for revision surgery). Applying the NHS reference costs to both patient cohorts, the optimal strategy at a WTP threshold of £20,000 per QALY was hybrid THR for both patient cohorts. This suggests that a key uncertainty of the model is the cost data inputs that have been used.

The decision by the manufacture to not report CEACs in the main text is questionable. CEACs were included in the appendix and the reader was instructed to view them with caution. CEACs should be used to characterise the current decision problem as the treatment options are mutually exclusive.

Sensitivity analysis The manufacturer undertook a range of univariate sensitivity analyses, probabilistic analyses and also additional scenario analyses. However, the scenario analysis, which used costs from Vale *et al.*,[19] as indicated in the submission, could not be identified in the report. Given that the cost of revision increased by only 45% and not double, the cost of revision should have been tested using inflated Vale *et al.*[19] costs.

Conclusions

The submitted evidence reflects the decision problem defined in the final scope and the manufacturer's submission is rigorous and complete with regard to relevant clinical studies and relevant data within those studies. The submission contains an unbiased evaluation of the literature in terms of treatment effects in relation to relevant populations, interventions, comparators and outcomes. There are uncertainties about the reliability of the clinical effectiveness evidence because of weaknesses highlighted related to transparency, synthesis and lack of quality assessment. The main shortcomings of the model concern the lack of a detailed methodology of how the model controlled for age and sex differences, the lack of a cost-effectiveness analysis based on age and sex and the minimal reporting of CEACs. The main conclusion of the cost-effectiveness analysis was that the DePuy devices are more cost-effective than all other prostheses. Hip RS was dominated by cemented THR, cementless THR, DePuy cementless THR (Corail/Pinnacle), DePuy cemented THR, hybrid THR and reverse hybrid THR in patients suitable for both procedures. It was also noted that DePuy cemented THR was the optimal treatment strategy in both patient populations in the base-case analysis. It should be noted that these conclusions cannot be verified as the cost data, displaying the greatest amount of uncertainty, were derived from a microcosting analysis, which was reported incompletely.

Smith & Nephew, Inc.

Contents of the submission

Smith & Nephew provided a 10-page non-systematic summary of the literature. Evidence was presented on the factors that should be included in the sensitivity analysis of a cost-effectiveness model. No methodology was reported and no economic evaluation was presented. The evidence was drawn from the

literature as well as the National Joint Registries of England and Australia. Smith & Nephew concluded that revision rates (and implant prices) drive the cost-effectiveness of THR and that bearing surfaces are known factors that impact revision rates following primary THR and should therefore be considered in sensitivity analyses of economic evaluations.

Literature search considerations
No details of any search methods were reported.

Strengths and limitations of the clinical effectiveness review

Strengths
The revision rates reported by bearing surface were extrapolated to 11 years.

Limitations
The submission by Smith & Nephew lacks a clearly defined research question and provides a non-systematic review of the clinical effectiveness literature with a clear focus on revision surgery only. RS is not considered as an intervention. Therefore, the population, intervention and comparator considered by Smith & Nephew only partially match those described in the final scope.

The review does not report any methodology nor does it specify any inclusion criteria. The clinical effectiveness review lacks a standardised quality assessment of the included studies and risk-of-bias assessment. The review does not report a list of excluded studies. Only revision following THR was considered as an outcome. Furthermore, the study and baseline population characteristics are not clearly presented and the results are not synthesised. Tables with study results were omitted and the manufacturer's submission does not include a section on equity considerations.

Cost-effectiveness review: overall quality considerations
Smith & Nephew provided a non-systematic coverage of the cost-effectiveness evidence concerning revision surgery post THR. The research question therefore only partially meets the decision problem under consideration. No methods were reported in terms of the literature search, inclusion criteria, data extraction and synthesis of evidence. No quality assessment of included studies was reported nor was a list of excluded studies reported. The cost-effectiveness review included a number of key papers but the list of included studies was not exhaustive, probably because of the focus on revision.

Conclusions
The report includes a subjective summary of the importance of bearing surfaces for revision rates and a justification for considering bearing surfaces in a sensitivity analysis within the cost-effectiveness model of the NICE report.[375] It concludes that known factors that modify revision rates, such as bearing surfaces, should be considered in analyses. It suggests that individual prostheses or design elements should be considered separately in analyses so that their impact on revision rates does not get lost when grouping new technology implants for analysis. Overall, the report lacks objectivity and transparency.

Stryker

Contents of the submission
Stryker provided a 22-page report that consisted of an executive summary and a review of the literature without any evidence of a systematic review. The report did not include any methodology on how the evidence was collected nor did it report any economic analysis. Stryker considered cemented and cementless THR as well as RS and summarised the complexity of available implants and corresponding revision rates, considering evidence from the literature and the National Joint Registries of England and Wales, Sweden, Norway and Australia. It concluded that the complexity of hip replacement procedures should be taken into consideration in economic evaluations and reported that it is currently working with a

group of researchers at the University of East Anglia and orthopaedic surgeons to develop a cost-effectiveness model to address the above-mentioned issues.

Literature search considerations

No details of the search methods were reported.

Strengths and limitations of the clinical effectiveness review

Strengths

The manufacturer's description of the underlying health problem and the overview of current service provision appear to be appropriate and relevant to the decision problem under consideration. The clinical evidence submitted by the manufacturer appears to reflect the characteristics of the patient population in England and Wales eligible for treatment. The interventions, comparators and outcomes described by the manufacturer match those described in the final scope although the list of included studies is not exhaustive. The review considered PROMs data for THR and revision rates were reported for 3 and 8 years.

Limitations

Stryker provided a non-systematic review of the clinical effectiveness literature using data referenced to the NJR 2011 annual report.[36] The review lacks a clearly formulated research question and does not specify any inclusion criteria. The clinical effectiveness review lacks a standardised quality assessment of the included studies and risk of bias assessment and the review does not report a list of excluded studies. Furthermore, no details are given on the methods of screening and data extraction. Study and baseline population characteristics are not clearly presented and the results are presented in a narrative fashion and are not synthesised. The conclusions are vague and no information on the validity of the findings, implications, knowledge gaps, future research needs and limitations/advantages of the review is presented. Finally, the manufacturer's submission does not include a section on equity considerations.

Cost-effectiveness review: overall quality considerations

Stryker provided a limited non-systematic coverage of the cost-effectiveness evidence concerning THR. A brief statement is made about the complexity of the cost-effectiveness modelling around THR. Stryker states that 'few cost-effectiveness studies have been published regarding THR compared to other broadly used surgical interventions'. In contrast, the current report has identified considerable evidence on the cost-effectiveness of different types of THR.

Conclusions

Stryker did not answer a clearly formulated question but presented a summary of a selection of the available evidence. Details are provided on the cemented procedure for the Exeter™ stem (Stryker). 'Very good' mid-term results are reported for the Exeter V40™ stem. Stryker also reported results from the 2011 AOANJRR[380] for the two stems listed above. Various published studies are listed that report positive results for these stems. The information in the report is limited and the submission lacks objectivity, transparency and clear conclusions.

JRI Orthopaedics Ltd

Contents of the submission

The JRI submission consisted of a 14-page report detailing a summary of JRI products and a price list of JRI components, with limited reference to the literature and data from the National Joint Registries of England and Wales, Sweden and Australia. The submission did not include an economic evaluation in the form of a model. The report compared JRI cementless THR with cemented, hybrid and cementless THR data from the NJR and concluded that revision rates for JRI cementless implants are lower than revision rates for all other cementless THRs, the majority of the hybrid THRs and two of the six categories of cemented THRs. Analysis of risk of revision by liner type and age showed that the risk of revision increased after the age of 70 years when using a polyethylene liner instead of a ceramic liner. Furthermore, a comparison of death

rates for cemented compared with cementless JRI implants demonstrated a slightly higher death rate for patients receiving a cemented JRI implant than for patients receiving a cementless JRI implant.

The JRI submission also included detailed clinical evaluation reports on four specific JRI brands including literature reviews and quality appraisals and four technical reports considering the JRI cemented and cementless components, coatings, details of the polyethylene used and specifications of the trunnion design. Finally, JRI submitted statistics from the NJR and complaints data by device collected by JRI.

Literature search considerations

A search strategy was developed for each brand to identify relevant literature over the last 5 years. The authors state that the majority of literature for the reviews was obtained online. Searches were undertaken in the *Journal of Bone & Joint Surgery*, Entrez PubMed (National Center for Biotechnology Information, Bethesda, MD, USA), the NJR and Google Scholar (Google Inc., Mountain View, CA, USA).

Strengths and limitations of the clinical effectiveness review

Strengths

The interventions, comparators and outcomes described by the manufacturer match those described in the final scope although the list of included studies is not exhaustive. The review included a quality assessment of the included studies. Finally, the submission provided a brief review of evidence highlighting data from the NJR including the number of JRI implants, revision rates for JRI cementless brands with comparative data, survival rates and risk of revision by age group for a JRI Furlong® H-A.C THR, trends in femoral head size, revision rates by liner type with different head sizes, revision rates by liner type and age group and mortality rates for JRI cemented and cementless implants.

Limitations

The JRI submission provided only brief scoping reviews of the clinical effectiveness literature for each JRI brand and the review lacks a clearly formulated research question. The review does not detail any methods concerning screening and data extraction nor does it specify any inclusion criteria or provide a list of excluded studies. Study and baseline population characteristics are not clearly presented. The submission only briefly discusses revision rates for cemented THR compared with cementless THR from three national joint registries. The submission does not include a section on equity considerations.

Cost-effectiveness review: overall quality considerations

The submission provided very limited information on cost-effectiveness.

Conclusions

In its submission, JRI presented an overview of its brands. Accompanying reports for each brand were provided as appendices. The average selling price per component was listed, which was useful. Overall, the report lacks transparency, objectivity and any clear conclusions.

Chapter 9 Discussion

Decision problem and objectives

The main objective was to undertake a clinical effectiveness and cost-effectiveness analysis of different types of THR and hip RS for the treatment of pain and disability in people with end-stage arthritis of the hip. Specific aims were to compare the clinical effectiveness and cost-effectiveness of (1) different types of primary THR and hip RS for people in whom both procedures are suitable and (2) different types of primary THR for people who are not suitable for hip RS.

Methods and summary of findings

We undertook systematic reviews of the clinical effectiveness of RS and THR and of registry reporting and cost-effectiveness studies in December 2012. For the clinical effectiveness review, searches were undertaken in 12 databases including MEDLINE, Science Citation Index, The Cochrane Library and Current Controlled Trials and were limited to studies published from 2008 onwards and including sample sizes of ≥ 100 participants. Two independent reviewers screened all records, extracted data and independently assessed risk of bias. Estimates of effectiveness were pooled and the quality of the evidence was assessed using the GRADE approach.

Although we appraised and summarised a very large amount of evidence, much of it was inconclusive because of poor reporting, missing data, inconsistent results, inappropriate pooling methods, inconsistent summary findings and uncertainty in treatment effect estimates. Improvements post surgery were reported for functional/clinical measures and quality-of-life measures regardless of the type of THR or RS. Evidence on the relative benefits of RS compared with THR or of different types of THR was largely lacking. Certain types of THR appeared to confer some benefit, including larger femoral head size, use of a cemented cup, use of a cross-linked polyethylene cup liner and use of a ceramic-on-ceramic as opposed to a metal-on-polyethylene articulation, although the findings were not conclusive or reflected short-term follow-up. Systematic reviews of cost-effectiveness and registry studies worldwide provided costs for revision and follow-up, corroboratory utility data and registry data for validating the survival analysis. For both research questions we drew on our systematic reviews of clinical effectiveness and cost-effectiveness and registry data to identify inputs for the models to compare the clinical effectiveness and cost-effectiveness of RS with that of different types of THR and different types of THR with each other.

For the cost-effectiveness analyses we used the NJR to identify populations undergoing the various types of interventions. We identified the group undergoing RS but it became clear that there was a very large possible number of categories for those undergoing THR. Using a series of cross-tabulations by combinations of components, we identified the top four most commonly used categories of THR ($> 25{,}000$ in the database) and our clinical advisors recommended the inclusion of a further fifth mutually exclusive category. We identified time to revision for all categories by age and sex using NJR data and investigated a large number of methods for extrapolating beyond observed data and tested goodness of fit.

We built a Markov, multistate model to investigate both RS and THR. Health states included successful primary surgery, revision surgery, successful revision surgery and death. Cycle length was 1 year. We adopted a 10-year and a lifetime horizon from the perspective of the NHS and PSS. We applied an annual discount rate of 3.5% to both costs and outcomes and ran the model deterministically and probabilistically. We undertook a large number of sensitivity analyses. The economic model was independently reviewed and adjusted in response to this.

We found that the ages and sexes of RS and THR patients overlapped substantially such that with the data available it was impossible to identify mutually exclusive cohorts eligible for both THR and RS.

We therefore used propensity matching to compare RS with THR, drawing age–sex matched pairs from the RS data set and from the five categories of THR combined. We used NHS Supply Chain costs for 'major hip procedures', drawing on the same nationally available HRG4 reference costs for both RS and THR for follow-up and revision. We used age- and sex-adjusted utility values from the PROMs data set, using the same utility values for both procedures for before and after hip replacement and for revision because no separate utility values were reported for RS.

We used age- and sex-specific PROMs data and assessed estimates of cost-effectiveness for men and women aged 40, 50 and 60 years using lifetime revision rates and undertook sensitivity analyses stratified by sex and controlled for age.

We compared the five categories of THR with each other, investigating patients eligible for THR (all patients) and those less eligible for RS (aged > 65 years). For the base case we used costs supplied by the manufacturers for each of the components of THR. We used alternative costs including those supplied by local trusts when manufacturer costs were not available and alternative manufacturers' costs in sensitivity analyses. We used age- and sex-adjusted PROMs utility values for health state utilities.

We undertook sensitivity analyses and analyses of cost drivers including investigating changes in age and sex categories, stratifying by age (< 65 years and > 65 years), investigating different methods of extrapolation of revision rates (using a log-normal model) and varying prosthesis costs (using NHS list prices) and discount rates.

The NJR included just fewer than 420,000 patients. Approximately 31,000 (7.4%) patients had undergone RS. Our identified categories of THR covered 62% of the THR population. In total, 90% of RS patients and 23% of THR category patients were aged < 65 years. Bathtub models (predicting an increasing likelihood of revision over time) gave the best fit to the observed data. PROMs data showed that utility differences were dramatic, that is, 0.35 pre intervention to 0.78 post intervention and 0.53 pre revision to 0.78 post revision.

Revision rates for all RS were always higher than those for THR (all THR, all of our identified categories of THR combined, each of our THR categories separately). The mean cost of RS was £2672 and the weighted mean cost of THR was £2571.

Costs for RS were higher than those for THR and mean QALYs gained were lower. The ICER showed that RS was dominated by THR (over a lifetime horizon in the base-case analysis, the incremental cost of RS was £11,490 and the incremental QALYs were –0.0879). Very similar results were obtained for the deterministic and probabilistic results for RS compared with THR and when THR was analysed separately in sensitivity analyses for all age and sex groups. RS remained clearly dominated by THR. CEACs showed that, for all patients, THR was almost 100% cost-effective at any WTP level.

The five categories of commonly used types of THR that we investigated are cemented–cemented with a polyethylene–metal articulation (CeMoP, category A) (125,285 patients); cementless–cementless with a polyethylene–metal articulation (CeLMoP, category B) (37,874 patients); cementless–cementless with a ceramic–ceramic articulation (CeLCoC, category C) (34,754 patients); hybrid (cementless–cemented) with a polyethylene–metal articulation (HyMoP, category D) (28,471 patients); and cemented–cemented with a polyethylene–ceramic articulation (CeCoP, category E) (12,705 patients).

There were age and sex differences between the populations receiving different types of THR and variations in revision rates between category E (1.6%) and category C (3.5%) at 9 years (for all interventions, revision rates at 9 years were well under 10%). The prosthesis cost varied between £1557.38 for category A (CeMoP) and £3868.80 for category C (CeLCoC).

In the base-case analysis, for all age and sex groups combined and using a bathtub model (indicating an increasing likelihood of need for revision with time) and a lifetime horizon, Category E dominated the other four categories. The mean cost for category E was slightly lower and the mean QALYs gained for category E were slightly higher than the corresponding values for all other THR categories for both the deterministic and the probabilistic analysis.

In the deterministic analysis, compared with category E, category A (CeMoP) cost £278 more (£14,801 vs. £14,523) and generated 0.0022 fewer QALYs (14.7887 vs. 14.7909) and the probabilistic results were very similar. Over a lifetime horizon, category E was 99.9% likely to be cost-effective whereas category A was 1% likely to be cost-effective at a WTP of £20,000 per QALY.

For patients aged > 65 years, over a 10-year time horizon, and at a WTP of £20,000 per QALY, category A was more likely to be cost-effective in all groups (category A: 99% probability of being cost-effective; categories B–E: < 1% probability of being cost-effective), although category E was more effective over a lifetime horizon for all groups (except for men aged 80 years for whom the QALYs generated by categories A and E were the same).

Sensitivity analysis for all age–sex groups combined using a log-normal model (indicating a decreasing risk of revision over time) and a lifetime horizon resulted in category A being 85% cost-effective at a WTP threshold of £20,000 per QALY. Further sensitivity analysis using an age- and sex-adjusted log-normal model demonstrated that, likewise, over a lifetime horizon and at a WTP of £20,000 per QALY, category A was 100% cost-effective at a WTP of £20,000 per QALY.

The main drivers of differences between category A and category E were found to be the costs of the components, discount rates and modelled revision rates.

Strengths and limitations

We undertook rigorous systematic reviews and we believe that we identified all relevant publications concerning the clinical effectiveness and cost-effectiveness of both THR and RS as well as all available registry results. However, given the wide scope and large amount of identified evidence, we limited our inclusion for clinical effectiveness studies to those with a sample size of ≥ 100 and those published since 2008. This decision was based on our sample size calculations for clinically important differences in the HHS and the fact that smaller studies tend to be underpowered to detect meaningful differences in continuous outcomes. We pooled data when possible and used the GRADE system for assessing overall quality.

We did not find any longer-term RCTs covering the comparison between RS and THR or between different types of THR that would allow us to model differences in revision rates for RS or THR relevant to a lifetime horizon. We therefore had to use nationally collected non-randomised clinical audit data from the NJR. The NJR has a high reported coverage with good quality assessment systems and NJR data were complete for patient age and sex at the time of receipt of the THR.

However, the non-randomised nature of the database means that selection bias may be operating within the data. Revision rates may be higher, for example those selected to receive one intervention rather than another (e.g. RS) may belong to a group who have an adverse profile in the population. We worked to reduce confounding by propensity matching RS patients with THR patients using NJR data and by undertaking extensive analyses by age and sex for the comparisons of different types of THR. However, we were of course unable to adjust for confounders of which we were unaware.

The number of unique prosthesis types used for THR patients was large, even without taking into account the variety of manufacturer brands available for the different components. It was necessary to reduce these to a smaller number for economic analysis. For the comparisons of different types of THR we therefore used cross-tabulations to generate the largest categories of THR. Selection was based on the frequency of use of different categories of prosthesis and on expert clinical opinion. The selection of the five THR categories was conducted pre hoc and before all analyses of revision rates. To our knowledge this is the first time that different THR components have been investigated in this comparative way – it allows for a more granular approach to assessing the cost-effectiveness of different types of THR than previously and has the advantage of more precisely reflecting current practice.

We were able to asses only a relatively small number of categories (five) as we needed to generate appropriate costings of subcomponents and to have enough patients in each category to model revision rates reliably. This meant that we were unable to include some of the less popular combinations of components for hip replacement (38% of THRs). However, we modelled revision rates and survival rates using all hip replacements to assess how our categories A–E compared with those for RS. We found that the overall revision rate was slightly higher than the revision rates for categories A–E. Given this finding we consider that our comparisons are likely to have focused on the more cost-effective THR options.

Age and sex distributions varied between categories. When populations were controlled for differences in age and sex or were stratified by sex and controlled for age, the lower revision rate for category E relative to the other categories remained. Also, when well-fitting models that predicted either an increasing or a decreasing hazard on extrapolation were used, the superiority of the category E revision rate was again upheld. There was insufficient information recorded consistently within the NJR for investigation of other potential confounders. For example, our clinical advisors suggested that selection of patients for RS may be made by surgeons based on activity levels (levels of physical fitness, athleticism, weight lifting, manual labour); however, the only characteristics that were reliably collected at the patient level in the NJR were age and sex. This means that we were unable to identify other characteristics or subpopulations in which RS might be more beneficial. However, age and sex may act as a proxy for physicality and it is of interest that revision rates for RS were higher in every age and sex group that we examined, including in the youngest category of men.

For revision rates the unit of analysis was the time to a patient's first revision. For patients who received a THR for both hips simultaneously, only the replacement that failed first was included as an event; for those who received a THR for both hips on separate occasions, only the first primary intervention entered the analysis. To model revision rates we followed NICE DSU[366] recommendations in first exploring exponential, Weibull, Gompertz, log-normal and log-logistic models of observed revision rates based on IPD. However, previous economic analyses of hip replacement, notably those of Briggs *et al.*,[38] Higashi and Barendregt[273] and Pennington *et al.*,[44] modelled revision rates on the assumption of a U-shaped hazard. In these an assumed high hazard for failure associated with surgery is followed by a decreasing hazard that eventually plateaus during an initial recovery period and is then followed by a gradually increasing hazard as host bone deteriorates with patient age and the prosthesis accumulates wear and tear. The resulting hazard curve is commonly termed a bathtub hazard. We therefore also explored bathtub models to extrapolate revision rates beyond the observed data.

For most age groups this offered the best fit to the observed data but, for patients aged > 85 years, during the observation period the revision rate was low and extrapolation with an increasing hazard becomes less appropriate. We derived the bathtub hazard directly using the Stata package developed by Crowther and Lambert.[360] Pennington *et al.*[44] employed a piecewise procedure to generate the U-shaped hazard; however, after extrapolation this predicted that > 100% of patients sustained revision and at this point the rate required capping. A strength of this work is that we tested a large number of methods for extrapolating the revision rate including competing risk analysis and flexible parametric models.

For RS a wide range of femoral head sizes are used and revision rates have been reported to vary according to head size.[15] Only a narrow range of head sizes are used for THR prostheses and expert clinical opinion indicated that these are unrelated to RS head sizes so that comparisons of RS and THR according to head size were not undertaken. It is of interest that we identified only one RCT investigating different THR head sizes. This demonstrated an advantage from a larger head size (36mm vs. 28mm) and had a low risk of bias, although so far follow-up has continued for only 1 year.

Utilities for both models for the base-case analysis were obtained from the national PROMs database, which is comprehensive. We were unable to link NJR and PROMs data; however, we adjusted EQ-5D scores for the successful primary health state and successful revision health state to reflect age and sex differences. In our economic model we assumed that costs and utilities were the same for both RS and THR. Our model is therefore likely to represent a fair comparison but is also likely to underestimate the prosthesis cost for RS, which has been reported to be more expensive than that for THR.[40] In spite of this assumption we found THR to be cost-effective (dominant) compared with RS for all age (40, 50 and 60 years) and sex groups.

Although we undertook a rigorous systematic search for cost-effectiveness studies, little information was available in the literature to estimate costs and resource usage. We could identify only one cost–utility analysis of RS compared with THR from a RCT.[40] The costs of follow-up in our model were based on this trial; however, we assumed that the costs of follow-up were the same for the first and subsequent years across the lifetime of the model. This may have overestimated the cost of follow-up although it was applied equally to both comparators in the model.

The cost of the prosthesis varied between THR categories. Category A was the least expensive but category E had lower revision rates and generated more QALYs over the lifetime horizon. We used prices for prosthesis components obtained from the NHS Supply Chain. We undertook a sensitivity analysis based on the highest (category C, £3868.80) and the lowest (category A, £1557.38) prices. So as not to disadvantage any one category, the costs of the prostheses used in revision surgery were assumed to be the same across categories. This is likely to underestimate differences in the costs of revision. We were unable to incorporate adverse events that were not severe enough to lead to revision, although we were able to weight revision costs by different reasons for revision.

Ideally, outcomes, including adverse events, costs and quality-of-life data, would be collected for each patient in a single audit database. This was not the case and we had to use separate databases for outcomes and quality of life without the possibility of linking these. However, we carried out sensitivity analyses to take account of possible cost and modelled revision rate differences. We based our economic model on previous research but a strength is that we had an independent critique and assessment of our model and altered its structure in relation to these external comments.

Chapter 10 Conclusions and implications for practice

Total hip replacement is a common operation and is clearly beneficial. Improvements post surgery were reported in the literature for functional/clinical and quality-of-life measures regardless of the type of THR or RS received. Overall, revision rates are low. However, although we appraised and summarised a very large amount of evidence, much of the published literature was inconclusive because of poor reporting, missing data, inconsistent results and uncertainty in treatment effect estimates. Evidence on the relative benefits of RS compared with THR or of different types of THR was largely lacking. Certain types of THR appeared to confer some benefit, including larger femoral head size, use of a cemented cup, use of a cross-linked polyethylene cup liner and use of a ceramic-on-ceramic as opposed to a metal-on-polyethylene articulation.

Resurfacing arthroplasty compared with total hip replacement

Compared with THR, revision rates for RS were higher, mean costs for RS were higher and mean QALYs gained were lower; RS was therefore dominated by THR.

Very similar results were obtained for the deterministic and the probabilistic analyses and for all age and sex groups and THR was almost 100% cost-effective at any WTP level.

Comparison between different types of total hip replacement

Revision rates for all types of THR were low. The costs of the prostheses varied depending partly on complexity (e.g. presence or absence of a liner). There were small but clear differences between categories in both costs and effectiveness as measured by QALYs and when age and sex were factored in. Category A was more cost-effective for older age groups in whom revision rates are lower. However, across all age–sex groups combined, in the base-case analysis the mean cost for category E (CeCoP) was slightly lower and the mean QALYs gained for category E were slightly higher than the corresponding values for all other THR categories. In both the deterministic and the probabilistic analyses, category E dominated the other four categories.

Recommendations for research

1. Randomised controlled trials with adequate follow-up were not available to guide us in evaluating these interventions for this very common and important problem. Consideration should be given to setting up RCTs with long-term follow-up.
2. We were not able to link PROMs data with NJR data or with costs. This linkage, coupled with resource use data and implemented routinely, would be extremely useful for future cost-effectiveness assessments. The NJR has embedded this utility data from 2013 and these data will be reported in the future.
3. We would welcome work to validate our new findings on the relative cost-effectiveness of different combinations of prosthesis components for THR.

Acknowledgements

The authors would like to thank the NJR and the NHS Information Centre for Health and Social Care for their continuing advice and support regarding the data analysis in the report.

We would like to thank our clinical and technical advisors who have assisted in protocol development, provided methodological, policy and clinical perspectives on the data and results and critiqued the report drafts. Clinical advisors were Professor Matt Costa, Professor of Trauma and Academic Orthopaedic Surgery, University of Warwick, Professor Ashley Blom, Head of the Orthopaedic Group, University of Bristol, and Professor Alistair Hart, University College London Chair of Orthopaedic Surgery, Consultant Orthopaedic Surgeon and Director of Research & Development. The biomechanical technical advisors were Professor Richie Gill, Professor of Healthcare Engineering, University of Bath, and Dr James Meredith, Senior Teaching Fellow, Warwick Manufacturing Group, University of Warwick.

We would like to thank our methodological advisor Dr Ewen Cummins (McMDC, Glasgow) for providing health economics advice during the study, in particular in the areas of multistate models, general evidence synthesis and statistical issues in health economic modelling and the application of statistical methods to various health-care settings.

We would also like to thank Mrs Hannah Fraser for her input in assembling material and formatting the report.

Michael Crowther is funded by a NIHR Doctoral Fellowship (DRF_2012–05–409).

Contribution of authors

Aileen Clarke (Professor of Public Health and Health Services Research) managed the design and analysis and wrote the abstract, summary and discussion.

Aileen Clarke, Amy Grove (Research Project Manager), **Karoline Freeman** (Research Fellow), **Alex Tsertsvadze** (Senior Research Fellow) and **Paul Sutcliffe** (Associate Professor), co-ordinated and conducted the clinical effectiveness, cost-effectiveness and registries systematic reviews. This included screening and retrieving papers, assessing papers against inclusion criteria, appraising the quality of papers and abstracting data from papers for synthesis.

Ruth Pulikottil-Jacob (Research Fellow) performed the economic modelling with support from **Martin Connock** (Senior Research Fellow).

Amy Grove co-ordinated the project.

Amy Grove, Alex Tsertsvadze, David Metcalfe (Academic Clinical Fellow) and **Sarah Morrow** (Medical Student) wrote the background section of the report.

Hema Mistry (Associate Professor) analysed the PROMs data and provided advice on the economic model.

Martin Connock (Senior Research Fellow), **Ngianga-Bakwin Kandala** and **Michal Crowther** (Research Associate) conducted the survival analysis, analysed the data and developed transition probabilities.

Rachel Court (Information Specialist) and **Samantha Johnson** (Information Specialist) developed the search strategy and undertook the searches.

Ngianga-Bakwin Kandala (Principal Research Fellow) and **Gaurav Suri** (Research Associate) performed the database analysis.

Matthew Costa (Professor of Trauma and Orthopaedic Surgery) provided clinical comment on the draft and had input into the data analysis and formation of the categories.

All authors were involved in writing draft and final versions of the report.

Publication

Pulikottil-Jacob R, Connock M, Kandala N-B, Mistry H, Grove A, Freeman K, *et al.* Cost-effectiveness of total hip replacements for women and men with hip osteoarthritis: comparison of devices with differing bearing surfaces and modes of fixation. *Bone Joint J* 2015; in press.

Kandala N-B, Connock M, Pulikottil-Jacob R, Sutcliffe P, Crowther M, Grove A, *et al*. Total hip replacement: setting a benchmark revision rate – analysis of registry evidence. *BMJ* 2015; in press.

Tsertsvadze A, Grove A, Freeman K, Court R, Johnson S, Connock M, *et al*. Total hip replacement for the treatment of end stage arthritis of the hip: a systematic review and meta-analysis. *PLOS ONE* 2014;**9**:e99804.

References

1. Sinusas K. Osteoarthritis: diagnosis and treatment. *Am Fam Physician* 2012;**85**:49–56.
2. Metcalfe D, Harte AL, Aletrari MO, Al Daghri NM, Al Disi D, Tripathi G, *et al.* Does endotoxaemia contribute to osteoarthritis in obese patients? *Clin Sci* 2012;**123**:627–34. http://dx.doi.org/10.1042/CS20120073
3. National Institute for Health and Care Excellence. *Osteoarthritis: The Care and Management of Osteoarthritis in Adults*. Clinical guidelines CG59. London: NICE; 2008. URL: www.nice.org.uk/CG59 (accessed 18 April 2013).
4. National Institute for Health and Care Excellence. *Rheumatoid Arthritis: The Management of Rheumatoid Arthritis in Adults*. Clinical guidelines CG79. London: NICE; 2009. URL: www.nice.org.uk/CG79 (accessed 18 April 2013).
5. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO III, *et al.* 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;**69**:1580–8. http://dx.doi.org/10.1136/ard.2010.138461
6. Aulakh TS, Kuiper JH, Dixey J, Richardson JB. Hip resurfacing for rheumatoid arthritis: independent assessment of 11-year results from an international register. *Int Orthop* 2011;**35**:803–8. http://dx.doi.org/10.1007/s00264-010-1046-0
7. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G, *et al.* Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann Rheum Dis* 2010;**69**:631–7. http://dx.doi.org/10.1136/ard.2009.123919
8. Chen A, Gupte C, Akhtar K, Smith P, Cobb J. The global economic cost of osteoarthritis: how the UK compares. *Arthritis* 2012;**2012**:698709. http://dx.doi.org/10.1155/2012/698709
9. Picavet HS, Hazes JM. Prevalence of self reported musculoskeletal diseases is high. *Ann Rheum Dis* 2003;**62**:644–50. http://dx.doi.org/10.1136/ard.62.7.644
10. Sulsky SI, Carlton L, Bochmann F, Ellegast R, Glitsch U, Hartmann B, *et al.* Epidemiological evidence for work load as a risk factor for osteoarthritis of the hip: a systematic review. *PLOS ONE* 2012;**7**:e31521. http://dx.doi.org/10.1371/journal.pone.0031521
11. de Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, *et al.* A systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip. *Health Technol Assess* 2008;**12**(26). http://dx.doi.org/10.3310/hta12260
12. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;**16**:137–62. http://dx.doi.org/10.1016/j.joca.2007.12.013
13. Symmons D, Mathers C, Pfleger B. Global burden of osteoarthritis in the year 2000. In *Global Burden of Disease 2000*. Geneva: World Health Organization; 2006. URL: www.who.int/healthinfo/statistics/bod_osteoarthritis.pdf (accessed 8 September 2014).
14. Hoaglund FT, Steinbach LS. Primary osteoarthritis of the hip: etiology and epidemiology. *J Am Acad Orthop Surg* 2001;**9**:320–7.

15. Smith AJ, Dieppe P, Howard PW, Blom AW. Failure rates of metal-on-metal hip resurfacings: analysis of data from the National Joint Registry for England and Wales. *Lancet* 2012;**380**:1759–66. http://dx.doi.org/10.1016/S0140-6736(12)60989-1

16. Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW, National Joint Registry of England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet* 2012;**379**:1199–204. http://dx.doi.org/10.1016/S0140-6736(12)60353-5

17. Pereira D, Peleteiro B, Araujo J, Branco J, Santos RA, Ramos E. The effect of osteoarthritis definition on prevalence and incidence estimates: a systematic review. *Osteoarthritis Cartilage* 2011;**19**:1270–85. http://dx.doi.org/10.1016/j.joca.2011.08.009

18. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. *Lancet* 2011;**377**:2115–26. http://dx.doi.org/10.1016/S0140-6736(11)60243-2

19. Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC. A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease. *Health Technol Assess* 2002;**6**(15). http://dx.doi.org/10.3310/hta6150

20. Buckwalter JA. Osteoarthritis and articular cartilage use, disuse, and abuse: experimental studies. *J Rheumatol Suppl* 1995;**43**:13–15.

21. Sharma L, Lou C, Cahue S, Dunlop DD. The mechanism of the effect of obesity in knee osteoarthritis: the mediating role of malalignment. *Arthritis Rheum* 2000;**43**:568–75. http://dx.doi.org/10.1002/1529-0131(200003)43:3<568::AID-ANR13>3.0.CO;2-E

22. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM, *et al.* Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med* 2000;**133**:635–46. http://dx.doi.org/10.7326/0003-4819-133-8-200010170-00016

23. Kujala UM, Kaprio J, Sarna S. Osteoarthritis of weight bearing joints of lower limbs in former elite male athletes. *BMJ* 1994;**308**:231–4. http://dx.doi.org/10.1136/bmj.308.6923.231

24. Haara MM, Heliovaara M, Kroger H, Arokoski JP, Manninen P, Karkkainen A, *et al.* Osteoarthritis in the carpometacarpal joint of the thumb. Prevalence and associations with disability and mortality. *J Bone Joint Surg Am* 2004;**86–A**:1452–7.

25. National Institute for Health and Care Excellence. *Guidance on the Use of Metal on Metal Hip Resurfacing Arthroplasty*. Technology appraisal guidance 44. London: NICE; 2002. URL: http://guidance.nice.org.uk/TA44 (accessed 5 March 2013).

26. Arthritis Care. *OA Nation: The Most Comprehensive UK Report of People with Osteoarthritis*. April 2004. URL: www.arthritiscare.org.uk/PublicationsandResources/Forhealthprofessionals/OANation (accessed 18 April 2013).

27. Ahmad MA, Xypnitos FN, Giannoudis PV. Measuring hip outcomes: common scales and checklists. *Injury* 2011;**42**:259–64. http://dx.doi.org/10.1016/j.injury.2010.11.052

28. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S, *et al.* Osteoarthritis index delivered by mobile phone (m-WOMAC) is valid, reliable and responsive. *Int Med J* 2010;**40**(Suppl. 3):6.

29. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS) – validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord* 2003;**4**:10. http://dx.doi.org/10.1186/1471-2474-4-10

30. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br* 1996;**78**:185–90.

31. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am* 1969;**51**:737–55.

32. d'Aubigné RM, Postel M. Functional results of hip arthroplasty with acrylic prosthesis. *J Bone Joint Surg Am* 1954;**5**:451–75.

33. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation – value in comparison with other assessment tests. *Scand J Rheumatol Suppl* 1987;**65**:85–9. http://dx.doi.org/10.3109/03009748709102182

34. Lequesne M. Indices of severity and disease activity for osteoarthritis. *Semin Arthritis Rheum* 1991;**20**(Suppl. 2):48–54. http://dx.doi.org/10.1016/0049-0172(91)90027-W

35. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. *J Rheumatol* 1997;**24**:779–81.

36. National Joint Registry. *National Joint Registry for England and Wales: 8th Annual Report*. Hemel Hempstead: NJR Centre; 2011.

37. Jenkins PJ, Clement ND, Hamilton DF, Gaston PF, Patton JT, Howie CR. Predicting the cost-effectiveness of total hip and knee replacement: a health economic analysis. *Bone Joint J* 2013;**95–B**:115–21. http://dx.doi.org/10.1302/0301-620X.95B1.29835

38. Briggs AF, Sculpher MF, Dawson JF, Fitzpatrick RF, Murray DF, Malchau H. The use of probabilistic decision models in technology assessment: the case of total hip replacement. *Appl Health Econ Health Policy* 2004;**3**:79–89. http://dx.doi.org/10.2165/00148365-200403020-00004

39. Morshed S, Bozic KJ, Ries MD, Malchau H, Colford JM Jr. Comparison of cemented and uncemented fixation in total hip replacement: a meta-analysis. *Acta Orthop* 2007;**78**:315–26. http://dx.doi.org/10.1080/17453670710013861

40. Edlin R, Tubeuf S, Achten J, Parsons N, Costa M. Cost-effectiveness of total hip arthroplasty versus resurfacing arthroplasty: economic evaluation alongside a clinical trial. *BMJ Open* 2012;**2**:e001162. http://dx.doi.org/10.1136/bmjopen-2012-001162

41. Pivec R, Johnson K, Mears S, Mont MA. Hip arthroplasty. *Lancet* 2012;**380**:1768–77. http://dx.doi.org/10.1016/S0140-6736(12)60607-2

42. South East Public Health Observatory. *NHS Atlas of Variation in Healthcare 2010*. 2010. URL: www.sepho.org.uk/extras/maps/NHSatlas/atlas.html (accessed 15 December 2012).

43. Davies C, Lorgelly P, Shemilt I, Mugford M, Tucker K, Macgregor A. Can choices between alternative hip prostheses be evidence based? A review of the economic evaluation literature. *Cost Eff Resour Alloc* 2010;**8**:20. http://dx.doi.org/10.1186/1478-7547-8-20

44. Pennington M, Grieve R, Sekhon JS, Gregg P, Black N, Van Der Meulen JH. Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis. *BMJ* 2013;**346**:f1026. http://dx.doi.org/10.1136/bmj.f1026

45. Appleby J. *Which English Hospital is Best at Hips?* London: The King's Fund; September 2010. URL: www.kingsfund.org.uk/blog/2010/09/which-english-hospital-best-hips (accessed 22 November 2012).

46. National Institute for Health and Care Excellence. *Guidance on the Selection of Prostheses for Primary Total Hip Replacement. Technology appraisal guidance 2*. London: NICE; 2000. URL: www.nice.org.uk/TA2 (accessed 18 April 2013).

47. Medicines and Healthcare Products Regulatory Agency. *Medical Device Alert: All Metal-on-Metal (MoM) Hip Replacements (MDA/2012/036)*. June 2012. URL: www.mhra.gov.uk/Publications/Safetywarnings/MedicalDeviceAlerts/CON155761 (accessed 18 April 2013).

48. National Joint Registry. *National Joint Registry for England and Wales: 9th Annual Report*. Hemel Hempstead: NJR Centre; 2012. URL: www.njrcentre.org.uk/NjrCentre/Portals/0/Documents/England/Reports/9th_annual_report/NJR%209th%20Annual%20Report%202012.pdf (accessed 18 April 2013).

49. Smith MA, Smith WT. The American Joint Replacement Registry. *Orthop Nurs* 2012;**31**:296–9. http://dx.doi.org/10.1097/NOR.0b013e31826649b6

50. Epinette JA, Manley MT. Uncemented stems in hip replacement – hydroxyapatite or plain porous: does it matter? Based on a prospective study of HA Omnifit stems at 15-years minimum follow-up. *Hip Int* 2008;**18**:69–74.

51. Martell JM, Verner JJ, Incavo SJ. Clinical performance of a highly cross-linked polyethylene at two years in total hip arthroplasty: a randomized prospective trial. *J Arthroplasty* 2003;**18**(Suppl. 1):55–9. http://dx.doi.org/10.1016/S0883-5403(03)00341-3

52. Dorr LD, Wan Z, Shahrdar C, Sirianni L, Boutary M, Yun A. Clinical performance of a Durasul highly cross-linked polyethylene acetabular liner for total hip arthroplasty at five years. *J Bone Joint Surg Am* 2005;**87**:1816–21. http://dx.doi.org/10.2106/JBJS.D.01915

53. Smith SL, Dowson D, Goldsmith AA. The effect of femoral head diameter upon lubrication and wear of metal-on-metal total hip replacements. *Proc Inst Mech Eng H* 2001;**215**:161–70. http://dx.doi.org/10.1243/0954411011533724

54. Dowson D, Hardaker C, Flett M, Isaac GH. A hip joint simulator study of the performance of metal-on-metal joints: part II: design. *J Arthroplasty* 2004;**19**(Suppl. 3):124–30.

55. Langton DJ, Jameson SS, Joyce TJ, Gandhi JN, Sidaginamale R, Mereddy P, *et al.* Accelerating failure rate of the ASR total hip replacement. *J Bone Joint Surg Br* 2011;**93**:1011–16. http://dx.doi.org/10.1302/0301-620X.93B8.26040

56. DePuy Synthes. *ASR™ Hip Recall Guide for Patients*. 2013. URL: http://asrrecall.depuy.com/ukpatient (accessed 14 May 2013).

57. Davies AP, Willert HG, Campbell PA, Learmonth ID, Case CP. An unusual lymphocytic perivascular infiltration in tissues around contemporary metal-on-metal joint replacements. *J Bone Joint Surg Am* 2005;**87**:18–27. http://dx.doi.org/10.2106/JBJS.C.00949

58. Yan Y, Neville A, Dowson D. Understanding the role of corrosion in the degradation of metal-on-metal implants. *Proc Inst Mech Eng H* 2006;**220**:173–81. http://dx.doi.org/10.1243/095441105X63246

59. Doorn PF, Campbell PA, Worrall J, Benya PD, McKellop HA, Amstutz HC. Metal wear particle characterization from metal on metal total hip replacements: transmission electron microscopy study of periprosthetic tissues and isolated particles. *J Biomed Mater Res* 1998;**42**:103–11. http://dx.doi.org/10.1002/(SICI)1097-4636(199810)42:1<103::AID-JBM13>3.0.CO;2-M

60. Amstutz HC, Grigoris P. Metal on metal bearings in hip arthroplasty. *Clin Orthop Related Res* 1996;**329**(Suppl.):S11–34. http://dx.doi.org/10.1097/00003086-199608001-00003

61. Basle MF, Bertrand G, Guyetant S, Chappard D, Lesourd M. Migration of metal and polyethylene particles from articular prostheses may generate lymphadenopathy with histiocytosis. *J Biomed Mater Res* 1996;**30**:157–63. http://dx.doi.org/10.1002/(SICI)1097-4636(199602)30:2<157::AID-JBM4>3.0.CO;2-Q

62. Hallab NJ, Anderson S, Caicedo M, Skipor A, Campbell P, Jacobs JJ. Immune responses correlate with serum-metal in metal-on-metal hip arthroplasty. *J Arthroplasty* 2004;**19**(Suppl. 3):88–93. http://dx.doi.org/10.1016/j.arth.2004.09.012

63. Pandit H, Glyn-Jones S, McLardy-Smith P, Gundle R, Whitwell D, Gibbons CL, *et al.* Pseudotumours associated with metal-on-metal hip resurfacings. *J Bone Joint Surg Br* 2008;**90**:847–51. http://dx.doi.org/10.1302/0301-620X.90B7.20213

64. Case CP, Langkamer VG, James C, Palmer MR, Kemp AJ, Heap PF, *et al.* Widespread dissemination of metal debris from implants. *J Bone Joint Surg Br* 1994;**76**:701–12.

65. Keegan GM, Learmonth ID, Case CP. A systematic comparison of the actual, potential, and theoretical health effects of cobalt and chromium exposures from industry and surgical implants. *Crit Rev Toxicol* 2008;**38**:645–74. http://dx.doi.org/10.1080/10408440701845534

66. Parry MC, Bhabra G, Sood A, Machado F, Cartwright L, Saunders M, *et al.* Thresholds for indirect DNA damage across cellular barriers for orthopaedic biomaterials. *Biomaterials* 2010;**31**:4477–83. http://dx.doi.org/10.1016/j.biomaterials.2010.02.038

67. Daley B, Doherty AT, Fairman B, Case CP. Wear debris from hip or knee replacements causes chromosomal damage in human cells in tissue culture. *J Bone Joint Surg Br* 2004;**86**:598–606.

68. Mäkelä KT, Visuri T, Pulkkinen P, Eskelinen A, Remes V, Virolainen P, *et al.* Risk of cancer with metal-on-metal hip replacements: population based study. *BMJ* 2012;**345**:e4646. http://dx.doi.org/10.1136/bmj.e4646

69. Smith AJ, Dieppe P, Porter M, Blom AW, National Joint Registry of England and Wales. Risk of cancer in first seven years after metal-on-metal hip replacement compared with other bearings and general population: linkage study between the National Joint Registry of England and Wales and hospital episode statistics. *BMJ* 2012;**344**:e2383. http://dx.doi.org/10.1136/bmj.e2383

70. Food and Drug Administration. *Safety Communication: Metal-on Metal Hip Implants*. Issued 17 January 2013. URL: www.fda.gov/MedicalDevices/Safety/AlertsandNotices/ucm335775.htm (accessed 12 May 2014).

71. British Orthopaedic Association and British Hip Society. *Metal-on-Metal, Guidance and Information to Surgeons*. 2013. URL: www.njrcentre.org.uk/njrcentre/metalonmetalhipimplants/tabid/237/default.aspx (accessed 2014).

72. Bozic KJ, Ong K, Lau E, Kurtz SM, Vail TP, Rubash HE, *et al.* Risk of complication and revision total hip arthroplasty among Medicare patients with different bearing surfaces. *Clin Orthop Relat Res* 2010;**468**:2357–62. http://dx.doi.org/10.1007/s11999-010-1262-3

73. Health Net. *National Medical Policy: Hip Resurfacing*. March 2013. URL: www.healthnet.com/static/general/unprotected/pdfs/national/policies/HipResurfacing.pdf (accessed 15 May 2013).

74. Johnson AJ, Zywiel MG, Hooper H, Mont MA. Narrowed indications improve outcomes for hip resurfacing arthroplasty. *Bull NYU Hosp Jt Dis* 2011;**69**(Suppl. 1):S27–9.

75. Havelin LI. The Norwegian Joint Registry. *Bull Hosp Jt Dis* 1999;**58**:139–47.

76. Bozic KJ, Kurtz SM, Lau E, Ong K, Vail TP, Berry DJ. The epidemiology of revision total hip arthroplasty in the United States. *J Bone Joint Surg Am* 2009;**91**:128–33. http://dx.doi.org/10.2106/JBJS.H.00155

77. Jafari SM, Coyle C, Mortazavi SM, Sharkey PF, Parvizi J. Revision hip arthroplasty: infection is the most common cause of failure. *Clin Orthop Related Res* 2010;**468**:2046–51. http://dx.doi.org/10.1007/s11999-010-1251-6

78. Ebramzadeh E, Campbell PA, Takamura KM, Lu Z, Sangiorgio SN, Kalma JJ, *et al.* Failure modes of 433 metal-on-metal hip implants: how, why, and wear. *Orthop Clin North Am* 2011;**42**:241–50. http://dx.doi.org/10.1016/j.ocl.2011.01.001

79. De Haan R, Pattyn C, Gill HS, Murray DW, Campbell PA, De Smet K. Correlation between inclination of the acetabular component and metal ion levels in metal-on-metal hip resurfacing replacement. *J Bone Joint Surg Br* 2008;**90**:1291–7. http://dx.doi.org/10.1302/0301-620X.90B10.20533

80. Parvizi J, Kim KI, Goldberg G, Mallo G, Hozack WJ. Recurrent instability after total hip arthroplasty: beware of subtle component malpositioning. *Clin Orthop Relat Res* 2006;**447**:60–5. http://dx.doi.org/10.1097/01.blo.0000218749.37860.7c

81. Parvizi J, Wade FA, Rapuri V, Springer BD, Berry DJ, Hozack WJ. Revision hip arthroplasty for late instability secondary to polyethylene wear. *Clin Orthop Relat Res* 2006;**447**:66–9. http://dx.doi.org/10.1097/01.blo.0000218751.14989.a6

82. Restrepo C, Mortazavi SM, Brothers J, Parvizi J, Rothman RH. Hip dislocation: are hip precautions necessary in anterior approaches? *Clin Orthop Relat Res* 2011;**469**:417–22. http://dx.doi.org/10.1007/s11999-010-1668-y

83. Glyn-Jones S, McLardy-Smith P, Gill HS, Murray DW. The creep and wear of highly cross-linked polyethylene: a three-year randomised, controlled trial using radiostereometric analysis. *J Bone Joint Surg Br* 2008;**90**:556–61. http://dx.doi.org/10.1302/0301-620X.90B5.20545

84. Iannotti JP, Balderston RA, Booth RE, Rothman RH, Cohn JC, Pickens G. Aseptic loosening after total hip arthroplasty. Incidence, clinical significance, and etiology. *J Arthroplasty* 1986;**1**:99–107. http://dx.doi.org/10.1016/S0883-5403(86)80047-X

85. Callanan MC, Jarrett B, Bragdon CR, Zurakowski D, Rubash HE, Freiberg AA, *et al.* The John Charnley Award: risk factors for cup malpositioning: quality improvement through a joint registry at a tertiary hospital. *Clin Orthop Relat Res* 2011;**469**:319–29. http://dx.doi.org/10.1007/s11999-010-1487-1

86. Hartman CW, Garvin KL. Femoral fixation in revision total hip arthroplasty. *J Bone Joint Surg Am* 2011;**93**:2311–22. http://dx.doi.org/10.2106/JBJS.9324icl

87. Lindahl H, Garellick G, Regner H, Herberts P, Malchau H. Three hundred and twenty-one periprosthetic femoral fractures. *J Bone Joint Surg Am* 2006;**88**:1215–22. http://dx.doi.org/10.2106/JBJS.E.00457

88. Sarvilinna R, Huhtala HS, Sovelius RT, Halonen PJ, Nevalainen JK, Pajamaki KJ. Factors predisposing to periprosthetic fracture after hip arthroplasty: a case ($n = 31$)–control study. *Acta Orthop Scand* 2004;**75**:16–20. http://dx.doi.org/10.1080/00016470410001708030

89. Clift B. Periprosthetic fracture of the femur. *J Bone Joint Surg Am* 2000;**82**:446–7.

90. Ogawa H, Ito Y, Takigami I, Shimizu K. Revision total hip arthroplasty for a Vancouver type B3 periprosthetic fracture using an allograft-cemented stem composite by the telescoping technique. *J Arthroplasty* 2011;**26**:665–8. http://dx.doi.org/10.1016/j.arth.2010.05.012

91. Urquhart DM, Hanna FS, Brennan SL, Wluka AE, Leder K, Cameron PA, *et al.* Incidence and risk factors for deep surgical site infection after primary total hip arthroplasty: a systematic review. *J Arthroplasty* 2010;**25**:1216–22. http://dx.doi.org/10.1016/j.arth.2009.08.011

92. Darwiche H, Barsoum WK, Klika A, Krebs VE, Molloy R. Retrospective analysis of infection rate after early reoperation in total hip arthroplasty. *Clin Orthop Relat Res* 2010;**468**:2392–6. http://dx.doi.org/10.1007/s11999-010-1325-5

93. Bottner F, Su E, Nestor B, Azzis B, Sculco TP, Bostrom M. Radiostereometric analysis: the hip. *HSS J* 2005;**1**:94–9. http://dx.doi.org/10.1007/s11420-005-0114-2

94. Lavernia CJ. Cost-effectiveness of early surgical intervention in silent osteolysis. *J Arthroplasty* 1998;**13**:277–9. http://dx.doi.org/10.1016/S0883-5403(98)90172-3

95. Australian Orthopaedic Association National Joint Replacement Registry. *Australian Orthopaedic Association National Joint Replacement Registry: 2012 Annual Report*. February 2013. URL: https://aoanjrr.dmac.adelaide.edu.au/annual-reports-2012 (accessed 18 April 2013).

96. Garellick G, Kärrholm J, Rogmark C, Herberts P. *Swedish Hip Arthroplasty Register: Annual Report 2010*. October 2011. URL: www.shpr.se/Libraries/Documents/AnnualReport-2010–2-eng.sflb.ashx (accessed 18 April 2013).

97. Health and Social Care Information Centre. *Patient Reported Outcome Measures (PROMs)*. 2013. URL: www.hscic.gov.uk/proms (accessed 16 May 2013).

98. NHS Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. CRD report no. 4. York: NHS Centre for Reviews and Dissemination, University of York; 1999.

99. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535. http://dx.doi.org/10.1136/bmj.b2535

100. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;**343**:d5928. http://dx.doi.org/10.1136/bmj.d5928

101. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, *et al.* Development of AMSTAR: a measurement tool to assess the methodological quality of systematic review. *BMC Med Res Methodol* 2007;**7**:10. http://dx.doi.org/10.1186/1471-2288-7-10

102. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, *et al.* GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings. *J Clin Epidemiol* 2011;**64**:383–94. http://dx.doi.org/10.1016/j.jclinepi.2010.04.026

103. Khan KS, Ter Riet G, Glanville J, Snowdon AJ, Kleijnen J. *Undertaking Systematic Reviews of Research on Effectiveness. CRD's Guidance for Carrying Out or Commissioning Reviews*, 2nd edn. CRD report no. 4. York: NHS Centre for Reviews and Dissemination, University of York; 2001.

104. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;**7**:177–88. http://dx.doi.org/10.1016/0197-2456(86)90046-2

105. The Cochrane Collaboration. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. March 2011. URL: http://handbook.cochrane.org (accessed 24 April 2013).

106. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34. http://dx.doi.org/10.1136/bmj.315.7109.629

107. Achten J, Parsons NR, Edlin RP, Griffin DR, Costa ML. A randomised controlled trial of total hip arthroplasty versus resurfacing arthroplasty in the treatment of young patients with arthritis of the hip joint. *BMC Musculoskelet Disord* 2010;**11**:8. http://dx.doi.org/10.1186/1471-2474-11-8

108. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, *et al.* Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. *Ann Rheum Dis* 2005;**64**:29–33. http://dx.doi.org/10.1136/ard.2004.022905

109. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;**14**:1523–32. http://dx.doi.org/10.1007/s11136-004-7713-0

110. Bjørgul K, Novicoff WM, Andersen ST, Brevig K, Thu F, Wiig M, *et al.* No differences in outcomes between cemented and uncemented acetabular components after 12–14 years: results from a randomized controlled trial comparing Duraloc with Charnley cups. *J Orthop Traumatol* 2010;**11**:37–45. http://dx.doi.org/10.1007/s10195-010-0082-2

111. Bjørgul K, Novicoff WM, Andersen ST, Brevig K, Thu F, Wiig M, *et al.* The Charnley stem: clinical, radiological and survival data after 11–14 years. *Orthop Traumatol Surg Res* 2010;**96**:97–103. http://dx.doi.org/10.1016/j.otsr.2009.11.002

112. Angadi DS, Brown S, Crawfurd EJ. Cemented polyethylene and cementless porous-coated acetabular components have similar outcomes at a mean of seven years after total hip replacement: a prospective randomised study. *J Bone Joint Surg Br* 2012;**94**:1604–10. http://dx.doi.org/10.1302/0301-620X.94B12.28060

113. Engh CA Jr, Hopper RH Jr, Huynh C, Ho H, Sritulanondha S, Engh CA Sr. A prospective, randomized study of cross-linked and non-cross-linked polyethylene for total hip arthroplasty at 10-year follow-up. *J Arthroplasty* 2012;**27**(Suppl.):2–7. http://dx.doi.org/10.1016/j.arth.2012.03.048

114. Engh CA Jr, Stepniewski AS, Ginn SD, Beykirch SE, Sychterz-Terefenko CJ, Hopper RH Jr, *et al.* A randomized prospective evaluation of outcomes after total hip arthroplasty using cross-linked marathon and non-cross-linked Enduron polyethylene liners. *J Arthroplasty* 2006;**21**(Suppl. 2):17–25. http://dx.doi.org/10.1016/j.arth.2006.05.002

115. Capello WN, D'Antonio JA, Feinberg JR, Manley MT, Naughton M. Ceramic-on-ceramic total hip arthroplasty: update. *J Arthroplasty* 2008;**23**(Suppl.):39–43. http://dx.doi.org/10.1016/j.arth.2008.06.003

116. D'Antonio J, Capello W, Manley M, Naughton M, Sutton K. Alumina ceramic bearings for total hip arthroplasty: five-year results of a prospective randomized study. *Clin Orthop Relat Res* 2005;**436**:164–71. http://dx.doi.org/10.1097/01.blo.0000162995.50971.39

117. D'Antonio J, Capello W, Manley M. Alumina ceramic bearings for total hip arthroplasty. *Orthopedics* 2003;**26**:39–46.

118. Mesko JW, D'Antonio JA, Capello WN, Bierbaum BE, Naughton M. Ceramic-on-ceramic hip outcome at a 5- to 10-year interval. *J Arthroplasty* 2011;**26**:172–7. http://dx.doi.org/10.1016/j.arth.2010.04.029

119. Corten K, Bourne RB, Charron KD, Au K, Rorabeck CH. Comparison of total hip arthroplasty performed with and without cement: a randomized trial. A concise follow-up, at twenty years, of previous reports. *J Bone Joint Surg Am* 2011;**93**:1335–8. http://dx.doi.org/10.2106/JBJS.J.00448

120. Laupacis A, Bourne R, Rorabeck C, Feeny D, Tugwell P, Wong C. Comparison of total hip arthroplasty performed with and without cement: a randomized trial. *J Bone Joint Surg Am* 2002;**84–A**:1823–8.

121. Bourne RB, Corten K. Cemented versus cementless stems: a verdict is in. *Orthopedics* 2010;**33**:638.

122. Corten K, Bourne RB, Charron KD, Au K, Rorabeck CH. What works best, a cemented or cementless primary total hip arthroplasty?: minimum 17-year followup of a randomized controlled trial. *Clin Orthop Relat Res* 2011;**469**:209–17. http://dx.doi.org/10.1007/s11999-010-1459-5

123. Howie DW, Holubowycz OT, Middleton R, Large Articulation Study Group. Large femoral heads decrease the incidence of dislocation after total hip arthroplasty: a randomized controlled trial. *J Bone Joint Surg Am* 2012;**94**:1095–102. http://dx.doi.org/10.2106/JBJS.K.00570

124. Lewis PM, Moore CA, Olsen M, Schemitsch EH, Waddell JP. Comparison of mid-term clinical outcomes after primary total hip arthroplasty with oxinium vs cobalt chrome femoral heads. *Orthopedics* 2008;**31**(Suppl. 2):114–22.

125. Amanatullah DF, Landa J, Strauss EJ, Garino JP, Kim SH, Di Cesare PE. Comparison of surgical outcomes and implant wear between ceramic–ceramic and ceramic–polyethylene articulations in total hip arthroplasty. *J Arthroplasty* 2011;**26**(Suppl.):72–7. http://dx.doi.org/10.1016/j.arth.2011.04.032

126. Kadar T, Hallan G, Aamodt A, Indrekvam K, Badawy M, Skredderstuen A, *et al.* Wear and migration of highly cross-linked and conventional cemented polyethylene cups with cobalt chrome or oxinium femoral heads: a randomized radiostereometric study of 150 patients. *J Orthop Res* 2011;**29**:1222–9. http://dx.doi.org/10.1002/jor.21389

127. Healy WL, Tilzey JF, Iorio R, Specht LM, Sharma S. Prospective, randomized comparison of cobalt-chrome and titanium trilock femoral stems. *J Arthroplasty* 2009;**24**:831–6. http://dx.doi.org/10.1016/j.arth.2008.06.035

128. Kim YH, Choi Y, Kim JS. Comparison of bone mineral density changes around short, metaphyseal-fitting, and conventional cementless anatomical femoral components. *J Arthroplasty* 2011;**26**:931–40. http://dx.doi.org/10.1016/j.arth.2010.10.001

129. Kim YH, Kim JS, Park JW, Joo JH. Comparison of total hip replacement with and without cement in patients younger than 50 years of age: the results at 18 years. *J Bone Joint Surg Br* 2011;**93**:449–55. http://dx.doi.org/10.1302/0301-620X.93B4.26149

130. Costa ML, Achten J, Parsons NR, Edlin RP, Foguet P, Prakash U, *et al.* Total hip arthroplasty versus resurfacing arthroplasty in the treatment of patients with arthritis of the hip joint: single centre, parallel group, assessor blinded, randomised controlled trial. *BMJ* 2012;**344**:e2147. http://dx.doi.org/10.1136/bmj.e2147

131. Garbuz DS, Tanzer M, Greidanus NV, Masri BA, Duncan CP. The John Charnley Award: metal-on-metal hip resurfacing versus large-diameter head metal-on-metal total hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res* 2010;**468**:318–25. http://dx.doi.org/10.1007/s11999-009-1029-x

132. Vendittoli PA, Ganapathi M, Roy AG, Lusignan D, Lavigne M. A comparison of clinical results of hip resurfacing arthroplasty and 28 mm metal on metal total hip arthroplasty: a randomised trial with 3–6 years follow-up. *Hip Int* 2010;**20**:1–13.

133. Vendittoli PA, Lavigne M, Girard J, Roy AG. A randomised study comparing resection of acetabular bone at resurfacing and total hip replacement. *J Bone Joint Surg Br* 2006;**88**:997–1002. http://dx.doi.org/10.1302/0301-620X.88B8.17615

134. Girard J, Lavigne M, Vendittoli PA, Roy AG. Biomechanical reconstruction of the hip: a randomised study comparing total hip resurfacing and total hip arthroplasty. *J Bone Joint Surg Br* 2006;**88**:721–6. http://dx.doi.org/10.1302/0301-620X.88B6.17447

135. Rama KR, Vendittoli PA, Ganapathi M, Borgmann R, Roy A, Lavigne M. Heterotopic ossification after surface replacement arthroplasty and total hip arthroplasty: a randomized study. *J Arthroplasty* 2009;**24**:256–62. http://dx.doi.org/10.1016/j.arth.2007.12.004

136. Vendittoli PA, Lavigne M, Roy AG, Lusignan D. A prospective randomized clinical trial comparing metal-on-metal total hip arthroplasty and metal-on-metal total hip resurfacing in patients less than 65 years old. *Hip Int* 2006;**16**(Suppl. 4):73–81.

137. Voigt JD, Mosier MC. Cemented all-polyethylene acetabular implants vs other forms of acetabular fixation: a systematic review and meta-analysis of randomized controlled trials. *J Arthroplasty* 2012;**27**:1544–53. http://dx.doi.org/10.1016/j.arth.2011.12.008

138. Pakvis D, van Hellemondt G, de Visser E, Jacobs W, Spruit M. Is there evidence for a superior method of socket fixation in hip arthroplasty? A systematic review. *Int Orthop* 2011;**35**:1109–18. http://dx.doi.org/10.1007/s00264-011-1234-6

139. Clement ND, Biant LC, Breusch SJ. Total hip arthroplasty: to cement or not to cement the acetabular socket? A critical review of the literature. *Arch Orthop Trauma Surg* 2012;**132**:411–27. http://dx.doi.org/10.1007/s00402-011-1422-2

140. Sedrakyan A, Normand S-L, Dabic S, Jacobs S, Graves S, Marinac-Dabic D. Comparative assessment of implantable hip devices with different bearing surfaces: systematic appraisal of evidence. *BMJ* 2011;**343**:d7434. http://dx.doi.org/10.1136/bmj.d7434

141. Yoshitomi H, Shikata S, Ito H, Nakayama T, Nakamura T. Manufacturers affect clinical results of THA with zirconia heads: a systematic review. *Clin Orthop Relat Res* 2009;**467**:2349–55. http://dx.doi.org/10.1007/s11999-009-0709-x

142. Jiang Y, Zhang K, Die J, Shi Z, Zhao H, Wang K. A systematic review of modern metal-on-metal total hip resurfacing vs standard total hip arthroplasty in active young patients. *J Arthroplasty* 2011;**26**:419–26. http://dx.doi.org/10.1016/j.arth.2010.07.008

143. Smith TO, Nichols R, Donell ST, Hing CB. The clinical and radiological outcomes of hip resurfacing versus total hip arthroplasty: a meta-analysis and systematic review. *Acta Orthop* 2010;**81**:684–95. http://dx.doi.org/10.3109/17453674.2010.533933

144. Springer BD, Connelly SE, Odum SM, Fehring TK, Griffin WL, Mason JB, *et al.* Cementless femoral components in young patients: review and meta-analysis of total hip arthroplasty and hip resurfacing. *J Arthroplasty* 2009;**24**:2–8. http://dx.doi.org/10.1016/j.arth.2009.04.032

145. McCalden RW, MacDonald SJ, Rorabeck CH, Bourne RB, Chess DG, Charron KD. Wear rate of highly cross-linked polyethylene in total hip arthroplasty. A randomized controlled trial. *J Bone Joint Surg Am* 2009;**91**:773–82. http://dx.doi.org/10.2106/JBJS.H.00244

146. Bhan S, Pankaj A, Malhotra R. One- or two-stage bilateral total hip arthroplasty: a prospective, randomised, controlled study in an Asian population. *J Bone Joint Surg Br* 2006;**88**:298–303. http://dx.doi.org/10.1302/0301-620X.88B3.17048

147. Conroy JL, Chawda M, Kaushal R, Whitehouse SL, Crawford RW, English H. Does use of a 'rim cutter' improve quality of cementation of the acetabular component of cemented exeter total hip arthroplasty? *J Arthroplasty* 2009;**24**:71–6. http://dx.doi.org/10.1016/j.arth.2008.01.130

148. Coyle D, Coyle K, Vale L, Verteuil R, Imamura M, Glazener C, *et al. Minimally Invasive Arthroplasty in the Management of Hip Arthritic Disease: Systematic Review and Economic Evaluation*. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health (CADTH); 2008.

149. D'Arrigo C, Speranza A, Monaco E, Carcangiu A, Ferretti A. Learning curve in tissue sparing total hip replacement: comparison between different approaches. *J Orthop Traumatol* 2009;**10**:47–54. http://dx.doi.org/10.1007/s10195-008-0043-1

150. Dorr LD, Maheshwari AV, Long WT, Wan Z, Sirianni LE. Early pain relief and function after posterior minimally invasive and conventional total hip arthroplasty. A prospective, randomized, blinded study. *J Bone Joint Surg Am* 2007;**89**:1153–60. http://dx.doi.org/10.2106/JBJS.F.00940

151. Dutka J, Sosin P, Libura M, Skowronek P. Total hip arthroplasty through a minimally invasive lateral approach – our experience and early results. *Ortop Traumatol Rehabil* 2007;**9**:39–45.

152. Flivik G, Kristiansson I, Kesteris U, Ryd L. Is removal of subchondral bone plate advantageous in cemented cup fixation? A randomized RSA study. *Clin Orthop Relat Res* 2006;**448**:164–72. http://dx.doi.org/10.1097/01.blo.0000203479.27044.d3

153. Flivik G. Fixation of the cemented acetabular component in hip arthroplasty. *Acta Orthop* 2005;**76**(Suppl.):3–30. http://dx.doi.org/10.1080/03008820510040685

154. Foucher KC, Wimmer MA, Moisio KC, Hildebrand M, Berli MC, Walker MR, *et al.* Time course and extent of functional recovery during the first postoperative year after minimally invasive total hip arthroplasty with two different surgical approaches – a randomized controlled trial. *J Biomech* 2011;**44**:372–8. http://dx.doi.org/10.1016/j.jbiomech.2010.10.026

155. Goosen JHM, Kollen BJ, Castelein RM, Kuipers BM, Verheyen CC. Minimally invasive versus classic procedures in total hip arthroplasty: a double-blind randomized controlled trial. *Clin Orthop Relat Res* 2011;**469**:200–8. http://dx.doi.org/10.1007/s11999-010-1331-7

156. Honl M, Dierk O, Gauck C, Carrero V, Lampe F, Dries S, *et al.* Comparison of robotic-assisted and manual implantation of a primary total hip replacement. A prospective study. *J Bone Joint Surg Am* 2003;**85–A**:1470–8.

157. Jolles BM, Bogoch ER. Posterior versus lateral surgical approach for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database Syst Rev* 2006;**3**:CD003828.

158. Jolles BM, Bogoch ER. Surgical approach for total hip arthroplasty: direct lateral or posterior? *J Rheumatol* 2004;**31**:1790–6.

159. Khan RJ, Maor D, Hofmann M, Haebich S. A comparison of a less invasive piriformis-sparing approach versus the standard posterior approach to the hip: a randomised controlled trial. *J Bone Joint Surg Br* 2012;**94**:43–50. http://dx.doi.org/10.1302/0301-620X.94B1.27001

160. Kim Y-H. Comparison of primary total hip arthroplasties performed with a minimally invasive technique or a standard technique. A prospective and randomized study. *J Arthroplasty* 2006;**21**:1092–8. http://dx.doi.org/10.1016/j.arth.2006.01.015

161. Speranza A, Iorio R, Ferretti M, D'Arrigo C, Ferretti A. A lateral minimal-incision technique in total hip replacement: a prospective, randomizes, controlled trial. *Hip Int* 2007;**17**:4–8.

162. Tanavalee A, Jaruwannapong S, Yuktanandana P, Itiravivong P. Early outcomes following minimally invasive total hip arthroplasty using a two-incision approach versus a mini-posterior approach. *Hip Int* 2006;**16**(Suppl. 4):17–22.

163. Tang Z. *Minimally Invasive Total Hip Replacement*. No. 4. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment (CCOHTA); 2004.

164. Haverkamp D, Van Den Bekerom MPJ, Harmse I, Schafroth MU. One stage bilateral total hip arthroplasty, is it safe? A meta-analysis. *Hip Int* 2010;**20**:440–6.

165. Beckmann J, Stengel D, Tingart M, Gotz J, Grifka J, Luring C. Navigated cup implantation in hip arthroplasty: a meta-analysis. *Acta Orthop* 2009;**80**:538–44. http://dx.doi.org/10.3109/17453670903350073

166. Li N, Deng Y, Chen L. Comparison of complications in single-incision minimally invasive THA and conventional THA. *Orthopedics* 2012;**35**:e1152–8. http://dx.doi.org/10.3928/01477447-20120725-12

167. Martin R, Clayson PE, Troussel S, Fraser BP, Docquier PL. Anterolateral minimally invasive total hip arthroplasty: a prospective randomized controlled study with a follow-up of 1 year. *J Arthroplasty* 2011;**26**:1362–72. http://dx.doi.org/10.1016/j.arth.2010.11.016

168. Mayr E, Nogler M, Benedetti MG, Kessler O, Reinthaler A, Krismer M, *et al.* A prospective randomized assessment of earlier functional recovery in THA patients treated by minimally invasive direct anterior approach: a gait analysis study. *Clin Biomech* 2009;**24**:812–18. http://dx.doi.org/10.1016/j.clinbiomech.2009.07.010

169. Mazoochian F, Weber P, Schramm S, Utzschneider S, Fottner A, Jansson V. Minimally invasive total hip arthroplasty: a randomized controlled prospective trial. *Arch Orthop Trauma Surg* 2009;**129**:1633–9. http://dx.doi.org/10.1007/s00402-009-0870-4

170. Muller M, Schwachmeyer V, Tohtz S, Tohtz S, Duda GN, Perka C, *et al.* The direct lateral approach: impact on gait patterns, foot progression angle and pain in comparison with a minimally invasive anterolateral approach. *Arch Orthop Trauma Surg* 2012;**132**:725–31. http://dx.doi.org/10.1007/s00402-012-1467-x

171. Muller M, Tohtz S, Springer I, Dewey M, Perka C. Randomized controlled trial of abductor muscle damage in relation to the surgical approach for primary total hip replacement: minimally invasive anterolateral versus modified direct lateral approach. *Arch Orthop Trauma Surg* 2011;**131**:179–89. http://dx.doi.org/10.1007/s00402-010-1117-0

172. Nakamura N, Sugano N, Nishii T, Kakimoto A, Miki H. A comparison between robotic-assisted and manual implantation of cementless total hip arthroplasty. *Clin Orthop Relat Res* 2010;**468**:1072–81. http://dx.doi.org/10.1007/s11999-009-1158-2

173. Ogonda L, Wilson R, Archbold P, Lawlor M, Humphreys P, O'Brien S, *et al.* A minimal-incision technique in total hip arthroplasty does not improve early postoperative outcomes. A prospective, randomized, controlled trial. *J Bone Joint Surg Am* 2005;**87**:701–10. http://dx.doi.org/10.2106/JBJS.D.02645

174. Pakvis D, Luites J, Hellemondt G, Spruit M. A cementless, elastic press-fit socket with and without screws. *Acta Orthop* 2012;**83**:481–7. http://dx.doi.org/10.3109/17453674.2012.720116

175. Parvizi J, Tarity TD, Sheikh E, Sharkey PF, Hozack WJ, Rothman RH. Bilateral total hip arthroplasty: one-stage versus two-stage procedures. *Clin Orthop Relat Res* 2006;**453**:137–41. http://dx.doi.org/10.1097/01.blo.0000246529.14135.2b

176. Pospischill M, Kranzl A, Attwenger B, Knahr K. Minimally invasive compared with traditional transgluteal approach for total hip arthroplasty: a comparative gait analysis. *J Bone Joint Surg Am* 2010;**92**:328–37. http://dx.doi.org/10.2106/JBJS.H.01086

177. Reininga IH, Wagenmakers R, van den Akker-Scheek I, Stant AD, Groothoff JW, Bulstra SK, *et al.* Effectiveness of computer-navigated minimally invasive total hip surgery compared to conventional total hip arthroplasty: design of a randomized controlled trial. *BMC Musculoskelet Disord* 2007;**8**:4. http://dx.doi.org/10.1186/1471-2474-8-4

178. Reininga IHF, Zijlstra W, Wagenmakers R, Boerboom AL, Huijbers BP, Groothoff JW, *et al.* Minimally invasive and computer-navigated total hip arthroplasty: a qualitative and systematic review of the literature. *BMC Musculoskelet Disord* 2010;**11**:92. http://dx.doi.org/10.1186/1471-2474-11-92

179. Restrepo C, Parvizi J, Pour AE, Hozack WJ. Prospective randomized study of two surgical approaches for total hip arthroplasty. *J Arthroplasty* 2010;**25**:671–9. http://dx.doi.org/10.1016/j.arth.2010.02.002

180. Saito S, Tokuhashi Y, Ishii T, Mori S, Hosaka K, Taniguchi S. One- versus two-stage bilateral total hip arthroplasty. *Orthopedics* 2010;**33**. http://dx.doi.org/10.3928/01477447-20100625-07

181. Sariali E, Mauprivez R, Khiami F, Pascal-Mousselard H, Catonne Y. Accuracy of the preoperative planning for cementless total hip arthroplasty. A randomised comparison between three-dimensional computerised planning and conventional templating. *Orthop Traumatol Surg Res* 2012;**98**:151–8. http://dx.doi.org/10.1016/j.otsr.2011.09.023

182. Wassilew GI, Perka C, Janz V, Konig C, Asbach P, Hasart O. Use of an ultrasound-based navigation system for an accurate acetabular positioning in total hip arthroplasty. A prospective, randomized, controlled study. *J Arthroplasty* 2012;**27**:687–94. http://dx.doi.org/10.1016/j.arth.2011.06.038

183. Witzleb WC, Stephan L, Krummenauer F, Neuke A, Gunther KP. Short-term outcome after posterior versus lateral surgical approach for total hip arthroplasty – a randomized clinical trial. *Eur J Med Res* 2009;**14**:256–63. http://dx.doi.org/10.1186/2047-783X-14-6-256

184. Yang B, Li H, He X, Wang G, Xu S. Minimally invasive surgical approaches and traditional total hip arthroplasty: a meta-analysis of radiological and complications outcomes. *PLOS ONE* 2012;**7**:e37947. http://dx.doi.org/10.1371/journal.pone.0037947

185. Krych AJ, Pagnano MW, Coleman WK, Meneghini RM, Kaufman K. No strength or gait benefit of two-incision THA: a brief followup at 1 year. *Clin Orthop Relat Res* 2011;**469**:1110–18. http://dx.doi.org/10.1007/s11999-010-1660-6

186. Krych AJ, Pagnano MW, Wood KC, Meneghini RM, Kaufmann K. No benefit of the two-incision THA over mini-posterior THA: a pilot study of strength and gait. *Clin Orthop Relat Res* 2010;**468**:565–70. http://dx.doi.org/10.1007/s11999-009-0780-3

187. Brodner W, Bitzan P, Meisinger V, Kaider A, Gottsauner-Wolf F, Kotz R. Serum cobalt levels after metal-on-metal total hip arthroplasty. *J Bone Joint Surg Am* 2003;**85–A**:2168–73.

188. Carlsson LV, Albrektsson BE, Albrektsson BG, Albrektsson TO, Jacobsson CM, Macdonald W, *et al.* Stepwise introduction of a bone-conserving osseointegrated hip arthroplasty using RSA and a randomized study: I. Preliminary investigations – 52 patients followed for 3 years. *Acta Orthop* 2006;**77**:549–58. http://dx.doi.org/10.1080/17453670610012601

189. Carlsson LV, Albrektsson T, Albrektsson BE, Jacobsson CM, Macdonald W, Regner L, *et al.* Stepwise introduction of a bone-conserving osseointegrated hip arthroplasty using RSA and a randomized study: II. Clinical proof of concept – 40 patients followed for 2 years. *Acta Orthop* 2006;**77**:559–66. http://dx.doi.org/10.1080/17453670610012610

190. Digas G, Karrholm J, Thanner J. Different loss of BMD using uncemented press-fit and whole polyethylene cups fixed with cement: repeated DXA studies in 96 hips randomized to 3 types of fixation. *Acta Orthop* 2006;**77**:218–26. http://dx.doi.org/10.1080/17453670610045948

191. Digas G, Thanner J, Anderberg C, Karrholm J. Bioactive cement or ceramic/porous coating vs. conventional cement to obtain early stability of the acetabular cup – randomised study of 96 hips followed with radiostereometry. *J Orthop Res* 2004;**22**:1035–43. http://dx.doi.org/10.1016/j.orthres.2003.09.012

192. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am* 2004;**86–A**:963–74.

193. Geerdink CH, Grimm B, Ramakrishnan R, Rondhuis J, Verburg AJ, Tonino AJ. Crosslinked polyethylene compared to conventional polyethylene in total hip replacement: pre-clinical evaluation, in-vitro testing and prospective clinical follow-up study. *Acta Orthop* 2006;**77**:719–25. http://dx.doi.org/10.1080/17453670610012890

194. Grant P, Aamodt A, Falch JA, Nordsletten L. Differences in stability and bone remodeling between a customized uncemented hydroxyapatite coated and a standard cemented femoral stem. A randomized study with use of radiostereometry and bone densitometry. *J Orthop Res* 2005;**23**:1280–5.

195. Grubl A, Weissinger M, Brodner W, Gleiss A, Giurea A, Gruber M, *et al.* Serum aluminium and cobalt levels after ceramic-on-ceramic and metal-on-metal total hip replacement. *J Bone Joint Surg Br* 2006;**88**:1003–5. http://dx.doi.org/10.1302/0301-620X.88B8.17870

196. Hartl A, Schillinger M, Wanivenhaus A. Cemented versus cementless total hip arthroplasty for osteoarthrosis and other non-traumatic diseases. *Cochrane Database Syst Rev* 2004;**3**:CD004850.

197. Howie DW, McGee MA, Costi K, Graves SE. Metal-on-metal resurfacing versus total hip replacement – the value of a randomized clinical trial. *Orthop Clin North Am* 2005;**36**:195–201. http://dx.doi.org/10.1016/j.ocl.2004.12.001

198. Karachalios T, Tsatsaronis C, Efraimis G, Papadelis P, Lyritis G, Diakoumopoulos G. The long-term clinical relevance of calcar atrophy caused by stress shielding in total hip arthroplasty: a 10-year, prospective, randomized study. *J Arthroplasty* 2004;**19**:469–75. http://dx.doi.org/10.1016/j.arth.2003.12.081

199. Karrholm J, Anderberg C, Snorrason F, Thanner J, Langeland N, Malchau H, *et al.* Evaluation of a femoral stem with reduced stiffness. A randomized study with use of radiostereometry and bone densitometry. *J Bone Joint Surg Am* 2002;**84–A**:1651–8.

200. Kim Y-H, Yoon S-H, Kim J-S. Changes in the bone mineral density in the acetabulum and proximal femur after cementless total hip replacement. *J Bone Joint Surg Br* 2007;**89**:174–9. http://dx.doi.org/10.1302/0301-620X.89B2.18634

201. Kim YH. Comparison of polyethylene wear associated with cobalt-chromium and zirconia heads after total hip replacement. A prospective, randomized study. *J Bone Joint Surg Am* 2005;**87**:1769–76. http://dx.doi.org/10.2106/JBJS.D.02572

202. Kraay MJ, Thomas RD, Rimnac CM, Fitzgerald SJ, Goldberg VM. Zirconia versus Co-Cr femoral heads in total hip arthroplasty: early assessment of wear. *Clin Orthop Relat Res* 2006;**453**:86–90. http://dx.doi.org/10.1097/01.blo.0000246544.95316.1f

203. Lombardi J, Mallory TH, Cuckler JM, Williams J, Berend KR, Smith TM. Mid-term results of a polyethylene-free metal-on-metal articulation. *J Arthroplasty* 2004;**19**(Suppl. 2):42–7. http://dx.doi.org/10.1016/j.arth.2004.06.016

204. MacDonald SJ, McCalden RW, Chess DG, Bourne RB, Rorabeck CH, Cleland A, *et al.* Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res* 2003;**406**:282–96. http://dx.doi.org/10.1097/00003086-200301000-00039

205. McCombe PF, Williams SA. A comparison of polyethylene wear rates between cemented and cementless cups. A prospective, randomised trial. *J Bone Joint Surg Br* 2004;**86**:344–9. http://dx.doi.org/10.1302/0301-620X.86B3.14567

206. Medical Advisory Secretariat. Metal-on-metal total hip resurfacing arthroplasty: an evidence-based analysis. *Ont Health Technol Assess Ser* 2006;**6**:1–57.

207. Meek RD, Allan DB. Cemented versus cementless surgical approach for total hip arthroplasty revision. *Cochrane Database Syst Rev* 2005;**2**:CD005322.

208. Alberta Heritage Foundation for Medical Research. *Metal-on-Metal Hip Resurfacing for Young, Active Adults with Degenerative Hip Disease*. Technote TN 33. Edmonton, AB: Alberta Heritage Foundation for Medical Research (AHFMR); 2002.

209. Palm L, Olofsson J, Astrom SE, Ivarsson I. No difference in migration or wear between cemented low-profile cups and standard cups: a randomized radiostereographic study of 53 patients over 3 years. *Acta Orthop* 2007;**78**:479–84. http://dx.doi.org/10.1080/17453670710014112

210. Pitto RP, Blanquaert D, Hohmann D. Alternative bearing surfaces in total hip arthroplasty: zirconia–alumina pairing. Contribution or caveat? *Acta Orthop Belg* 2002;**68**:242–50.

211. Pitto RP, Schikora N, Willmann G, Graef B, Schmidt R. Radiostereoanalysis of press-fit cups with alumina liner – a randomized clinical trial. *Bioceramics* 2003;**15**:817–22.

212. Sonny BB, Aleto TJ, Garino JP, Toni A, Hendricks KJ. Ceramic-on-ceramic versus ceramic-on-polyethylene bearings in total hip arthroplasty: results of a multicenter prospective randomized study and update of modern ceramic total hip trials in the United States. *Hip Int* 2005;**15**:129–35.

213. Strom H, Kolstad K, Mallmin H, Sahlstedt B, Milbrink J. Comparison of the uncemented Cone and the cemented Bimetric hip prosthesis in young patients with osteoarthritis: an RSA, clinical and radiographic study. *Acta Orthop* 2006;**77**:71–8. http://dx.doi.org/10.1080/17453670610045713

214. von Schewelov T, Sanzen L, Onsten I, Carlsson A, Besjakov J. Total hip replacement with a zirconium oxide ceramic femoral head: a randomised roentgen stereophotogrammetric study. *J Bone Joint Surg Br* 2005;**87**:1631–5. http://dx.doi.org/10.1302/0301-620X.87B12.16873

215. Wyness L, Vale L, McCormack K, Grant A, Brazzelli M. The effectiveness of metal on metal hip resurfacing: a systematic review of the available evidence published before 2002. *BMC Health Serv Res* 2004;**4**:39. http://dx.doi.org/10.1186/1472-6963-4-39

216. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis Cartilage* 2007;**15**:981–1000. http://dx.doi.org/10.1016/j.joca.2007.06.014

217. Zhou ZK, Li MG, Borlin N, Wood DJ, Nivbrant B. No increased migration in cups with ceramic-on-ceramic bearing: an RSA study. *Clin Orthop Relat Res* 2006;**448**:39–45. http://dx.doi.org/10.1097/01.blo.0000223999.10389.c9

218. Ayers DC, Hays PL, Drew JM, Eskander MS, Osuch D, Bragdon CR. Two-year radiostereometric analysis evaluation of femoral head penetration in a challenging population of young total hip arthroplasty patients. *J Arthroplasty* 2009;**24**(Suppl.):9–14. http://dx.doi.org/10.1016/j.arth.2009.05.027

219. Cai P, Hu Y, Xie J. Large-diameter delta ceramic-on-ceramic versus common-sized ceramic-on-polyethylene bearings in THA. *Orthopedics* 2012;**35**:e1307–13. http://dx.doi.org/10.3928/01477447-20120822-14

220. Dahlstrand H, Stark A, Anissian L, Hailer NP. Elevated serum concentrations of cobalt, chromium, nickel, and manganese after metal-on-metal alloarthroplasty of the hip: a prospective randomized study. *J Arthroplasty* 2009;**24**:837–45. http://dx.doi.org/10.1016/j.arth.2008.07.019

221. Digas G, Karrholm J, Thanner J, Herberts P. 5-year experience of highly cross-linked polyethylene in cemented and uncemented sockets: two randomized studies using radiostereometric analysis. *Acta Orthop* 2007;**78**:746–54. http://dx.doi.org/10.1080/17453670710014518

222. Digas GF, Karrholm JF, Thanner JF, Malchau HF, Herberts P. Highly cross-linked polyethylene in cemented THA: randomized study of 61 hips. *Clin Orthop Relat Res* 2003;**417**:126–38.

223. Digas G, Karrholm J, Thanner J, Malchau H, Herberts P. The Otto Aufranc Award. Highly cross-linked polyethylene in total hip arthroplasty: randomized evaluation of penetration rate in cemented and uncemented sockets using radiostereometric analysis. *Clin Orthop Relat Res* 2004;**429**:6–16. http://dx.doi.org/10.1097/01.blo.0000150314.70919.e3

224. Johanson PE, Digas G, Herberts P, Thanner J, Karrholm J. Highly crosslinked polyethylene does not reduce aseptic loosening in cemented THA: 10-year findings of a randomized study. *Clin Orthop Relat Res* 2012;**470**:3083–93. http://dx.doi.org/10.1007/s11999-012-2400-x

225. Glyn-Jones S, Isaac S, Hauptfleisch J, McLardy-Smith P, Murray DW, Gill HS. Does highly cross-linked polyethylene wear less than conventional polyethylene in total hip arthroplasty? A double-blind, randomized, and controlled trial using roentgen stereophotogrammetric analysis. *J Arthroplasty* 2008;**23**:337–43. http://dx.doi.org/10.1016/j.arth.2006.12.117

226. Thomas GER, Simpson DJ, Mehmood S, Taylor A, McLardy-Smith P, Gill HS, *et al.* The seven-year wear of highly cross-linked polyethylene in total hip arthroplasty: a double-blind, randomized controlled trial using radiostereometric analysis. *J Bone Joint Surg Am* 2011;**93**:716–22. http://dx.doi.org/10.2106/JBJS.J.00287

227. Hailer NP, Blaheta RA, Dahlstrand H, Stark A. Elevation of circulating HLA DR(+) CD8(+) T-cells and correlation with chromium and cobalt concentrations 6 years after metal-on-metal hip arthroplasty. *Acta Orthop* 2011;**82**:6–12. http://dx.doi.org/10.3109/17453674.2010.548028

228. Jensen C, Aagaard P, Overgaard S. Recovery in mechanical muscle strength following resurfacing vs. standard total hip arthroplasty – a randomised clinical trial. *Osteoarthritis Cartilage* 2011;**19**:1108–16. http://dx.doi.org/10.1016/j.joca.2011.06.011

229. Lavigne M, Therrien M, Nantel J, Roy A, Prince F, Vendittoli PA. The John Charnley Award: the functional outcome of hip resurfacing and large-head THA is the same: a randomized, double-blind study. *Clin Orthop Relat Res* 2010;**468**:326–36. http://dx.doi.org/10.1007/s11999-009-0938-z

230. Lewis PM, Al-Belooshi A, Olsen M, Schemitch EH, Waddell JP. Prospective randomized trial comparing alumina ceramic-on-ceramic with ceramic-on-conventional polyethylene bearings in total hip arthroplasty. *J Arthroplasty* 2010;**25**:392–7. http://dx.doi.org/10.1016/j.arth.2009.01.013

231. Malviya A, Ramaskandhan JR, Bowman R, Hashmi M, Holland JP, Kometa S, *et al.* What advantage is there to be gained using large modular metal-on-metal bearings in routine primary hip replacement? A preliminary report of a prospective randomised controlled trial. *J Bone Joint Surg Br* 2011;**93**:1602–9. http://dx.doi.org/10.1302/0301-620X.93B12.27533

232. McCalden RW, Charron KD, Yuan X, Bourne RB, Naudie DD, MacDonald SJ. Randomised controlled trial comparing early migration of two collarless polished cemented stems using radiostereometric analysis. *J Bone Joint Surg Br* 2010;**92**:935–40. http://dx.doi.org/10.1302/0301-620X.92B7.24462

233. Nikolaou VS, Edwards MR, Bogoch E, Schemitsch EH, Waddell JP. A prospective randomised controlled trial comparing three alternative bearing surfaces in primary total hip replacement. *J Bone Joint Surg Br* 2012;**94**:459–65. http://dx.doi.org/10.1302/0301-620X.94B4.27735

234. Nysted M, Benum P, Klaksvik J, Foss O, Aamodt A. Periprosthetic bone loss after insertion of an uncemented, customized femoral stem and an uncemented anatomical stem. A randomized DXA study with 5-year follow-up. *Acta Orthop* 2011;**82**:410–16. http://dx.doi.org/10.3109/17453674.2011.588860

235. Petersen MK, Andersen NT, Mogensen P, Voight M, Soballe K. Gait analysis after total hip replacement with hip resurfacing implant or Mallory-head Exeter prosthesis: a randomised controlled trial. *Int Orthop* 2011;**35**:667–74. http://dx.doi.org/10.1007/s00264-010-1040-6

236. Schouten R, Malone AA, Tiffen C, Frampton CM, Hooper G. A prospective, randomised controlled trial comparing ceramic-on-metal and metal-on-metal bearing surfaces in total hip replacement. *J Bone Joint Surg Br* 2012;**94**:1462–7. http://dx.doi.org/10.1302/0301-620X.94B11.29343

237. Smolders JM, Hol A, Rijnberg WJ, van Susante JL. Metal ion levels and functional results after either resurfacing hip arthroplasty or conventional metal-on-metal hip arthroplasty. *Acta Orthop* 2011;**82**:559–66. http://dx.doi.org/10.3109/17453674.2011.625533

238. Smolders JM, Hol AF, Rijnders TF, van Susante JL. Changes in bone mineral density in the proximal femur after hip resurfacing and uncemented total hip replacement: a prospective randomised controlled study. *J Bone Joint Surg Br* 2010;**92**:1509–14. http://dx.doi.org/10.1302/0301-620X.92B11.24785

239. Weissinger M, Grubl A, Poll G. Serum-cobalt levels with metal-on-metal bearings in the cement-free total hip arthroplasty results covering two years; prospective study. *Acta Chir Orthop Traumatol Cech* 2011;**78**:410–15.

240. Zijlstra WP, van den Akker-Scheek I, Zee MJ, van Raay JJ. No clinical difference between large metal-on-metal total hip arthroplasty and 28-mm-head total hip arthroplasty? *Int Orthop* 2011;**35**:1771–6. http://dx.doi.org/10.1007/s00264-011-1233-7

241. Zijlstra WP, Bos N, van Raaij JJ. Large head metal-on-metal cementless total hip arthroplasty versus 28 mm metal-on-polyethylene cementless total hip arthroplasty: design of a randomized controlled trial. *BMC Musculoskelet Disord* 2008;**9**:136. http://dx.doi.org/10.1186/1471-2474-9-136

242. Voleti PB, Baldwin KD, Lee GC. Metal-on-metal vs conventional total hip arthroplasty: a systematic review and meta-analysis of randomized controlled trials. *J Arthroplasty* 2012;**27**:1844–9. http://dx.doi.org/10.1016/j.arth.2012.05.023

243. Pailhe R, Sharma A, Reina N, Cavaignac E, Chiron P, Laffosse JM. Hip resurfacing: a systematic review of literature. *Int Orthop* 2012;**36**:2399–410. http://dx.doi.org/10.1007/s00264-012-1686-3

244. Li S, Huang B, Chen Y, Gao H, Fan Q, Zhao J *et al.* Hydroxyapatite-coated femoral stems in primary total hip arthroplasty: a meta-analysis of randomized controlled trials. *Int J Surg* 2013;**11**:477–82. http://dx.doi.org/10.1016/j.ijsu.2013.04.003

245. Abdulkarim A, Ellanti P, Motterlini N, Fahey T, O'Byrne JM. Cemented versus uncemented fixation in total hip replacement: a systematic review and meta-analysis of randomized controlled trials. *Orthoped Rev* 2013;**5**:e8. http://dx.doi.org/10.4081/or.2013.e8

246. Nuesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, *et al.* Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. *BMJ* 2010;**341**:c3515. http://dx.doi.org/10.1136/bmj.c3515

247. Glanville J, Kaunelis D, Mensinkai S. How well do search filters perform in identifying economic evaluations in MEDLINE and EMBASE. *Int J Technol Assess Health Care* 2009;**25**:522–9. http://dx.doi.org/10.1017/S0266462309990523

248. Paisley S, Booth A, Mensinkai S. Health-related quality of life studies. In National Information Center on Health Services Research and Health Care Technology (NICHSR), editor. Etext on *Health Technology Assessment (HTA) Information Resources*. National Library of Medicine; 2005. URL: www.nlm.nih.gov/archive/20060905/nichsr/ehta/chapter12.html (accessed 6 March 2013).

249. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005;**21**:240–54.

250. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality of assessment. *Pharmacoeconomics* 2006;**24**:355–71. http://dx.doi.org/10.2165/00019053-200624040-00006

251. Pollard TCB, Baker RP, Eastaugh-Waring SJ, Bannister GC. Treatment of the young active patient with osteoarthritis of the hip. A five- to seven-year comparison of hybrid total hip arthroplasty and metal-on-metal resurfacing. *J Bone Joint Surg Br* 2008;**90–B**(Suppl. III):504.

252. Baker RP, Pollard TCB, Eastaugh-Waring SJ, Bannister GC. A medium-term comparison of hybrid hip replacement and Birmingham hip resurfacing in active young patients. *J Bone Joint Surg Br* 2011;**93**:158–63. http://dx.doi.org/10.1302/0301-620X.93B2.25625

253. Bozic KJ, Pui CM, Ludeman MJ, Vail TP, Silverstein MD. Do the potential benefits of metal-on-metal hip resurfacing justify the increased cost and risk of complications? *Clin Orthop Relat Res* 2010;**468**:2301–12. http://dx.doi.org/10.1007/s11999-010-1301-0

254. Amman S, Cizik A, Leopold SS, Manner PA. Two-incision minimally invasive vs standard total hip arthroplasty: comparison of component position and hospital costs. *J Arthroplasty* 2012;**27**:1569–74. http://dx.doi.org/10.1016/j.arth.2012.03.006

255. Duwelius PJ, Brenner JS, Reyner DP, George JC. Cost effectiveness of minimally invasive total hip arthroplasty. *Sem Arthroplasty* 2008;**19**:186–93. http://dx.doi.org/10.1053/j.sart.2008.02.005

256. Straumann D, Valderrabano V, Eckstein M, Dick W, Dora C. Cost–benefit analysis of MIS THA: model-based analysis of the consequences for Switzerland. *Hip Int* 2006;**16**(Suppl. 4):54–7.

257. Fordham R, Skinner J, Wang X, Nolan J, Exeter Primary Outcome Study Group. The economic benefit of hip replacement: a 5-year follow-up of costs and outcomes in the Exeter Primary Outcomes Study. *BMJ Open* 2012;**2**:e000752. http://dx.doi.org/10.1136/bmjopen-2011-000752

258. Hulleberg G, Aamodt A, Espehaug B, Benum P. A clinical and radiographic 13-year follow-up study of 138 Charnley hip arthroplasties in patients 50–70 years old: comparison of university hospital data and registry data. *Acta Orthop* 2008;**79**:609–17. http://dx.doi.org/10.1080/17453670810016614

259. Bozic KJ, Morshed S, Silverstein MD, Rubash HE, Kahn JG. Use of cost-effectiveness analysis to evaluate new technologies in orthopaedics: the case of alternative bearing surfaces in total hip arthroplasty. *J Bone Joint Surg* 2006;**88**:706–14. http://dx.doi.org/10.2106/JBJS.E.00614

260. Marinelli M, Soccetti A, Panfoli N, de Palma L. Cost-effectiveness of cemented versus cementless total hip arthroplasty. A Markov decision analysis based on implant cost. *J Orthop Traumatol* 2008;**9**:23–8. http://dx.doi.org/10.1007/s10195-008-0100-9

261. di Tanna GL, Ferro S, Cipriani F, Bordini B, Stea S, Toni A, *et al.* Modeling the cost-effectiveness for cement-less and hybrid prosthesis in total hip replacement in Emilia Romagna, Italy. *J Surg Res* 2011;**169**:227–33. http://dx.doi.org/10.1016/j.jss.2009.10.031

262. Cummins JS, Tomek IM, Kantor SR, Furnes O, Engesaeter LB, Finlayson SR. Cost-effectiveness of antibiotic-impregnated bone cement used in primary total hip arthroplasty. *J Bone Joint Surg Am* 2009;**91**:634–41. http://dx.doi.org/10.2106/JBJS.G.01029

263. Larsen K, Sorensen OG, Hansen TB, Thomsen PB, Soballe K. Accelerated perioperative care and rehabilitation intervention for hip and knee replacement is effective: a randomized clinical trial involving 87 patients with 3 months of follow-up. *Acta Orthop* 2008;**79**:149–59. http://dx.doi.org/10.1080/17453670710014923

264. Cullen J, Bramley D, Armstrong D, Butler L, Rouse P, Ashton T. Increasing productivity, reducing cost and improving quality in elective surgery in New Zealand: the Waitemata District Health Board joint arthroplasty pilot. *Int Med J* 2012;**42**:620–6. http://dx.doi.org/10.1111/j.1445-5994.2012.02815.x

265. Brunenberg DE, van Steyn MJ, Sluimer JC, Bekebrede LL, Bulstra SK, Joore MA. Joint recovery programme versus usual care – an economic evaluation of a clinical pathway for joint replacement surgery. *Med Care* 2005;**43**:1018–26. http://dx.doi.org/10.1097/01.mlr.0000178266.75744.35

266. Bak P, Muller WD, Bocker B, Smolenski UC. Short-term patterns of recovery from total hip and knee arthroplasty after multidisciplinary inpatient rehabilitation. *Phys Med Rehab Kuror* 2008;**18**:11–18. http://dx.doi.org/10.1055/s-2007-993195

267. Batsis JA, Naessens JM, Keegan MT, Wagie AE, Huddleston PM, Huddleston JM. Impact of body mass on hospital resource use in total hip arthroplasty. *Public Health Nutr* 2009;**12**:1122–32. http://dx.doi.org/10.1017/S1368980009005072

268. Jibodh SR, Gurkan I, Wenz JF. In-hospital outcome and resource use in hip arthroplasty: influence of body mass. *Orthopedics* 2004;**27**:594–601.

269. Segal L, Day SE, Chapman AB, Osborne RH. Can we reduce disease burden from osteoarthritis? *Med J Aust* 2004;**180**(Suppl.):S11–17.

270. Tien WC, Kao HY, Tu YK, Chiu HC, Lee KT, Shi HY. A population-based study of prevalence and hospital charges in total hip and knee replacement. *Int Orthop* 2009;**33**:949–54. http://dx.doi.org/10.1007/s00264-008-0612-1

271. Dutka J, Dutka L, Janiszewski M, Hajduk G. Cost analysis and sociomedical aspects of the conservative and surgical treatment of hip osteoarthritis. *Ortop Traumatol Rehabil* 2008;**10**:537–46.

272. Fujita K, Makimoto K, Higo T, Shigematsu M, Hotokebuchi T. Changes in the WOMAC, EuroQol and Japanese lifestyle measurements among patients undergoing total hip arthroplasty. *Osteoarthritis Cartilage* 2009;**17**:848–55. http://dx.doi.org/10.1016/j.joca.2008.11.012

273. Higashi H, Barendregt JJ. Cost-effectiveness of total hip and knee replacements for the Australian population with osteoarthritis: discrete-event simulation model. *PLOS ONE* 2011;**6**:e25403. http://dx.doi.org/10.1371/journal.pone.0025403

274. Jimenez-Garcia R, Villanueva-Martinez M, Fernandez-de-Las-Penas C, Hernandez-Barrera V, Rios-Luna A, Garrido PC, *et al.* Trends in primary total hip arthroplasty in Spain from 2001 to 2008: evaluating changes in demographics, comorbidity, incidence rates, length of stay, costs and mortality. *BMC Musculoskelet Disord* 2011;**12**:43. http://dx.doi.org/10.1186/1471-2474-12-43

275. Mota REM. Cost-effectiveness analysis of early versus late total hip replacement in Italy. *Value Health* 2013;**16**:267–79. http://dx.doi.org/10.1016/j.jval.2012.10.020

276. Scheerlinck T, Duquet W, Casteleyn P-P. Socioeconomic aspects of total hip arthroplasty: a one-year survey in a Belgian university hospital. *Acta Orthop Belg* 2004;**70**:525–33.

277. Rana AJ, Iorio R, Healy WL. Hospital economics of primary THA decreasing reimbursement and increasing cost, 1990 to 2008. *Clin Orthop Relat Res* 2011;**469**:355–61. http://dx.doi.org/10.1007/s11999-010-1526-y

278. O'Shea K, Bale E, Murray P. Cost analysis of primary total hip replacement. *Ir Med J* 2002;**95**:177–80.

279. Zhang Y, Zhang H, Clarke HD, Hattrup SJ. Analysis of total joint arthroplasty costs in Chinese patients. *J Arthroplasty* 2012;**27**:1423–8. http://dx.doi.org/10.1016/j.arth.2012.01.014

280. Stargardt T. Health service costs in Europe: cost and reimbursement of primary hip replacement in nine countries. *Health Econ* 2008;**17**(Suppl. 1):S9–20. http://dx.doi.org/10.1002/hec.1328

281. Judge A, Cooper C, Arden NK, Williams S, Hobbs N, Dixon D, *et al.* Pre-operative expectation predicts 12-month post-operative outcome among patients undergoing primary total hip replacement in European orthopaedic centres. *Osteoarthritis Cartilage* 2011;**19**:659–67. http://dx.doi.org/10.1016/j.joca.2011.03.009

282. Fielden JM, Cumming JM, Horne JG, Devane PA, Slack A, Gallagher LM. Waiting for hip arthroplasty: economic costs and health outcomes. *J Arthroplasty* 2005;**20**:990–7. http://dx.doi.org/10.1016/j.arth.2004.12.060

283. Tuominen U, Sintonen H, Hirvonen J, Seitsalo S, Paavolainen P, Lehto M, *et al.* The effect of waiting time on health and quality of life outcomes and costs of medication in hip replacement patients: a randomized clinical trial. *Osteoarthritis Cartilage* 2009;**17**:1144–50. http://dx.doi.org/10.1016/j.joca.2009.03.014

284. Bozic KJ, Chiu VW, Slover JD, Immerman I, Kahn JG. Health state utility in patients with osteoarthritis of the hip and total hip arthroplasty. *J Arthroplasty* 2011;**26**(Suppl.):129–32. http://dx.doi.org/10.1016/j.arth.2011.03.033

285. Bohm ER, Dunbar MJ, Frood JJ, Johnson TM, Morris KA. Rehospitalizations, early revisions, infections, and hospital resource use in the first year after hip and knee arthroplasties. *J Arthroplasty* 2012;**27**:232–7. http://dx.doi.org/10.1016/j.arth.2011.05.004

286. March L, Cross M, Tribe K, Lapsley H, Courtenay B, Brooks P. Cost of joint replacement surgery for osteoarthritis: the patients' perspective. *J Rheumatol* 2002;**29**:1006–14.

287. Montin L, Suominen T, Katajisto J, Lepisto J, Leino-Kilpi H. Economic outcomes from patients' perspective and health-related quality of life after total hip arthroplasty. *Scand J Caring Sci* 2009;**23**:11–20. http://dx.doi.org/10.1111/j.1471-6712.2007.00580.x

288. de Palma L, Procaccini R, Soccetti A, Marinelli M. Hospital cost of treating early dislocation following hip arthroplasty. *Hip Int* 2012;**22**:62–7. http://dx.doi.org/10.5301/HIP.2012.9059

289. Burns AWR, Bourne RB. Economics of revision total hip arthroplasty. *Curr Orthop* 2006;**20**:203–7. http://dx.doi.org/10.1016/j.cuor.2006.02.007

290. Urquhart DM, Hanna F, Graves S, Wang Y, Cameron P, Hannaford A, *et al.* In-hospital outcomes and hospital resource utilization of hip replacement procedures. *ANZ J Surg* 2008;**78**:875–80. http://dx.doi.org/10.1111/j.1445-2197.2008.04684.x

291. Kurtz SM, Lau E, Watson H, Schmier JK, Parvizi J. Economic burden of periprosthetic joint infection in the United States. *J Arthroplasty* 2012;**27**(Suppl.):61–5. http://dx.doi.org/10.1016/j.arth.2012.02.022

292. Vanhegan IS, Malik AK, Jayakumar P, Ul Islam S, Haddad FS. A financial analysis of revision hip arthroplasty: the economic burden in relation to the national tariff. *J Bone Joint Surg Br* 2012;**94**:619–23. http://dx.doi.org/10.1302/0301-620X.94B5.27073

293. Parvizi J, Pawasarat IM, Azzam KA, Joshi A, Hansen EN, Bozic KJ. Periprosthetic joint infection: the economic impact of methicillin-resistant infections. *J Arthroplasty* 2010;**25**(Suppl.):103–7. http://dx.doi.org/10.1016/j.arth.2010.04.011

294. Feeny D, Wu L, Eng K. Comparing short form 6D, standard gamble, and Health Utilities Index Mark 2 and Mark 3 utility scores: results from total hip arthroplasty patients. *Qual Life Res* 2004;**13**:1659–70. http://dx.doi.org/10.1007/s11136-004-6189-2

295. Dawson J, Fitzpatrick R, Frost S, Gundle R, McLardy-Smith P, Murray D. Evidence for the validity of a patient-based instrument for assessment of outcome after revision hip replacement. *J Bone Joint Surg Br* 2001;**83**:1125–9. http://dx.doi.org/10.1302/0301-620X.83B8.11643

296. Jones CA, Pohar S. Health-related quality of life after total joint arthroplasty: a scoping review. *Clin Geriatr Med* 2012;**28**:395–429. http://dx.doi.org/10.1016/j.cger.2012.06.001

297. Ostendorf M, van Stel HF, Buskens E, Schrijvers AJ, Marting LN, Verbout AJ, *et al.* Patient-reported outcome in total hip replacement. A comparison of five instruments of health status. *J Bone Joint Surg Br* 2004;**86**:801–8. http://dx.doi.org/10.1302/0301-620X.86B6.14950

298. Rolfson O, Karrholm J, Dahlberg LE, Garellick G. Patient-reported outcomes in the Swedish Hip Arthroplasty Register: results of a nationwide prospective observational study. *J Bone Joint Surg Br* 2011;**93**:867–75. http://dx.doi.org/10.1302/0301-620X.93B7.25737

299. McKenzie L, Vale L, Stearns S, McCormack K. Metal on metal hip resurfacing arthroplasty. An economic analysis. *Eur J Health Econ* 2003;**4**:122–9. http://dx.doi.org/10.1007/s10198-002-0158-x

300. Clement RC, Kamath AF, Derman PB, Garino JP, Lee GC. Bipolar sealing in revision total hip arthroplasty for infection: efficacy and cost analysis. *J Arthroplasty* 2012;**27**:1376–81. http://dx.doi.org/10.1016/j.arth.2011.11.016

301. Feeny D, Blanchard CM, Mahon JL, Bourne R, Rorabeck C, Stitt L, *et al.* The stability of utility scores: test–retest reliability and the interpretation of utility scores in elective total hip arthroplasty. *Qual Life Res* 2004;**13**:15–22. http://dx.doi.org/10.1023/B:QURE.0000015307.33811.2d

302. Jansson KA, Granath F. Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthop* 2011;**82**:82–9. http://dx.doi.org/10.3109/17453674.2010.548026

303. Larsen K, Hansen TB, Thomsen PB, Christiansen T, Soballe K. Cost-effectiveness of accelerated perioperative care and rehabilitation after total hip and knee arthroplasty. *J Bone Joint Surg* 2009;**91**:761–72. http://dx.doi.org/10.2106/JBJS.G.01472

304. Lemon MA, Hamilton PD, Field RE. Comparing total hip and knee replacement costs. *Br J Healthcare Manag* 2008;**14**:108–12. http://dx.doi.org/10.12968/bjhc.2008.14.3.108

305. Rasch A, Dalen N, Berg HE. Muscle strength, gait, and balance in 20 patients with hip osteoarthritis followed for 2 years after THA. *Acta Orthop* 2010;**81**:183–8. http://dx.doi.org/10.3109/17453671003793204

306. Rolfson O, Dahlberg LE, Nilsson J, Malchau H, Garellick G. Variables determining outcome in total hip replacement surgery. *J Bone Joint Surg Br* 2009;**91**:157–61. http://dx.doi.org/10.1302/0301-620X.91B2.20765

307. Tso P, Walker K, Mahomed N, Coyte PC, Rampersaud YR. Comparison of lifetime incremental cost:utility ratios of surgery relative to failed medical management for the treatment of hip, knee and spine osteoarthritis modelled using 2-year postsurgical values. *Can J Surg* 2012;**55**:181–90. http://dx.doi.org/10.1503/cjs.033910

308. Xie F, Thumboo J, Fong KY, Lo NN, Yeo SJ, Yang KY, *et al.* Direct and indirect costs of osteoarthritis in Singapore: a comparative study among multiethnic Asian patients with osteoarthritis. *J Rheumatol* 2007;**34**:165–71.

309. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. New York, NY: Oxford University Press; 2005.

310. Akiyama H, Hoshino A, Iida H, Shindo H, Takakura Y, Miura H, *et al.* A pilot project for the Japan arthroplasty register. *J Orthop Sci* 2012;**17**:358–69. http://dx.doi.org/10.1007/s00776-012-0229-5

311. Buergi ML, Walter WL. Hip resurfacing arthroplasty: the Australian experience. *J Arthroplasty* 2007;**22**(Suppl. 3):61–5. http://dx.doi.org/10.1016/j.arth.2007.05.021

312. Conroy JL, Whitehouse SL, Graves SE, Pratt NL, Ryan P, Crawford RW. Risk factors for revision for early dislocation in total hip arthroplasty. *J Arthroplasty* 2008;**23**:867–72. http://dx.doi.org/10.1016/j.arth.2007.07.009

313. Corten K, MacDonald SJ. Hip resurfacing data from national joint registries: what do they tell us? What do they not tell us? *Clin Orthop Relat Res* 2010;**468**:351–7. http://dx.doi.org/10.1007/s11999-009-1157-3

314. Luo R, Brekke A, Noble PC. The financial impact of joint registries in identifying poorly performing implants. *J Arthroplasty* 2012;**27**(Suppl.):66–71. http://dx.doi.org/10.1016/j.arth.2012.03.046

315. Sexton SA, Walter WL, Jackson MP, De Steiger R, Stanford T. Ceramic-on-ceramic bearing surface and risk of revision due to dislocation after primary total hip replacement. *J Bone Joint Surg Br* 2009;**91**:1448–53. http://dx.doi.org/10.1302/0301-620X.91B11.22100

316. Allami MK, Fender D, Khaw FM, Sandher DR, Esler C, Harper WM, *et al.* Outcome of Charnley total hip replacement across a single health region in England: the results at ten years from a regional arthroplasty register. *J Bone Joint Surg* 2006;**88**:1293–8. http://dx.doi.org/10.1302/0301-620X.88B10.17933

317. Jameson SS, Baker PN, Mason J, Gregg PJ, Brewster N, Deehan DJ, *et al.* The design of the acetabular component and size of the femoral head influence the risk of revision following 34 721 single-brand cemented hip replacements: a retrospective cohort study of medium-term data from a National Joint Registry. *J Bone Joint Surg Br* 2012;**94**:1611–17. http://dx.doi.org/10.1302/0301-620X.94B12.30040

318. McMinn DJ, Snell KI, Daniel J, Treacy RB, Pynsent PB, Riley RD. Mortality and implant revision rates of hip arthroplasty in patients with osteoarthritis: registry based cohort study. *BMJ* 2012;**344**:e3319. [Corrections published in *BMJ* 2012;**344**:e4371 and *BMJ* 2012;**345**:e6282]. http://dx.doi.org/10.1136/bmj.e3319

319. Meek RMD, Allan DB, McPhillips G, Kerr L, Howie CR. Late dislocation after total hip arthroplasty. *Clin Med Res* 2008;**6**:17–23. http://dx.doi.org/10.3121/cmr.2008.770

320. Stea S, Bordini B, De Clerico M, Petropulacos K, Toni A. First hip arthroplasty register in Italy: 55,000 cases and 7 year follow-up. *Int Orthop* 2009;**33**:339–46. http://dx.doi.org/10.1007/s00264-007-0465-z

321. Eskelinen A, Remes V, Helenius I, Pulkkinen P, Nevalainen J, Paavolainen P. Uncemented total hip arthroplasty for primary osteoarthritis in young patients: a mid-to long-term follow-up study from the Finnish Arthroplasty Register. *Acta Orthop* 2006;**77**:57–70. http://dx.doi.org/10.1080/17453670610045704

322. Eskelinen A, Paavolainen P, Helenius I, Pulkkinen P, Remes V. Total hip arthroplasty for rheumatoid arthritis in younger patients: 2,557 replacements in the Finnish Arthroplasty Register followed for 0–24 years. *Acta Orthop* 2006;**77**:853–65. http://dx.doi.org/10.1080/17453670610013132

323. Eskelinen A, Remes V, Helenius I, Pulkkinen P, Nevalainen J, Paavolainen P. Total hip arthroplasty for primary osteoarthrosis in younger patients in the Finnish arthroplasty register. 4,661 primary replacements followed for 0–22 years. *Acta Orthop* 2005;**76**:28–41. http://dx.doi.org/10.1080/00016470510030292

324. Makela K, Eskelinen A, Pulkkinen P, Paavolainen P, Remes V. Cemented total hip replacement for primary osteoarthritis in patients aged 55 years or older: results of the 12 most common cemented implants followed for 25 years in the Finnish Arthroplasty Register. *J Bone Joint Surg Br* 2008;**90**:1562–9. http://dx.doi.org/10.1302/0301-620X.90B12.21151

325. Makela KT, Eskelinen A, Pulkkinen P, Paavolainen P, Remes V. Total hip arthroplasty for primary osteoarthritis in patients fifty-five years of age or older. *J Bone Joint Surg Am* 2008;**90**:2160–70. http://dx.doi.org/10.2106/JBJS.G.00870

326. Makela KT, Eskelinen A, Paavolainen P, Pulkkinen P, Remes V. Cementless total hip arthroplasty for primary osteoarthritis in patients aged 55 years and older. *Acta Orthop* 2010;**81**:42–52. http://dx.doi.org/10.3109/17453671003635900

327. Makela KT, Eskelinen A, Pulkkinen P, Virolainen P, Paavolainen P, Remes V. Cemented versus cementless total hip replacements in patients fifty-five years of age or older with rheumatoid arthritis. *J Bone Joint Surg Am* 2011;**93**:178–86. http://dx.doi.org/10.2106/JBJS.I.01283

328. Makela KT, Eskelinen A, Pulkkinen P, Paavolainen P, Remes V. Results of 3,668 primary total hip replacements for primary osteoarthritis in patients under the age of 55 years. *Acta Orthop* 2011;**82**:521–9. http://dx.doi.org/10.3109/17453674.2011.618908

329. Seppanen M, Makela K, Virolainen P, Remes V, Pulkkinen P, Eskelinen A. Hip resurfacing arthroplasty: short-term survivorship of 4,401 hips from the Finnish Arthroplasty Register. *Acta Orthop* 2012;**83**:207–13. http://dx.doi.org/10.3109/17453674.2012.693016

330. Skytta ET, Jarkko L, Antti E, Huhtala H, Ville R. Increasing incidence of hip arthroplasty for primary osteoarthritis in 30- to 59-year-old patients. *Acta Orthop* 2011;**82**:1–5. http://dx.doi.org/10.3109/17453674.2010.548029

331. Espehaug B, Furnes O, Engester LB, Havelin LI. 18 years of results with cemented primary hip prostheses in the Norwegian Arthroplasty Register: concerns about some newer implants. *Acta Orthop* 2009;**80**:1–11. http://dx.doi.org/10.3109/17453670903161124

332. Espehaug B, Furnes O, Engesaeter LB, Havelin LI. Hip arthroplasty in Norway 1989–2008. *Tidsskr Nor Laegeforen* 2011;**131**:1543–8. http://dx.doi.org/10.4045/tidsskr.09.1091

333. Fevang B-T, Lie SA, Havelin LI, Engesaeter LB, Furnes O. Improved results of primary total hip replacement: results from the Norwegian Arthroplasty Register, 1987–2007. *Acta Orthop* 2010;**81**:649–59. http://dx.doi.org/10.3109/17453674.2010.537807

334. Kadar T, Dybvik E, Hallan G, Furnes O, Havelin LI. Head material influences survival of a cemented total hip prosthesis in the Norwegian Arthroplasty Register. *Clin Orthop Relat Res* 2012;**470**:3007–13. http://dx.doi.org/10.1007/s11999-012-2396-2

335. Schrama JC, Espehaug B, Hallan G, Engesaeter LB, Furnes O, Havelin LI, *et al.* Risk of revision for infection in primary total hip and knee arthroplasty in patients with rheumatoid arthritis compared with osteoarthritis: a prospective, population-based study on 108,786 hip and knee joint arthroplasties from the Norwegian Arthroplasty Register. *Arthritis Care Res* 2010;**62**:473–9. http://dx.doi.org/10.1002/acr.20036

336. Franklin PD, Allison JJ, Ayers DC. Beyond joint implant registries: a patient-centered research consortium for comparative effectiveness in total joint replacement. *JAMA* 2012;**308**:1217–18. http://dx.doi.org/10.1001/jama.2012.12568

337. Namba RS, Inacio MC, Paxton EW. Risk factors associated with surgical site infection in 30,491 primary total hip replacements. *J Bone Joint Surg Br* 2012;**94**:1330–8. http://dx.doi.org/10.1302/0301-620X.94B10.29184

338. Paxton EW, Inacio M, Slipchenko T, Fithian DC. The Kaiser Permanente National Total Joint Replacement Registry. *Permanente J* 2008;**12**:12–16. http://dx.doi.org/10.7812/TPP/08-008

339. Johnsen SP, Sorensen HT, Lucht U, Soballe K, Overgaard S, Pedersen AB. Patient-related predictors of implant failure after primary total hip replacement in the initial, short- and long-terms. A nationwide Danish follow-up study including 36,984 patients. *J Bone Joint Surg Br* 2006;**88**:1303–8. http://dx.doi.org/10.1302/0301-620X.88B10.17399

340. Pedersen AB, Johnsen SP, Overgaard S, Soballe K, Sorensen HT, Lucht U. Total hip arthroplasty in Denmark: incidence of primary operations and revisions during 1996–2002 and estimated future demands. *Acta Orthop* 2005;**76**:182–9. http://dx.doi.org/10.1080/00016470510030553

341. Pedersen AB, Baron JA, Overgaard S, Johnsen SP. Short- and long-term mortality following primary total hip replacement for osteoarthritis: a Danish nationwide epidemiological study. *J Bone Joint Surg Br* 2011;**93**:172–7. http://dx.doi.org/10.1302/0301-620X.93B2.25629

342. Rud-Sorensen C, Pedersen AB, Johnsen SP, Riis AH, Overgaard S. Survival of primary total hip arthroplasty in rheumatoid arthritis patients. *Acta Orthop* 2010;**81**:60–5. http://dx.doi.org/10.3109/17453671003685418

343. Lazarinis S, Krarholm J, Hailer NP. Increased risk of revision of acetabular cups coated with hydroxyapatite: a Swedish Hip Arthroplasty Register study involving 8,043 total hip replacements. *Acta Orthop* 2010;**81**:53–9. http://dx.doi.org/10.3109/17453670903413178

344. Weiss RJ, Hailer NP, Stark A, Karrholm J. Survival of uncemented acetabular monoblock cups: evaluation of 210 hips in the Swedish Hip Arthroplasty Register. *Acta Orthop* 2012;**83**:214–19. http://dx.doi.org/10.3109/17453674.2012.688726

345. Nečas L, Katina S, Krivanek S, Uhlárová J. Review of the annual report of the Slovakian Arthroplasty Register – 2010. *Acta Chir Orthop Traumatol Cech* 2011;**78**(Suppl. B):1–59.

346. Boyer P, Boutron I, Ravaud P. Scientific production and impact of national registers: the example of orthopaedic national registers. *Osteoarthritis Cartilage* 2011;**19**:858–63. http://dx.doi.org/10.1016/j.joca.2011.02.006

347. Graves SE, Rothwell A, Tucker K, Jacobs JJ, Sedrakyan A. A multinational assessment of metal-on-metal bearings in hip replacement. *J Bone Joint Surg Am* 2011;**93**(Suppl. 3):43–7. http://dx.doi.org/10.2106/JBJS.K.01220

348. Havelin LI, Fenstad AM, Salomonsson R, Mehnert F, Furnes O, Overgaard S, *et al.* The Nordic Arthroplasty Register Association: a unique collaboration between 3 national hip arthroplasty registries with 280,201 THRs. *Acta Orthop* 2009;**80**:393–401. http://dx.doi.org/10.3109/17453670903039544

349. Johanson PE, Fenstad AM, Furnes O, Garellick G, Havelin LI, Overgaard S, *et al.* Inferior outcome after hip resurfacing arthroplasty than after conventional arthroplasty. Evidence from the Nordic Arthroplasty Register Association (NARA) database, 1995 to 2007. *Acta Orthop* 2010;**81**:535–41. http://dx.doi.org/10.3109/17453674.2010.525193

350. Sadoghi P, Schroder C, Fottner A, Steinbruck A, Betz O, Muller PE, *et al.* Application and survival curve of total hip arthroplasties: a systematic comparative analysis using worldwide hip arthroplasty registers. *Int Orthop* 2012;**36**:2197–203. http://dx.doi.org/10.1007/s00264-012-1614-6

351. Schuh R, Neumann D, Rauf R, Hofstaetter J, Boehler N, Labek G. Revision rate of Birmingham hip resurfacing arthroplasty: comparison of published literature and arthroplasty register data. *Int Orthop* 2012;**36**:1349–54. http://dx.doi.org/10.1007/s00264-012-1502-0

352. Waddell J, Johnson K, Hein W, Raabe J, FitzGerald G, Turibio F. Orthopaedic practice in total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). *Am J Orthop (Belle Meade NJ)* 2010;**39**(Suppl.):5–13.

353. Jameson SS, Baker PN, Mason J, Porter ML, Deehan DJ, Reed MR. Independent predictors of revision following metal-on-metal hip resurfacing: a retrospective cohort study using National Joint Registry data. *J Bone Joint Surg Br* 2012;**94**:746–54. http://dx.doi.org/10.1302/0301-620X.94B6.29239

354. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70**:41–55. http://dx.doi.org/10.1093/biomet/70.1.41

355. Leuven E, Sianesi B. *PSMATCH2: Stata Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing, and Covariate Imbalance Testing*. Statistical Software Components S432001. Boston College Department of Economics, MA; 2003.

356. Coviello V, Boggess M. Cumulative incidence estimation in the presence of competing risks. *Stata J* 2004;**4**:103–12.

357. Collet D. *Modelling Survival Data in Medical Research*. London: Chapman and Hall/CRC; 2003.

358. Peace KE, editor. *Design and Analysis of Clinical Trials with Time to Event Endpoints*. Boca Raton, FL: CRC; 2009. http://dx.doi.org/10.1201/9781420066401

359. Machin D, Cheung YB, Parmar MKB. *Survival Analysis: A Practical Approach*. Chichester: John Wiley; 2006. http://dx.doi.org/10.1002/0470034572

360. Crowther MJ, Lambert PC. stgenreg: a Stata package for general parametric survival analysis. *J Stat Software* 2013;**53**.

361. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 2002;**21**:2175–97. http://dx.doi.org/10.1002/sim.1203

362. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J* 2009;**9**:265–90.

363. Boyer KA, Beaupre GS, Andriacchi TP. Gender differences exist in the hip joint moments of healthy older walkers. *J Biomech* 2008;**41**:3360–5. http://dx.doi.org/10.1016/j.jbiomech.2008.09.030

364. National Institute for Health and Care Excellence. *Consideration of Consultation Responses on Review Proposal: Review of TA2; Hip Disease Replacement Prostheses and TA44; Hip Disease Metal on Metal Resurfacing*. London: NICE; 2011. URL: www.nice.org.uk/nicemedia/live/11386/55850/55850.pdf (accessed 19 June 2013).

365. Latimer NR. Survival analysis for economic evaluations alongside clinical trials – extrapolation with patient level data: inconsistencies, limitations and a practical guide. *Med Decis Making* 2013;**33**:743–54. http://dx.doi.org/10.1177/0272989X12472398

366. Latimer N. *NICE DSU Technical Support Document 14: Survival Analysis for Economic Evaluations Alongside Clinical Trials – Extrapolation with Patient-Level Data*. NICE Decision Support Unit; March 2013. URL: www.nicedsu.org.uk/NICE%20DSU%20TSD%20Survival%20analysis.updated%20March%202013.pdf (accessed 9 April 2013).

367. Briggs A, Sculpher M, Dawson J, Fitzpatrick R, Murray D, Malchau H. *Modelling the Cost-Effectiveness of Primary Hip Replacement: How Cost-effective is the Spectron Compared to the Charnley Prosthesis?* CHE Technical Paper Series 28. York: University of York; 2003. URL: www.york.ac.uk/media/che/documents/papers/technicalpapers/CHE%20Technical%20Paper%2028.pdf (accessed 16 May 2013).

368. Schulte KR, Callaghan JJ, Kelley SS, Johnston RC. The outcome of Charnley total hip arthroplasty with cement after a minimum twenty-year follow-up. The results of one surgeon. *J Bone Joint Surg Am* 1993;**75**:961–75.

369. Madey SM, Callaghan JJ, Olejniczak JP, Goetz DD, Johnston RC. Charnley total hip arthroplasty with use of improved techniques of cementing. The results after a minimum of fifteen years of follow-up. *J Bone Joint Surg Am* 1997;**79**:53–64.

370. Callaghan JJ, Albright JC, Goetz DD, Olejniczak JP, Johnston RC. Charnley total hip arthroplasty with cement. Minimum twenty-five-year follow-up. *J Bone Joint Surg Am* 2000;**82**:487–97.

371. Callaghan JJ, Templeton JE, Liu SS, Pedersen DR, Goetz DD, Sullivan PM, *et al.* Results of Charnley total hip arthroplasty at a minimum of thirty years. A concise follow-up of a previous report. *J Bone Joint Surg Am* 2004;**86**:690–5.

372. Schreurs BW, Bolder SB, Gardeniers JW, Verdonschot N, Slooff TJ, Veth RP. Acetabular revision with impacted morsellised cancellous bone grafting and a cemented cup. A 15- to 20-year follow-up. *J Bone Joint Surg Br* 2004;**86**:492–7.

373. Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al.* Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses. *Health Technol Assess* 1998;**2**(20). http://dx.doi.org/10.3310/hta2200

374. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: NICE; 2008. URL: www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf (accessed 29 May 2012).

375. National Institute for Health and Care Excellence. *Total Hip Replacement and Resurfacing Arthroplasty for End-Stage Arthritis of the Hip (Review of Technology Appraisal Guidance 2 and 44)*. 2012. URL: www.nice.org.uk/guidance/ta304/resources/arthritis-of-the-hip-end-stage-hip-replacement-total-and-surface-replacement-appendix-a-draft-scope2 (accessed 8 September 2014).

376. Canterbury District Health Board. *New Zealand Joint Registry: Thirteen Year Report January 1999 to December 2011*. 2012. URL: www.cdhb.govt.nz/njr/reports/A2D65CA3.pdf (accessed 16 May 2013).

377. Curtis L. *Unit Costs of Health and Social Care 2011*. Canterbury: Personal Social Services Research Unit, University of Kent; 2011.

378. Baad-Hansen T, Kold S, Nielsen PT, Laursen MB, Christensen PH, Soballe K. Comparison of trabecular metal cups and titanium fiber-mesh cups in primary hip arthroplasty: a randomized RSA and bone mineral densitometry study of 50 hips. *Acta Orthop* 2011;**82**:155–60. http://dx.doi.org/10.3109/17453674.2011.572251

379. Department of Health. *NHS Reference Costs*. URL: www.gov.uk/government/collections/nhs-reference-costs (accessed 8 September 2014).

380. Australian Orthopaedic Associated National Joint Registry. *Australian Orthopaedic Associated National Joint Registry: 2011 Annual Report*. URL https://aoanjrr.dmac.adelaide.edu.au/documents/10180/44800/Annual%20Report%202011?version=1.2&t=1347337258367 (accessed 18 April 2013).

381. Baad-Hansen T, Kold S, Olsen N, Christensen F, Soballe K. Excessive distal migration of fiber-mesh coated femoral stems. *Acta Orthop* 2011;**82**:308–14. http://dx.doi.org/10.3109/17453674.2011.574562

382. Bernath V. *Hip Resurfacing in Patients with Osteoarthritis*. Clayton, Victoria: Centre for Clinical Effectiveness; 2002.

383. Beswick A, Wylde V, Blom A, Gooberman-Hill R, Dieppe PA. Pain after hip or knee joint replacement for osteoarthritis: a systematic review. *Arthritis Rheum* 2011;**63**:S413.

384. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;**2**:e000435. http://dx.doi.org/10.1136/bmjopen-2011-000435

385. Bisseling P, Smolders JMH, Hol A, Van Susante JLC. No clear influence of preference bias on satisfaction and early functional outcome in resurfacing hip arthroplasty. *Acta Orthop* 2011;**82**:161–5. http://dx.doi.org/10.3109/17453674.2011.566140

386. Boden H, Adolphson P. No adverse effects of early weight bearing after uncemented total hip arthroplasty: a randomized study of 20 patients. *Acta Orthop Scand* 2004;**75**:21–9. http://dx.doi.org/10.1080/00016470410001708040

387. Boe BG, Rohrl SM, Heier T, Snorrason F, Nordsletten L. A prospective randomized study comparing electrochemically deposited hydroxyapatite and plasma-sprayed hydroxyapatite on titanium stems. *Acta Orthop* 2011;**82**:13–19. http://dx.doi.org/10.3109/17453674.2010.548027

388. Butler RA, Rosenzweig S, Myers L, Barrack RL. The Frank Stinchfield Award: the impact of socioeconomic factors on outcome after THA: a prospective, randomized study. *Clin Orthop Relat Res* 2011;**469**:339–47. http://dx.doi.org/10.1007/s11999-010-1519-x

389. Cobb J. The functional outcome of hip resurfacing and large-head THA is the same: a randomized, double-blind study. *Clin Orthop Relat Res* 2010;**468**:3134. http://dx.doi.org/10.1007/s11999-010-1512-4

390. Corbett KL, Losina E, Nti AA, Prokopetz JJZ, Katz JN. Population-based rates of revision of primary total hip arthroplasty: a systematic review. *PLOS ONE* 2010;**5**:e13520. http://dx.doi.org/10.1371/journal.pone.0013520

391. D'Angelo F, Murena L, Zatti G, Cherubino P. The unstable total hip replacement. *Ind J Orthop* 2008;**42**:252–9. http://dx.doi.org/10.4103/0019-5413.39667

392. Digas G, Karrholm J, Thanner J. Addition of fluoride to acrylic bone cement does not improve fixation of a total hip arthroplasty stem. *Clin Orthop Relat Res* 2006;**448**:58–66. http://dx.doi.org/10.1097/01.blo.0000224014.35045.7b

393. Digas G, Thanner J, Anderberg C, Karrholm J. Fluoride-containing acrylic bone cement in total hip arthroplasty. Randomized evaluation of 97 stems using radiostereometry and dual-energy x-ray absorptiometry. *J Arthroplasty* 2005;**20**:784–92. http://dx.doi.org/10.1016/j.arth.2004.12.056

394. Edwards SJ, Hamilton V, Nherera L, Arber M. Systematic review of the impact different metal femoral stems (MFSS) have on patient outcomes in total hip replacement (THR) due to osteoarthritis (OA). *Value Health* 2011;**14**:A245. http://dx.doi.org/10.1016/j.jval.2011.08.081

395. Eingartner C, Piel S, Weise K. Results of a cemented straight titanium alloy femoral stem after mean follow-up of 13 years. *Eur J Orthop Surg Traumatol* 2007;**17**:587–93. http://dx.doi.org/10.1007/s00590-007-0232-2

396. Fenandez-Lopez JC, Gossec L, Dougados M. Magnitude of the symptomatic at 3, 6 and 12 months after total articular replacement in hip and knee osteoarthritis: a systematic review and meta-analysis. *Arthritis Rheum* 2008;**58**:S245–6.

397. Fick DP, Nivbrant B. Minimally invasive surgical approaches for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database Syst Rev* 2004;**2**:CD004798.

398. Freund KG, Houshian S, Riegels-Nielsen P. Occlusion and stability of two different femoral canal plugs in cemented hip arthroplasty. A prospective and randomized study, with a two year follow-up. *Hip Int* 2003;**13**:142–7.

399. Gallart X, Riba J, Garcia S, Combalia A, Esteban PL, Marmolejo C. Time saving during acrylic bone cement setting in femoral stem implantation of hip arthroplasty: a prospective, double-blind, randomised study. *Hip Int* 2005;**15**:143–8.

400. Hallan G, Aamodt A, Furnes O, Skredderstuen A, Haugan K, Havelin LI. Palamed G compared with Palacos R with gentamicin in Charnley total hip replacement. *J Bone Joint Surg Br* 2006;**88**:1143–8. http://dx.doi.org/10.1302/0301-620X.88B9.18008

401. Hamadouche M, Baque F, Lefevre N, Kerboull M. Minimum 10-year survival of Kerboull cemented stems according to surface finish. *Clin Orthop Relat Res* 2008;**466**:332–9. http://dx.doi.org/10.1007/s11999-007-0074-6

402. Haverkamp D, Klinkenbijl MN, Somford MP, Albers GHR, van der Vis HM. Obesity in total hip arthroplasty – does it really matter? A meta-analysis. *Acta Orthop* 2011;**82**:417–22. http://dx.doi.org/10.3109/17453674.2011.588859

403. HAYES, Inc. *Ganz Trochanteric Flip Osteotomy Approach to Hip Resurfacing for Treatment of Osteoarthritis*. Healthcare Technology Brief Publication. Lansdale, PA: HAYES, Inc.; 2012.

404. Hoebink E, Struijs Peter AA. Effects of different bearing surface materials on aseptic loosening of total hip arthroplasty in patients with osteoarthritis and other non-traumatic diseases of the hip. *Cochrane Database Syst Rev* 2008;**4**:CD007494.

405. Husby OS, Haugan K, Benum P, Foss OA. A prospective randomised radiostereometric analysis trial of SmartSet HV and Palacos R bone cements in primary total hip arthroplasty. *J Orthop Traumatol* 2010;**11**:29–35. http://dx.doi.org/10.1007/s10195-010-0087-x

406. Ise K, Kawanabe K, Tamura J, Akiyama H, Goto K, Nakamura T. Clinical results of the wear performance of cross-linked polyethylene in total hip arthroplasty. Prospective randomized trial. *J Arthroplasty* 2009;**24**:1216–20. http://dx.doi.org/10.1016/j.arth.2009.05.020

407. ISPOR 14th Annual European Congress Research Abstracts. *Value Health* 2011;**14**:A233–574.

408. Jager M, Begg MJ, Ready J, Bittersohl B, Millis M, Krauspe R, *et al.* Primary total hip replacement in childhood, adolescence and young patients: quality and outcome of clinical studies. *Technol Health Care* 2008;**16**:195–214.

409. Jandric S, Jovicic Z, Novakovic S. Differences in pain between women and men in patients with total hip arthroplasty. *Eur J Pain* 2009;**13**(Suppl. 1):S133. http://dx.doi.org/10.1016/S1090-3801(09)60446-8

410. Jandric S, Manojlovic S. Quality of life of men and women with osteoarthritis of the hip and arthroplasty assessment by WOMAC questionnaire. *Am J Phys Med Rehabil* 2009;**88**:328–35. http://dx.doi.org/10.1097/PHM.0b013e318194fa24

411. Jensen C, Aagaard P, Overgaard S. Recovery in horizontal gait after hip resurfacing vs. total hip arthroplasty at 6-month follow-up – a randomized clinical trial. *Osteoarthritis Cartilage* 2012;**20**(Suppl. 1):S99. http://dx.doi.org/10.1016/j.joca.2012.02.106

412. Jolles BM, Grzesiak A, Eudier A, Dejnabadi H, Voracek C, Pichonnaz C, *et al.* A randomised controlled clinical trial and gait analysis of fixed- and mobile-bearing total knee replacements with a five-year follow-up. *J Bone Joint Surg Br* 2012;**94**:648–55. http://dx.doi.org/10.1302/0301-620X.94B5.27598

413. Jolles BM, Michel J, Burnand B, Leyvraz P. Surgical treatment for advanced stage of avascular necrosis of the femoral head in adults. *Cochrane Database Syst Rev* 2006;**3**:CD006079.

414. Kadar T, Hallan G, Aamodt A, Indrekvam K, Badawy M, Havelin LI, *et al.* A randomized study on migration of the Spectron EF and the Charnley flanged 40 cemented femoral components using radiostereometric analysis at 2 years. *Acta Orthop* 2011;**82**:538–44. http://dx.doi.org/10.3109/17453674.2011.618914

415. Karas S. Outcomes of Birmingham hip resurfacing: a systematic review. *Asian J Sports Med* 2012;**3**:1–7.

416. Kenanidis EI, Potoupnis ME, Papavasiliou KA, Sayegh FE, Kapetanos GA. Re: Prospective randomized study of two surgical approaches for total hip arthroplasty. *J Arthroplasty* 2011;**26**:821. http://dx.doi.org/10.1016/j.arth.2011.02.017

417. Kim S, Losina E, Solomon DH, Wright J, Katz JN. Effectiveness of clinical pathways for total knee and total hip arthroplasty: literature review. *J Arthroplasty* 2003;**18**:69–74. http://dx.doi.org/10.1054/arth.2003.50030

418. Kim YH, Kim JS, Joo JH, Park JW. Is hydroxyapatite coating necessary to improve survivorship of porous-coated titanium femoral stem? *J Arthroplasty* 2012;**27**:559–63. http://dx.doi.org/10.1016/j.arth.2011.06.020

419. Kim YH, Oh JH. A comparison of a conventional versus a short, anatomical metaphyseal-fitting cementless femoral stem in the treatment of patients with a fracture of the femoral neck. *J Bone Joint Surg Br* 2012;**94**:774–81. http://dx.doi.org/10.1302/0301-620X.94B6.29152

420. Kim YH, Oh SW, Kim JS. Prevalence of fat embolism following bilateral simultaneous and unilateral total hip arthroplasty performed with or without cement: a prospective, randomized clinical study. *J Bone Joint Surg Am* 2002;**84**:1372–9.

421. Lane NE. Osteoarthritis of the hip. *New Engl J Med* 2007;**357**:1413–21. http://dx.doi.org/10.1056/NEJMcp071112

422. Laursen MB, Nielsen PT, Soballe K. Bone remodelling around HA-coated acetabular cups: a DEXA study with a 3-year follow-up in a randomised trial. *Int Orthop* 2007;**31**:199–204. http://dx.doi.org/10.1007/s00264-006-0148-1

423. Lavigne M, Masse V, Girard J, Roy AG, Vendittoli PA. [Return to sport after hip resurfacing or total hip arthroplasty: a randomized study]. *Rev Chir Orthop Reparatrice Appar Mot* 2008;**94**:361–7. http://dx.doi.org/10.1016/j.rco.2007.12.009

424. Mallmin H, Wolf O, Larsson S, Milbrink J, Mattsson P. Body composition and BMD after total hip arthroplasty. A randomised clinical trial of two different postoperative regimes with 5 years of follow-up. *Bone* 2011;**48**(Suppl. 2):S267. http://dx.doi.org/10.1016/j.bone.2011.03.658

425. Markmiller M, Weiss T, Kreuz P, Ruter A, Konrad G. Partial weightbearing is not necessary after cementless total hip arthroplasty: a two-year prospective randomized study on 100 patients. *Int Orthop* 2011;**35**:1139–43. http://dx.doi.org/10.1007/s00264-010-1089-2

426. Montin L, Leino-Kilpi H, Suominen T, Lepisto J. A systematic review of empirical studies between 1966 and 2005 of patient outcomes of total hip arthroplasty and related factors. *J Clin Nurs* 2008;**17**:40–5. http://dx.doi.org/10.1111/j.1365-2702.2007.01944.x

427. Moskal JT, Capps SG. Acetabular component positioning in total hip arthroplasty: an evidence-based analysis. *J Arthroplasty* 2011;**26**:1432–7. http://dx.doi.org/10.1016/j.arth.2010.11.011

428. Mouilhade F, Matsoukis J, Oger P, Mandereau C, Brzakala V, Dujardin F. Component positioning in primary total hip replacement: a prospective comparative study of two anterolateral approaches, minimally invasive versus gluteus medius hemimyotomy. *Orthop Traumatol Surg Res* 2011;**97**:14–21. http://dx.doi.org/10.1016/j.otsr.2010.05.013

429. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty? A systematic review. *Clin Orthop Relat Res* 2010;**468**:1891–904. http://dx.doi.org/10.1007/s11999-009-1135-9

430. Nantel J, Termoz N, Vendittoli PA, Lavigne M, Prince F. Gait patterns after total hip arthroplasty and surface replacement arthroplasty. *Arch Phys Med Rehabil* 2009;**90**:463–9. http://dx.doi.org/10.1016/j.apmr.2008.08.215

431. Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RG. Good diagnostic performance of early migration as a predictor of late aseptic loosening of acetabular cups: results from ten years of follow-up with Roentgen stereophotogrammetric analysis (RSA). *J Bone Joint Surg Am* 2012;**94**:874–80. http://dx.doi.org/10.2106/JBJS.K.00305

432. Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RG. The Exeter femoral stem continues to migrate during its first decade after implantation: 10–12 years of follow-up with radiostereometric analysis (RSA). *Acta Orthop* 2012;**83**:129–34. http://dx.doi.org/10.3109/17453674.2012.672093

433. Nygaard M, Elling F, Bastholm L, Soballe K, Borgwardt A. No difference in early cellular response of the pseudo-synovial membrane after total hip arthroplasty: comparison of 3 combinations of bearing materials. *Acta Orthop* 2006;**77**:402–12. http://dx.doi.org/10.1080/17453670610046325

434. Nygaard M, Zerahn B, Bruce C, Soballe K, Borgwardt A. Early periprosthetic femoral bone remodelling using different bearing material combinations in total hip arthroplasties: a prospective randomised study. *Eur Cells Mater* 2004;**8**:65–73.

435. Pabinger C, Kroner A, Lange A, Eyb R. Cemented titanium stems show high migration: transprosthetic drainage system has no advantage over third-generation cementation technique. *Arch Orthop Trauma Surg* 2004;**124**:489–94. http://dx.doi.org/10.1007/s00402-004-0705-2

436. Palm L, Jacobsson S-A, Ivarsson I. Hydroxyapatite coating improves 8- to 10-year performance of the Link RS cementless femoral stem. *J Arthroplasty* 2002;**17**:172–5. http://dx.doi.org/10.1054/arth.2002.29395

437. Patel D, Parvizi J, Sharkey PF. Alternative bearing surface options for revision total hip arthroplasty. *Instr Course Lect* 2011;**60**:257–67.

438. Petersen MK, Andersen NT, Soballe K. Self-reported functional outcome after primary total hip replacement treated with two different perioperative regimes: a follow-up study involving 61 patients. *Acta Orthop* 2008;**79**:160–7. http://dx.doi.org/10.1080/17453670710014932

439. Pitto RP, Hamer H, Fabiani R, Radespiel-Troeger M, Koessler M. Prophylaxis against fat and bone-marrow embolism during total hip arthroplasty reduces the incidence of postoperative deep-vein thrombosis: a controlled, randomized clinical trial. *J Bone Joint Surgery Am* 2002;**84**:39–48.

440. Pivec R, Johnson AJ, Mont MA. Results of total hip arthroplasty in patients who have rapidly progressive hip disease: a systematic review of the literature. *Expert Rev Med Devices* 2012;**9**:257–62. http://dx.doi.org/10.1586/erd.12.12

441. Prudhon JL. Dual-mobility cup and cemented femoral component: 6 year follow-up results. *Hip Int* 2011;**21**:713–17. http://dx.doi.org/10.5301/HIP.2011.8846

442. Rasanen P, Paavolainen P, Sintonen H, Koivisto AM, Blom M, Ryynanen OP, *et al.* Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthop* 2007;**78**:108–15. http://dx.doi.org/10.1080/17453670610013501

443. Rasquinha VJ, Ranawat CS, Dua V, Ranawat AS, Rodriguez JA. A prospective, randomized, double-blind study of smooth versus rough stems using cement fixation: minimum 5-year follow-up. *J Arthroplasty* 2004;**19**(Suppl. 2):2–9. http://dx.doi.org/10.1016/j.arth.2004.06.005

444. Ratko TA, Aronson N, Ziegler KM, Bonnell CJ. *Metal-on-Metal Total Hip Resurfacing*. Chicago, IL: Blue Cross and Blue Shield Association, Technology Evaluation Center. Assessment Program; 2007.

445. Renkawitz T, Santori FS, Grifka J, Valverde C, Morlock MM, Learmonth ID. A new short uncemented, proximally fixed anatomic femoral implant with a prominent lateral flare: design rationals and study design of an international clinical trial. *BMC Musculoskelet Disord* 2008;**9**:147. http://dx.doi.org/10.1186/1471-2474-9-147

446. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Care Res* 2008;**59**:876–83. http://dx.doi.org/10.1002/art.23706

447. Rohrl SM, Nivbrant B, Strom H, Nilsson KG. Effect of augmented cup fixation on stability, wear, and osteolysis: a 5-year follow-up of total hip arthroplasty with RSA. *J Arthroplasty* 2004;**19**:962–71.

448. Santaguida PL, Hawker GA, Hudak PL, Glazier R, Mahomed NN, Kreder HJ, *et al.* Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can J Surg* 2008;**51**:428–36.

449. Schauss SM, Hinz M, Mayr E, Bach CM, Krismer M, Fischer M. Inferior stability of a biodegradable cement plug. 122 total hip replacements randomized to degradable or non-degradable cement restrictor. *Arch Orthop Trauma Surg* 2006;**126**:324–9. http://dx.doi.org/10.1007/s00402-006-0132-7

450. Schmidutz F, Dull T, Voges O, Grupp T, Muller P, Jansson V. Secondary cement injection technique reduces pulmonary embolism in total hip arthroplasty. *Int Orthop* 2012;**36**:1575–81. http://dx.doi.org/10.1007/s00264-012-1537-2

451. Scott D. Osteoarthritis of the hip. *Clin Evid Handbook* December 2009:398–9.

452. Seyler TM, Bonutti PM, Shen J, Naughton M, Kester M. Use of an alumina-on-alumina bearing system in total hip arthroplasty for osteonecrosis of the hip. *J Bone Joint Surg Am* 2006;**88**(Suppl. 3):116–25. http://dx.doi.org/10.2106/JBJS.F.00775

453. Sharma V, Morgan PM, Cheng EY. Factors influencing early rehabilitation after THA: a systematic review. *Clin Orthop Relat Res* 2009;**467**:1400–11. http://dx.doi.org/10.1007/s11999-009-0750-9

454. Shetty V, Shitole B, Shetty G, Thakur H, Bhandari M. Optimal bearing surfaces for total hip replacement in the young patient: a meta-analysis. *Int Orthop* 2011;**35**:1281–7. http://dx.doi.org/10.1007/s00264-010-1104-7

455. Singh JA. Epidemiology of knee and hip arthroplasty: a systematic review. *Open Orthop J* 2011;**5**:80–5. http://dx.doi.org/10.2174/1874325001105010080

456. Singh JA, Kundukulam J, Riddle DL, Strand V, Tugwell P. Early postoperative mortality following joint arthroplasty: a systematic review. *J Rheumatol* 2011;**38**:1507–13. http://dx.doi.org/10.3899/jrheum.110280

457. Sluimer JC, Hoefnagels NH, Emans PJ, Kuijer R, Geesink RG. Comparison of two hydroxyapatite-coated femoral stems: clinical, functional, and bone densitometry evaluation of patients randomized to a regular or modified hydroxyapatite-coated stem aimed at proximal fixation. *J Arthroplasty* 2006;**21**:344–52. http://dx.doi.org/10.1016/j.arth.2005.06.015

458. Stanat SJC, Capozzi JD. Squeaking in third- and fourth-generation ceramic-on-ceramic total hip arthroplasty. Meta-analysis and systematic review. *J Arthroplasty* 2012;**27**:445–53. http://dx.doi.org/10.1016/j.arth.2011.04.031

459. Stilling M, Rahbek O, Soballe K. Inferior survival of hydroxyapatite versus titanium-coated cups at 15 years. *Clin Orthop Relat Res* 2009;**467**:2872–9. http://dx.doi.org/10.1007/s11999-009-0796-8

460. Suda AJ, Knahr K. Early results with the cementless Variall hip system. *Exp Rev Med Devices* 2009;**6**:21–5. http://dx.doi.org/10.1586/17434440.6.1.21

461. Tarasevicius S, Robertsson O, Wingstrand H. Posterior soft tissue repair in total hip arthroplasty: a randomized controlled trial. *Orthopedics* 2010;**33**:871. http://dx.doi.org/10.3928/01477447-20101021-11

462. ten Broeke RHM, Hendrickx RPM, Leffers P, Jutten LMC, Geesink RGT. Randomised trial comparing bone remodelling around two uncemented stems using modified Gruen zones. *Hip Int* 2012;**22**:41–9. http://dx.doi.org/10.5301/HIP.2012.9103

463. Thien TM, Thanner J, Karrholm J. Randomized comparison between 3 surface treatments of a single anteverted stem design: 84 hips followed for 5 years. *J Arthroplasty* 2010;**25**:437–44. http://dx.doi.org/10.1016/j.arth.2009.01.015

464. Thien TM, Thanner J, Karrholm J. Fixation and bone remodeling around a low-modulus stem. Seven-year follow-up of a randomized study with use of radiostereometry and dual-energy X-ray absorptiometer. *J Arthroplasty* 2012;**27**:134–42. http://dx.doi.org/10.1016/j.arth.2011.03.029

465. Thomas W, Tafuro L, Thomas S. Osteoinductive gel in cementless hip joint replacement: a randomized prospective study. *Curr Orthop Pract* 2009;**20**:655–9. http://dx.doi.org/10.1097/BCO.0b013e3181a56cff

466. Timperley AJ, Whitehouse SL, Hourigan PG. The influence of a suction device on fixation of a cemented cup using RSA. *Clin Orthop Relat Res* 2009;**467**:792–8. http://dx.doi.org/10.1007/s11999-008-0574-z

467. Ullmark G, Sorensen J, Nilsson O. Analysis of bone formation on porous and calcium phosphate-coated acetabular cups: a randomised clinical [18F]fluoride PET study. *Hip Int* 2012;**22**:172–8. http://dx.doi.org/10.5301/HIP.2012.9233

468. Vail TP, Goetz D, Tanzer M, Fisher DA, Mohler CG, Callaghan JJ. A prospective randomized trial of cemented femoral components with polished versus grit-blasted surface finish and identical stem geometry. *J Arthroplasty* 2003;**18**:95–102. http://dx.doi.org/10.1016/S0883-5403(03)00298-5

469. Van der Wal BC, Rahmy AI, Grimm B, Blake GM, Heyligers IC, Tonino AJ. The influence of implant design on periprosthetic bone remodelling of two types of uncemented HA-coated hip stems. A two-year follow-up study using DEXA. *Hip Int* 2006;**16**:8–17.

470. Van Der Weegen W, Hoekstra HJ, Sijbesma T, Bos E, Schemitsch EH, Poolman RW. Survival of metal-on-metal hip resurfacing arthroplasty: a systematic review of the literature. *J Bone Joint Surg Br* 2011;**93**:298–306. http://dx.doi.org/10.1302/0301-620X.93B3.25594

471. van Gerwen M, Shaerf DA, Veen RM. Hip resurfacing arthroplasty: a systematic review of functional outcome. *Acta Orthop* 2010;**81**:680–3. http://dx.doi.org/10.3109/17453674.2010.501742

472. Veldstra R, van Dongen A, Kraaneveld EC. Comparing alumina-reduced and conventional surface grit-blasted acetabular cups in primary THA: early results from a randomised clinical trial. *Hip Int* 2012;**22**:296–301. http://dx.doi.org/10.5301/HIP.2012.9244

473. Vendittoli PA, Ganapathi M, Duval N, Lavoie P, Roy A, Lavigne M. Randomised controlled trial comparing two methods of acetabular cup positioning during total hip arthroplasty. *Hip Int* 2007;**17**:137–42.

474. Vissers MM, Bussmann JB, Verhaar JA, Arends LR, Furlan AD, Reijman M. Recovery of physical functioning after total hip arthroplasty: systematic review and meta-analysis of the literature. *Phys Ther* 2011;**91**:615–29. http://dx.doi.org/10.2522/ptj.20100201

475. Vissers MM, Reijman M, Bussmann HB, Arends LR, Verhaar JA. Recovery of physical functioning after total hip arthroplasty: a systematic review of the literature. *Osteoarthritis Cartilage* 2009;**17**(Suppl. 1):S287–8. http://dx.doi.org/10.1016/S1063-4584(09)60558-6

476. Yamauchi Y, Jinno T, Koga D, Asou Y, Morita S, Okawa A. Comparison of different distal designs of femoral components and their effects on bone remodeling in 1-stage bilateral total hip arthroplasty. *J Arthroplasty* 2012;**27**:1538–43. http://dx.doi.org/10.1016/j.arth.2012.01.031

477. Zagra L, Bianchi L, Licari V, Champlon C, Giacometti CR. Gait analysis of THA with different head diameters: a prospective randomized study. *J Orthop Traumatol* 2011;**12**(Suppl.):S149–50.

478. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis. Part III: changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;**18**:476–99. http://dx.doi.org/10.1016/j.joca.2010.01.013

479. Zhang Y, Yang T-T, Zhou Y, Ma B-A. Comparison of postoperative curative effect and the possible survival rate of prosthesis following cemented and cementless total hip replacement. *Chin J Clin Rehabil* 2006;**10**:10.

480. Zwartele RE, Witjes S, Doets HC, Stijnen T, Poll RG. Cementless total hip arthroplasty in rheumatoid arthritis: a systematic review of the literature. *Arch Orthop Trauma Surg* 2012;**132**:535–46. http://dx.doi.org/10.1007/s00402-011-1432-0

Appendix 1 Search strategies for the reviews of clinical effectiveness, cost-effectiveness and registry data

Clinical effectiveness search strategies

MEDLINE via Ovid interface

Date range searched: 1946 to October week 4 2012.

Searched: 5 November 2012.

1. exp Arthroplasty, Replacement, Hip/ (15,246)
2. exp Hip Prosthesis/ (18,304)
3. (tha or thr).tw. (23,312)
4. exp Hip Joint/ (20,108)
5. exp Hip/ (8480)
6. hip.tw. (79,606)
7. ("femur head*" or "femoral head*" or acetabul*).tw. (20,571)
8. exp Femur Head/ (7700)
9. exp Acetabulum/ (8243)
10. 4 or 5 or 6 or 7 or 8 or 9 (97,057)
11. (arthroplast* or replace* or implant* or prosthes*).tw. (514,865)
12. exp Joint Prosthesis/ (33,736)
13. exp "Prostheses and Implants"/ (355,910)
14. 11 or 12 or 13 (716,289)
15. 10 and 14 (35,876)
16. (surf* or resurf*).tw. (629,176)
17. 10 and 16 (5573)
18. 1 or 2 or 3 or 15 or 17 (61,490)
19. exp Arthritis, Rheumatoid/ or exp Arthritis/ (190,095)
20. exp Osteoarthritis, Hip/ or exp Osteoarthritis/ (39,813)
21. (arthrit* or osteoarthrit* or osteoarthrosis or "rheumatoid arthrit*").tw. (141,102)
22. 19 or 20 or 21 (221,909)
23. 18 and 22 (7739)
24. meta analysis.pt. (37,222)
25. randomized controlled trial.pt. (340,101)
26. (random* or "controlled trial*" or "clinical trial*" or rct).tw. (718,263)
27. (metaanalys* or "meta analys*" or "meta-analys*").tw. (42,924)
28. "systematic review*".tw. (34,474)
29. 24 or 25 or 26 or 27 or 28 (846,326)
30. 23 and 29 (614)
31. limit 30 to (english language and yr="2002 -Current") (443)

MEDLINE(R) In-Process & Other Non-Indexed Citations via Ovid interface

Date range searched: 2 November 2012.

Searched: 5 November 2012.

1. (tha or thr).tw. (773)
2. (hip* or "femoral head*" or "femur head*" or acetabul*).tw. (9533)
3. (arthroplast* or replace* or implant* or prosthes*).tw. (26,517)
4. (surf* or resurf*).tw. (87,833)
5. 3 or 4 (111,797)
6. 2 and 5 (1866)
7. 1 or 6 (2325)
8. meta analysis.pt. (34)
9. randomized controlled trial.pt. (460)
10. (random* or "controlled trial*" or "clinical trial*" or rct).tw. (50,523)
11. "systematic review*".tw. (4605)
12. (metaanalys* or "meta analys*" or "meta-analys*").tw. (4148)
13. 8 or 9 or 10 or 11 or 12 (55,127)
14. 7 and 13 (192)

EMBASE via Ovid interface

Date range searched: 1974 to 2012 week 44.

Searched: 5 November 2012.

1. hip arthroplasty/ (37,471)
2. exp total hip prosthesis/ (20,316)
3. (tha or thr).tw. (25,515)
4. exp hip/ (29,775)
5. hip*.tw. (257,572)
6. exp femur head/ (9209)
7. exp acetabulum/ (8120)
8. ("femoral head*" or "femur head*" or acetabul*).tw. (25,366)
9. 4 or 5 or 6 or 7 or 8 (274,870)
10. (arthroplast* or replace* or implant* or prosthes*).tw. (661,628)
11. exp joint prosthesis/ (43,247)
12. exp prosthesis/ or exp implant/ or exp "prostheses and orthoses"/ (380,676)
13. 10 or 11 or 12 (872,492)
14. 9 and 13 (48,019)
15. (surf* or resurf*).tw. (809,046)
16. 9 and 15 (10,412)
17. 1 or 2 or 3 or 14 or 16 (80,859)
18. exp arthritis/ or exp chronic arthritis/ or exp rheumatoid arthritis/ (310,664)
19. exp hip osteoarthritis/ or exp osteoarthritis/ (72,791)
20. (arthrit* or osteoarthrit* or osteoarthroses or "rheumatoid arthrit*").tw. (189,738)
21. 18 or 19 or 20 (338,180)
22. 17 and 21 (10,696)
23. meta analysis/ (66,936)
24. randomized controlled trial/ (334,512)
25. (metaanalys* or "meta analys*" or "meta-analys*").tw. (60,497)

26. (random* or "controlled trial*" or "clinical trial*" or rct).tw. (968,002)
27. "systematic review*".tw. (47,500)
28. 23 or 24 or 25 or 26 or 27 (1,114,673)
29. 22 and 28 (837)
30. limit 29 to (english language and yr="2002 -Current") (645)

Web of Science (Science Citation Index and Conference Proceedings Citation Index – Science)

Searched: 9 November 2012.

#11 #10 AND #9 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (691)

#10 (TS=(random* or "clinical trial*" or "controlled trial*" or rct or "metaanalys*" or "meta analys*" or "meta-analys*" or "systematic review*")) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (778,898)

#9 #8 AND #7 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (3944)

#8 (TS=(arthriti* or osteoarthrit* or "rheumatoid arthrit*" or osteoarthrosis)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (106,592)

#7 #6 OR #1 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (39,025)

#6 #5 AND #2 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (28,795)

#5 #4 OR #3 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (1,604,641)

#4 (TS=(surf* or resurf*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (1,273,961)

#3 (TS=(arthroplast* or implant* or replace* or prosthes*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (374,918)

#2 (TS=(hip* or "femoral head*" or "femur head*" or acetabul*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (151,640)

#1 (TS=(tha or thr)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-09, Lemmatization=On* (12,821)

The Cochrane Library (including Cochrane Database of Systematic Reviews Cochrane Central Register of Controlled Trials, Database of Abstracts of Reviews of Effects, NHS Economic Evaluation Database and Health Technology Assessment database)

Searched: 9 November 2012.

#1 MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees (1296)

#2 MeSH descriptor: [Hip Prosthesis] explode all trees (949)

#3 tha or thr (664)

#4 MeSH descriptor: [Hip Joint] explode all trees (678)

#5 MeSH descriptor: [Hip] explode all trees (258)

#6 hip* or femoral next head* or femur next head* or acetabul* (10,895)

#7 MeSH descriptor: [Acetabulum] explode all trees (103)

#8 MeSH descriptor: [Femur Head] explode all trees (57)

#9 #4 or #5 or #6 or #7 or #8 (10,895)

#10 (arthroplast* or replace* or implant* or prosthes*) (31,347)

#11 MeSH descriptor: [Joint Prosthesis] explode all trees (1478)

#12 MeSH descriptor: [Prostheses and Implants] explode all trees (11,766)

#13 #10 or #11 or #12 (35,161)

#14 #9 and #13 (4051)

#15 (surf* or resurf*) (14,646)

#16 #9 and #15 (429)

#17 #1 or #2 or #3 or #14 or #16 (4422)

#18 MeSH descriptor: [Arthritis] explode all trees (7651)

#19 MeSH descriptor: [Arthritis, Rheumatoid] explode all trees (3887)

#20 MeSH descriptor: [Osteoarthritis] explode all trees (2998)

#21 MeSH descriptor: [Osteoarthritis, Hip] explode all trees (490)

#22 (arthrit* or osteoarthrit* or osteoarthrosis or rheumatoid next arthritit*) (12,010)

#23 #18 or #19 or #20 or #21 or #22 (12,530)

#24 #17 and #23 (663)

#25 #24 from 2002 (493)

Note: of the total 493 papers identified, we did not import three from The Cochrane Library

ClinicalTrials.gov
Searched: 9 November 2012.

(arthritis OR arthitic OR arthritics OR osteoarthritis OR osteoarthritic OR osteoarthritics OR "rheumatoid arthritis" OR "rheumatoid arthritic" OR "rheumatoid arthritics" OR osteoarthrosis OR osteoarthrotic) [ALL-FIELDS] AND ((tha OR thr) OR ((hip OR hips OR "femoral head" OR "femoral heads" OR "femur head" OR "femur heads" OR acetabulum OR acetabula OR acetabulums OR acetabular) AND (tha OR thr OR arthroplasty OR arthroplasties OR implant OR implants OR replace OR replacement OR replacements OR replaced OR replaces OR implant OR implants OR prosthesis OR prostheses OR surface OR surfacing OR surfaced OR resurface OR resurfacing OR resurfaced))) [ALL-FIELDS] (315)

UK Clinical Research Network Portfolio
Searched: 12 December 2012.

1. TOPIC: Musculoskeletal
2. SPECIALTY GROUP: All
3. RESEARCH SUMMARY: hip femur femoral acetabula acetabulum (36)

Current Controlled Trials
Searched: 12 December 2012.

(arthrit* OR osteoarthrit* OR osteoarthro*) AND ((THR OR THA) OR ((hip* OR femur* OR femor* OR acetabul*) AND (arthroplast* OR implant* OR replace* OR prosthes* OR surfac* OR resurfac*))) (73)

Cost-effectiveness search strategies

MEDLINE via Ovid interface
Date range of search: 1946 to November week 3 2012.

Searched: 26 November 2012.

1. exp Arthroplasty, Replacement, Hip/ (15,463)
2. exp Hip Prosthesis/ (18,505)
3. (tha or thr).tw. (23,441)
4. exp Hip Joint/ (20,462)
5. exp Hip/ (8619)
6. hip.tw. (80,726)
7. ("femur head*" or "femoral head*" or acetabul*).tw. (20,866)
8. exp Femur Head/ (7859)
9. exp Acetabulum/ (8399)
10. 4 or 5 or 6 or 7 or 8 or 9 (98,375)
11. (arthroplast* or replace* or implant* or prosthes*).tw. (518,308)
12. exp Joint Prosthesis/ (34,040)
13. exp "Prostheses and Implants"/ (360,484)

14. 11 or 12 or 13 (722,831)
15. 10 and 14 (36,336)
16. (surf* or resurf*).tw. (632,344)
17. 10 and 16 (5616)
18. 1 or 2 or 3 or 15 or 17 (62,056)
19. exp Arthritis, Rheumatoid/ or exp Arthritis/ (190,890)
20. exp Osteoarthritis, Hip/ or exp Osteoarthritis/ (40,142)
21. (arthrit* or osteoarthrit* or osteoarthrosis or "rheumatoid arthrit*").tw. (141,817)
22. 19 or 20 or 21 (222,917)
23. 18 and 22 (7861)
24. *Economics/ or exp *"economics, hospital"/ or *economics, medical/ or *economics, nursing/ (27,342)
25. exp *"Costs and Cost Analysis"/ (42,120)
26. exp *"Cost of Illness"/ (6779)
27. exp *"Models, Economic"/ (3078)
28. (cost* or economic*).ti. (96,110)
29. exp *"Quality of Life"/ (46,255)
30. exp *"Quality-Adjusted Life Years"/ (1297)
31. (ICER or qaly* or eq5d* or "eq-5d*" or euroqol or "euro-qol" or "quality of well-being" or "quality of wellbeing" or "short-form 36" or "shortform 36" or "36-item short-form" or "36-item short form" or "sf-36" or sf36 or "short-form 12" or "short form 12" or "12-item short-form" or "12-item short form" or "sf12" or "sf-12").ti. (1823)
32. ("Stanford Health Assessment Questionnaire" or HAQ or "Western Ontario and McMaster University Osteoarthritis Index" or WOMAC or OAKHQOL or JAQQ or PSAQoL).tw. (3222)
33. (markov or "time trade off" or "time-trade-off" or standard gamble or utilit* or qol or hrql or hrqol or disutilit* or "net-benefit analysis").ti. (18,006)
34. (quality adj2 life).ti. (32,969)
35. (decision adj2 model).ti. (454)
36. ("resource use" or "resource utili?ation").ti. (1506)
37. exp *Health Status/ (45,849)
38. ("health state*" or "health status").ti. (7441)
39. 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 (247,461)
40. 23 and 39 (500)
41. limit 40 to (english language and yr="2002 -Current") (348)

MEDLINE(R) In-Process & Other Non-Indexed Citations via Ovid interface

Date range of search: 21 November 2012.

Searched: 26 November 2012.

1. (tha or thr).tw. (816)
2. (hip* or "femoral head*" or "femur head*" or acetabul*).tw. (9790)
3. (arthroplast* or replace* or implant* or prosthes*).tw. (27,240)
4. (surf* or resurf*).tw. (88,816)
5. 3 or 4 (113,450)
6. 2 and 5 (1940)
7. 1 or 6 (2422)
8. (cost* or economic*).ti. (5677)
9. (ICER or qaly* or eq5d* or "eq-5d*" or euroqol or "euro-qol" or "quality of well-being" or "quality of wellbeing" or "short-form 36" or "shortform 36" or "36-item short-form" or "36-item short form" or "sf-36" or sf36 or "short-form 12" or "short form 12" or "12-item short-form" or "12-item short form" or "sf12" or "sf-12").ti. (60)
10. ("Stanford Health Assessment Questionnaire" or HAQ or "Western Ontario and McMaster University Osteoarthritis Index" or WOMAC or OAKHQOL or JAQQ or PSAQoL).tw. (230)

11. (markov or "time trade off" or "time-trade-off" or standard gamble or utilit* or qol or hrql or hrqol or disutilit* or "net-benefit analysis").ti. (1507)
12. (quality adj2 life).ti. (2232)
13. (decision adj2 model).ti. (23)
14. ("resource use" or "resource utili?ation").ti. (82)
15. ("health state*" or "health status").ti. (289)
16. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (9814)
17. 7 and 16 (82)

EMBASE via Ovid interface

Date range of search: 1974 to 2012 week 46.

Searched: 26 November 2012.

1. exp hip arthroplasty/ (37,516)
2. exp total hip prosthesis/ (20,340)
3. (tha or thr).tw. (25,540)
4. exp hip/ (29,872)
5. hip*.tw. (258,132)
6. exp femur head/ (9240)
7. exp acetabulum/ (8131)
8. ("femoral head*" or "femur head*" or acetabul*).tw. (25,429)
9. 4 or 5 or 6 or 7 or 8 (275,465)
10. (arthroplast* or replace* or implant* or prosthes*).tw. (663,172)
11. exp joint prosthesis/ (43,311)
12. exp prosthesis/ or exp implant/ or exp "prostheses and orthoses"/ (381,252)
13. 10 or 11 or 12 (872,557)
14. 9 and 13 (48,566)
15. (surf* or resurf*).tw. (810,448)
16. 9 and 15 (10,437)
17. 1 or 2 or 3 or 14 or 16 (81,422)
18. exp arthritis/ or exp chronic arthritis/ or exp rheumatoid arthritis/ (310,935)
19. exp hip osteoarthritis/ or exp osteoarthritis/ (72,812)
20. (arthrit* or osteoarthrit* or osteoarthroses or "rheumatoid arthrit*").tw. (189,771)
21. 18 or 19 or 20 (338,437)
22. 17 and 21 (10,767)
23. *health economics/ (16,427)
24. exp *economic evaluation/ (30,055)
25. exp *"health care cost"/ (44,344)
26. exp *"quality of life"/ (48,125)
27. exp *health status/ (38,138)
28. (cost* or economic*).ti. (128,093)
29. (ICER or qaly* or eq5d* or "eq-5d*" or euroqol or "euro-qol" or "quality of well-being" or "quality of wellbeing" or "short-form 36" or "shortform 36" or "36-item short-form" or "36-item short form" or "sf-36" or sf36 or "short-form 12" or "short form 12" or "12-item short-form" or "12-item short form" or "sf12" or "sf-12").ti. (2376)
30. ("Stanford Health Assessment Questionnaire" or HAQ or "Western Ontario and McMaster University Osteoarthritis Index" or WOMAC or OAKHQOL or JAQQ or PSAQoL).tw. (5548)
31. (markov or "time trade off" or "time-trade-off" or standard gamble or utilit* or qol or hrql or hrqol or disutilit* or "net-benefit analysis").ti. (25,983)
32. (quality adj2 life).ti. (47,695)
33. (decision adj2 model).ti. (583)
34. ("resource use" or "resource utili?ation").ti. (2176)

35. ("health state*" or "health status").ti. (8766)
36. 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 (300,476)
37. 22 and 36 (640)
38. limit 37 to (english language and yr="2002 -Current") (489)

Web of Science (Science Citation Index and Conference Proceedings Citation Index – Science)

Searched: 30 November 2012.

#18 #17 AND #9 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (520)

#17 #16 OR #15 OR #14 OR #13 OR #12 OR #11 OR #10 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (169,247)

#16 (TI=("health state*" or "health status")) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (3487)

#15 (TI=("resource use" or "resource utili?ation")) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (2194)

#14 (TI=(decision NEAR/2 model)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (2366)

#13 (TI=(quality NEAR/2 life)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (31,919)

#12 (TI=(markov or "time trade off" or "time-trade-off" or standard gamble or utilit* or qol or hrql or hrqol or disutilit* or "net-benefit analysis")) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (31,421)

#11 (TS=("Stanford Health Assessment Questionnaire" or HAQ or "Western Ontario and McMaster University Osteoarthritis Index" or WOMAC or OAKHQOL or JAQQ or PSAQoL)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (2797)

#10 (TI=(cost* or economic* or ICER or qaly* or eq5d* or "eq-5d*" or euroqol or "euro-qol" or "quality of well-being" or "quality of wellbeing" or "short-form 36" or "shortform 36" or "36-item short-form" or "36-item short form" or "sf-36" or sf36 or "short-form 12" or "short form 12" or "12-item short-form" or "12-item short form" or "sf12" or "sf-12")) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (100,820)

#9 #8 AND #7 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (3969)

#8 (TS=(arthriti* or osteoarthrit* or "rheumatoid arthrit*" or osteoarthrosis)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (108,491)

#7 #6 OR #1 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (39,281)

#6 #5 AND #2 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (28,995)

#5 #4 OR #3 *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (1,617,466)

#4 (TS=(surf* or resurf*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (1,284,032)

#3 (TS=(arthroplast* or implant* or replace* or prosthes*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (378,013)

#2 (TS=(hip* or "femoral head*" or "femur head*" or acetabul*)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (152,745)

#1 (TS=(tha or thr)) AND Language=(English) *Databases=SCI-EXPANDED, CPCI-S Timespan=2002-01-01 - 2012-11-30*, *Lemmatization=On* (12,896)

NHS Economic Evaluation Database and Health Technology Assessment database, via The Cochrane Library

Searched: 30 November 2012.

#1 MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees (1297)

#2 MeSH descriptor: [Hip Prosthesis] explode all trees (949)

#3 tha or thr (667)

#4 MeSH descriptor: [Hip Joint] explode all trees (679)

#5 MeSH descriptor: [Hip] explode all trees (258)

#6 hip* or femoral next head* or femur next head* or acetabul* (10,914)

#7 MeSH descriptor: [Acetabulum] explode all trees (103)

#8 MeSH descriptor: [Femur Head] explode all trees (57)

#9 #4 or #5 or #6 or #7 or #8 (10,914)

#10 (arthroplast* or replace* or implant* or prosthes*) (31,428)

#11 MeSH descriptor: [Joint Prosthesis] explode all trees (1478)

#12 MeSH descriptor: [Prostheses and Implants] explode all trees (11,775)

#13 #10 or #11 or #12 (35,244)

#14 #9 and #13 (4060)

#15 (surf* or resurf*) (14,670)

#16 #9 and #15 (432)

#17 #1 or #2 or #3 or #14 or #16 (4435)

#18 MeSH descriptor: [Arthritis] explode all trees (7660)

#19 MeSH descriptor: [Arthritis, Rheumatoid] explode all trees (3891)

#20 MeSH descriptor: [Osteoarthritis] explode all trees (3004)

#21 MeSH descriptor: [Osteoarthritis, Hip] explode all trees (490)

#22 (arthrit* or osteoarthrit* or osteoarthrosis or rheumatoid next arthritit*) (12,032)

#23 #18 or #19 or #20 or #21 or #22 (12,552)

#24 #17 and #23 (666)

#25 #24 from 2002 (Word variations have been searched) (496)

Note: of the total 496, we imported 30 from NHS EED and eight from the HTA database.

Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials and Database of Abstracts of Reviews of Effects via The Cochrane Library

Searched: 30 November 2012.

#1 MeSH descriptor: [Arthroplasty, Replacement, Hip] explode all trees (1297)

#2 MeSH descriptor: [Hip Prosthesis] explode all trees (949)

#3 tha or thr (667)

#4 MeSH descriptor: [Hip Joint] explode all trees (679)

#5 MeSH descriptor: [Hip] explode all trees (258)

#6 hip* or femoral next head* or femur next head* or acetabul* (10,914)

#7 MeSH descriptor: [Acetabulum] explode all trees (103)

#8 MeSH descriptor: [Femur Head] explode all trees (57)

#9 #4 or #5 or #6 or #7 or #8 (10,914)

#10 (arthroplast* or replace* or implant* or prosthes*) (31,428)

#11 MeSH descriptor: [Joint Prosthesis] explode all trees (1478)

#12 MeSH descriptor: [Prostheses and Implants] explode all trees (11,775)

#13 #10 or #11 or #12 (35,244)

#14 #9 and #13 (4060)

#15 (surf* or resurf*) (14,670)

#16 #9 and #15 (432)

#17 #1 or #2 or #3 or #14 or #16 (4435)

#18 MeSH descriptor: [Arthritis] explode all trees (7660)

#19 MeSH descriptor: [Arthritis, Rheumatoid] explode all trees (3891)

#20 MeSH descriptor: [Osteoarthritis] explode all trees (3004)

#21 MeSH descriptor: [Osteoarthritis, Hip] explode all trees (490)

#22 (arthrit* or osteoarthrit* or osteoarthrosis or rheumatoid next arthritit*) (12,032)

#23 #18 or #19 or #20 or #21 or #22 (12,552)

#24 #17 and #23 (666)

#25 #24 from 2002 (Word variations have been searched) (496)

#26 MeSH descriptor: [Economics] this term only (50)

#27 MeSH descriptor: [Economics, Hospital] explode all trees (1385)

#28 MeSH descriptor: [Economics, Medical] this term only (33)

#29 MeSH descriptor: [Economics, Nursing] this term only (15)

#30 MeSH descriptor: [Costs and Cost Analysis] explode all trees (18,514)

#31 MeSH descriptor: [Models, Economic] explode all trees (1457)

#32 (cost* or economic*):ti,ab (30,879)

#33 MeSH descriptor: [Quality of Life] explode all trees (12,081)

#34 MeSH descriptor: [Quality-Adjusted Life Years] explode all trees (2768)

#35 (ICER or qaly* or eq5d* or eq-5d* or euroqol or euro-qol or "quality of well-being" or "quality of wellbeing" or "short-form 36" or "shortform 36" or "36-item short-form" or "36-item short form" or "sf-36" or sf36 or "short-form 12" or "short form 12" or "12-item short-form" or "12-item short form" or "sf12" or "sf-12"):ti,ab (4323)

#36 ("Stanford Health Assessment Questionnaire" or HAQ or "Western Ontario and McMaster University Osteoarthritis Index" or WOMAC or OAKHQOL or JAQQ or PSAQoL):ti,ab (845)

#37 (markov or "time trade off" or "time-trade-off" or "standard gamble" or utilit* or qol or hrql or hrqol or disutilit* or "net-benefit analysis"):ti,ab (8363)

#38 (quality near/2 life):ti,ab (18,690)

#39 (decision near/2 model):ti,ab (281)

#40 ("resource use" or resource next utili?ation):ti,ab (1048)

#41 MeSH descriptor: [Health Status] explode all trees (4568)

#42 (health next state* or "health status"):ti,ab (2397)

#43 #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 or #38 or #39 or #40 or #41 or #42 (62,008)

#44 #25 and #43 (178)

Note: of the total 178, we imported 64 from CDSR, five from DARE and 78 from CENTRAL.

Cost-effectiveness Analysis Registry (Articles)

Searched: 12 December 2012.

Hip (159)

Femoral (18)

Acetabula (2)

Acetabulum (0)

Femur (0)

Note: 26 references were selected across all of the searches.

ClinicalTrials.gov

Searched: 9 November 2012.

(arthritis OR arthitic OR arthritics OR osteoarthritis OR osteoarthritic OR osteoarthritics OR "rheumatoid arthritis" OR "rheumatoid arthritic" OR "rheumatoid arthritics" OR osteoarthrosis OR osteoarthrotic) [ALL-FIELDS] AND ((tha OR thr) OR ((hip OR hips OR "femoral head" OR "femoral heads" OR "femur head" OR "femur heads" OR acetabulum OR acetabula OR acetabulums OR acetabular) AND (tha OR thr OR arthroplasty OR arthroplasties OR implant OR implants OR replace OR replacement OR replacements OR replaced OR replaces OR implant OR implants OR prosthesis OR prostheses OR surface OR surfacing OR surfaced OR resurface OR resurfacing OR resurfaced))) [ALL-FIELDS] (315)

Saved search: Repeat the search by clicking on the following link: www.clinicaltrials.gov/ct2/results?term=arthritis+OR+arthitic+OR+arthritics+OR+osteoarthritis+OR+osteoarthritic+OR+osteoarthritics+OR+%22rheumatoid+arthritis%22+OR+%22rheumatoid+arthritic%22+OR+%22rheumatoid+arthritics%22+OR+osteoarthrosis+OR+osteoarthrotic&sntx=%28+tha+OR+thr+OR+%28+hip+OR+hips+OR+%22femoral+head%22+OR+%22femoral+heads%22+OR+%22femur+head%22+OR+%22femur+heads%22+OR+acetabulum+OR+acetabula+OR+acetabulums+OR+acetabular+%29+AND+%28+tha+OR+thr+OR+arthroplasty+OR+arthroplasties+OR+implant+OR+implants+OR+replace+OR+replacement+OR+replacements+OR+replaced+OR+replaces+OR+implant+OR+implants+OR+prosthesis+OR+prostheses+OR+surface+OR+surfacing+OR+surfaced+OR+resurface+OR+resurfacing+OR+resurfaced+%29+%29+%5BALL-FIELDS%5D&show_xprt=Y

UK Clinical Research Network Portfolio
Searched: 12 December 2012.

1. TOPIC: Musculoskeletal
2. SPECIALTY GROUP: All
3. RESEARCH SUMMARY: hip femur femoral acetabula acetabulum (36)

Current Controlled Trials
Searched: 12 December 2012.

(arthrit* OR osteoarthrit* OR osteoarthro*) AND ((THR OR THA) OR ((hip* OR femur* OR femor* OR acetabul*) AND (arthroplast* OR implant* OR replace* OR prosthes* OR surfac* OR resurfac*))) (73)

Saved search: Repeat the search by clicking on the following link: www.controlled-trials.com/isrctn/search.html?srch=%28arthrit*+OR+osteoarthrit*+OR+osteoarthro*%29+AND+%28%28THR+OR+THA%29+OR+%28%28hip*+OR+femur*+OR+femor*+OR+acetabul*%29+AND+%28arthroplast*+OR+implant*+OR+replace*+OR+prosthes*+OR+surfac*+OR+resurfac*%29%29%29&sort=3&dir=desc&max=10&Submit=SUBMIT

The National Library of Medicine Gateway
Searched: 12 December 2012.

(hip OR tha OR thr OR femur OR femor* OR acetabul*) (87 HSRProj references)

Registry search strategies

MEDLINE via Ovid interface
Date range of search: 1946 to November week 3 2012.

Searched: 4 December 2012.

1. exp Arthroplasty, Replacement, Hip/ (15,469)
2. exp Hip Prosthesis/ (18,508)
3. (tha or thr).tw. (23,458)
4. exp Hip Joint/ (20,473)
5. exp Hip/ (8627)
6. hip.tw. (80,767)
7. ("femur head*" or "femoral head*" or acetabul*).tw. (20,872)
8. exp Femur Head/ (7861)
9. exp Acetabulum/ (8400)
10. 4 or 5 or 6 or 7 or 8 or 9 (98,424)
11. (arthroplast* or replace* or implant* or prosthes*).tw. (518,589)
12. exp Joint Prosthesis/ (34,047)
13. exp "Prostheses and Implants"/ (360,616)
14. 11 or 12 or 13 (723,184)
15. 10 and 14 (36,345)
16. (surf* or resurf*).tw. (632,737)
17. 10 and 16 (5621)
18. 1 or 2 or 3 or 15 or 17 (62,084)
19. exp Arthritis, Rheumatoid/ or exp Arthritis/ (190,963)

20. exp Osteoarthritis, Hip/ or exp Osteoarthritis/ (40,163)
21. (arthrit* or osteoarthrit* or osteoarthrosis or "rheumatoid arthrit*").tw. (141,888)
22. 19 or 20 or 21 (223,011)
23. 18 and 22 (7863)
24. exp Registries/ (50,293)
25. (registry or registries).tw. (48,917)
26. (register or registers).tw. (34,521)
27. Databases as Topic/ (7953)
28. Databases, Factual/ (37,658)
29. 24 or 25 or 26 or 27 or 28 (145,740)
30. 23 and 29 (245)
31. limit 30 to (english language and yr="2002 -Current") (208)

EMBASE via Ovid interface

Date range of search: 1947 to 3 December 2012.

Searched: 4 December 2012.

1. exp hip arthroplasty/ (38,263)
2. exp total hip prosthesis/ (20,654)
3. (tha or thr).tw. (26,605)
4. exp hip/ (38,923)
5. hip*.tw. (271,815)
6. exp femur head/ (11,130)
7. exp acetabulum/ (9541)
8. ("femoral head*" or "femur head*" or acetabul*).tw. (28,308)
9. 4 or 5 or 6 or 7 or 8 (291,434)
10. (arthroplast* or replace* or implant* or prosthes*).tw. (710,537)
11. exp joint prosthesis/ (44,090)
12. exp prosthesis/ or exp implant/ or exp "prostheses and orthoses"/ (399,853)
13. 10 or 11 or 12 (926,495)
14. 9 and 13 (50,948)
15. (surf* or resurf*).tw. (856,383)
16. 9 and 15 (11,107)
17. 1 or 2 or 3 or 14 or 16 (85,465)
18. exp arthritis/ or exp chronic arthritis/ or exp rheumatoid arthritis/ (339,336)
19. exp hip osteoarthritis/ or exp osteoarthritis/ (79,262)
20. (arthrit* or osteoarthrit* or osteoarthroses or "rheumatoid arthrit*").tw. (205,812)
21. 18 or 19 or 20 (367,580)
22. 17 and 21 (11,575)
23. register/ (51,419)
24. (registry or registries).tw. (72,626)
25. (register or registers).tw. (45,675)
26. exp factual database/ (37,382)
27. 23 or 24 or 25 or 26 (164,947)
28. 22 and 27 (289)
29. limit 28 to (yr="2002 -Current") (263)

MEDLINE(R) In-Process & Other Non-Indexed Citations via Ovid interface

Date range of search: 6 December 2012.

Searched: 7 December 2012.

1. (tha or thr).tw. (869)
2. (hip* or "femoral head*" or "femur head*" or acetabul*).tw. (10,378)
3. (arthroplast* or replace* or implant* or prosthes*).tw. (28,722)
4. (surf* or resurf*).tw. (91,209)
5. 3 or 4 (117,173)
6. 2 and 5 (2040)
7. 1 or 6 (2552)
8. (registry or registries).tw. (3295)
9. (register or registers).tw. (2122)
10. 8 or 9 (5284)
11. 7 and 10 (70)

Appendix 2 Quality assessment of included randomised controlled trials and systematic reviews

Quality assessment of included randomised controlled trials: studies comparing different types of total hip replacement ($n = 13$)

Quality assessment was carried out using The Cochrane Collaboration's tool for assessing risk of bias for a RCT (adapted from Higgins *et al.*[100]).

Cup fixation

Angadi *et al.*[112]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Randomised using a number generation program	Low risk of bias
	Allocation concealment		Assignment card in a sealed opaque envelope. This envelope was opened before surgery	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Awareness of cup fixation unlikely to influence the result	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Awareness of cup fixation unlikely to influence the result	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Small attrition rate (< 5%)	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All specified outcomes in the protocol were reported in the results section – survival was unclear	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Source of funding unclear	Unclear risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Unclear risk of bias
Objective (list of outcomes): aseptic loosening, implant dislocation, implant survival rate, infection, osteolysis, revision rate	Low risk of bias

Bjørgul *et al.*[110,111]
Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Randomisation was concealed	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Patients were blinded to the allocated treatment (cup fixation method) until the 2-year follow-up	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Knowledge of the allocated treatment less likely to influence these outcomes	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Assessor not directly involved in the study	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	30% of the initial sample not analysed	High risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	See above	High risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All specified outcomes reported	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Baseline imbalance in Charnley classification; no ITT analysis	High risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Unclear risk of bias
Objective (list of outcomes): revision, implant survival, cup migration	Low risk of bias

Cup liner bearing surface

McCalden *et al.*[145]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Not described	Unclear risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Blinded study nurse who measured clinical performance with a number of different validated clinical outcomes	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	See above	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Radiographic analysis was performed by a single individual with considerable expertise with the technique who was blinded to the type of polyethylene that had been used for each patient	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Nine patients died and two were lost to follow-up so did not complete the questionnaire	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes are reported in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Financial support for this study was provided by Zimmer, Inc. (Warsaw, IN, USA)	High risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS, WOMAC score, SF-12 score	Unclear risk of bias
Objective (list of outcomes): femoral head penetration (mm/year), osteolysis, infection	Low risk of bias

Engh *et al.*[113,114]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Not described	Unclear risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described	Unclear risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Blinded independent assessor	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	10-year outcome data were available for 78% and 63% of the patients in the two treatment groups, respectively; differential attrition rates may have affected the results	High risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	See above	High risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes reported	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Not described	Unclear risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): pain severity, function, mobility	Unclear risk of bias
Objective (list of outcomes): head penetration, osteolysis, revisions, dislocation	Unclear risk of bias

Cup shell design

Capello *et al.*,[115] D'Antonio *et al.*[116,117] and Mesko *et al.*[118]
Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Not described	Unclear risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	The knowledge of the treatment is less likely to influence these outcomes	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	The knowledge of the treatment is less likely to influence these outcomes	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	8% of data incomplete	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		Head penetration specified in the methods section not reported in the results section	High risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Industry funded	High risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Unclear risk of bias
Objective (list of outcomes): implant survival rate, revision, head penetration, osteolysis, dislocation, fractures	Low risk of bias

Cup/stem fixation

Corten *et al.*,[119,122] Bourne and Corten[121] and Laupacis *et al.*[120]
Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Randomisation was computer generated	Low risk of bias
	Allocation concealment		An opaque envelope that indicated the type of prosthesis to be used was opened in the operating room	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Throughout the course of the study, neither the patient nor the research assistant assessing the outcomes was aware of the type of prosthesis that had been inserted	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	See above	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	See above	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Small loss to follow-up (< 15%)	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	Small loss to follow-up (< 15%)	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes reported	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No other concerns	Low risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS, WOMAC score, MACTAR score	Low risk of bias
Objective (list of outcomes): implant revision-free survival, revision rate	Low risk of bias

Femoral head size

Howie *et al.*[123]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Computer-generated random numbers; blocked randomisation	Low risk of bias
	Allocation concealment		Sealed envelopes	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	NA	NA
		Objective (e.g. mortality, radiography, dislocation)	The awareness of the implant size would not influence the results for dislocation, revision and mortality	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	NA	NA
		Objective (e.g. mortality, radiography, dislocation)	The awareness of the implant size would not influence the results for dislocation, revision and mortality	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	NA	NA
		Objective outcomes (e.g. mortality, radiography, dislocation)	Incomplete data because of 17 (3.0%) excluded patients was not substantial enough to influence the results	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes reported in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No major concerns	Low risk of bias

ITT, intention to treat; NA, not applicable; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): NR	NA
Objective (list of outcomes): revision rate, mortality, implant dislocation	Low risk of bias

NA, not applicable; NR, not reported.

Femoral head bearing

Lewis *et al.*[124]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Sealed envelopes	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described but unlikely to influence the results	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described but unlikely to influence the results	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Not described	Unclear risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	Not described	Unclear risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		None of the outcomes reported in the results section was mentioned in the methods section	High risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Not described	Unclear risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS, WOMAC score, SF-12 (physical and mental components) score	Unclear risk of bias
Objective (list of outcomes): stem survival rate (before dislocation), dislocation, loosening	Low risk of bias

Femoral head bearing-on-cup liner bearing

Amanatullah *et al.*[125]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Unclear methods of randomisation	Unclear risk of bias
	Allocation concealment		'Because the randomization process used sealed envelopes' (p. 73)	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described although the knowledge of treatment would not have influenced the result	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Radiographs were evaluated by a surgeon and then a radiologist who was blinded to intervention type; wear rates were assessed by observers blinded to intervention type	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Not described	Unclear risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	'There was no statistical difference ($p > 0.05$) in the rate of attrition from either group at any time interval' (p. 74) Limited information on the patients lost to follow-up Only present data for 61.6% of the sample at 5 years At the final follow-up, 39 hips were excluded from the linear and volumetric wear rate analysis because of poor radiographic quality	High risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		Only report 5-year data despite collecting data at other time points	High risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP) baseline imbalance in important characteristics		Significant age differences between groups	High risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Unclear risk of bias
Objective (list of outcomes): revision rate, osteolysis, infection, deep-vein thrombosis, dislocation, pulmonary embolus	Low risk of bias

Capello *et al.*,[115] D'Antonio *et al.*[116,117] and Mesko *et al.*[118]

See *Cup shell design* for risk of bias of this RCT.

Kadar 2011[126]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Sealed envelopes	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Blinded participants	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described but unlikely to influence the outcome ascertainment	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described but unlikely to influence the outcome ascertainment	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)		Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	15% of data incomplete	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	15% of data incomplete	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes presented in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No major concerns; small sample size	Low risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Low risk of bias
Objective (list of outcomes): femoral head penetration	Low risk of bias

Stem composition

Healy *et al.*[127]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Patients were prospectively assigned to two treatment groups using medical record numbers as follows: odd numbers received a cobalt–chromium femoral stem and even numbers received a titanium femoral stem	High risk of bias
	Allocation concealment		NA	High risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Not described but the results are unlikely to be influenced by the knowledge of the treatment group	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Examiners were not blinded to the type of femoral implant used	High risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Examiners were not blinded to the type of femoral implant used but the results are unlikely to be influenced by the knowledge of the treatment group	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	10% of data incomplete	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	5–10% of data incomplete	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		WOMAC and SF-36 scores specified in the methods section were not reported in the results section	High risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No important concerns	Low risk of bias

ITT, intention to treat; NA, not applicable; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): pain, WOMAC score, SF-36 score and HHS	High risk of bias (lack of blinding, selective outcome reporting)
Objective (list of outcomes): revision rate, stem survival and complications (dislocation, fracture, haematoma, infection)	Unclear risk of bias (although the knowledge of treatment is unlikely to influence the results for these outcomes, there was no true randomisation and it is unclear what would be the combined effect of these two factors)

Stem design

Kim *et al.*[128]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Computer generated	Low risk of bias
	Allocation concealment		Not described	Unclear risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Results unlikely to be influenced by the knowledge of the femoral component implanted	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Results unlikely to be influenced by the knowledge of the femoral component implanted	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	No losses to follow-up	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	No losses to follow-up	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		VAS (10-point) pain scores not reported in the results section	High risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No important concerns	Low risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS	Unclear risk of bias
Objective (list of outcomes): revision rate, mortality	Low risk of bias

Stem fixation

Kim *et al.*[129]
Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		Adequate concealment – envelope opened in the operating room before the skin incision had been made	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Awareness of the treatment is unlikely to influence the outcomes	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Awareness of the treatment is unlikely to influence the outcomes	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Analysed the same sample (95%) for all outcomes	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	Analysed the same sample (95%) for all outcomes	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes reported in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Source of funding unclear	Unclear risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS, pain score (VAS), UCLA activity score, WOMAC score	Unclear risk of bias
Objective (list of outcomes): osteolysis, aseptic loosening, infection, implant dislocation, revision rate, time to revision, implant survival rate	Low risk of bias

Guide for assessing summary risk of bias for an outcome within a study across domains

Risk of bias across key domains	Interpretation	Summary risk of bias
Low risk of bias for all key domains	Plausible bias unlikely to seriously alter the results	Low risk of bias
Unclear risk of bias for at least two key domains	Plausible bias that raises some doubt about the results	Unclear risk of bias
High risk of bias for at least two key domains	Plausible bias that seriously weakens confidence in the results	High risk of bias

Key (most important) domains: selection bias, performance bias and detection bias.

Quality assessment of included systematic reviews: studies comparing different types of total hip replacement ($n = 5$)

Cup fixation

Quality assessment criteria for systematic reviews: The AMSTAR tool for assessing methodological quality of systematic reviews

Name of first reviewer: Alexander Tsertsvadze
Name of second reviewer: Paul Sutcliffe
First author surname year of publication: Voigt 2012[134]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes ☐ No ☒ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?

The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☒ **High quality**

☐ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Name of first reviewer: Paul Sutcliffe
Name of second reviewer: Alexander Tsertsvadze
First author surname year of publication: Pakvis 2011[135]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☐ **High quality**

☒ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Name of first reviewer: Paul Sutcliffe
Name of second reviewer: Alexander Tsertsvadze
First author surname year of publication: Clement 2012[136]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☐ **High quality**

☐ **Medium quality**

☒ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Femoral head bearing-on-cup liner bearing

Name of first reviewer: Alexander Tsertsvadze
Name of second reviewer: Paul Sutcliffe
First author surname year of publication: Sedrakyan 2011[137]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☒ **High quality**

☐ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Name of first reviewer: Alexander Tsertsvadze
Name of second reviewer: Paul Sutcliffe
First author surname year of publication: Yoshitomi 2009[138]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes ☐ No ☒ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☐ No ☐ Can't answer ☒ Not applicable

9. Were the methods used to combine the findings of studies appropriate?

For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☐ **High quality**

☒ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Quality assessment of included randomised controlled trials: studies comparing total hip replacement with resurfacing arthroplasty (*n* = 3)

Quality assessment was carried out using The Cochrane Collaboration's tool for assessing risk of bias for a RCT (adapted from Higgins *et al.*[100]).

Costa et al.[130] *and Achten* et al.[107]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Computer-generated random numbers and stratified by the supervising orthopaedic surgeon to balance any potential surgeon effects	Low risk of bias
	Allocation concealment		Investigators blinded to the sequence of allocation	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Patients/personnel were aware of the allocated treatment and this was likely to influence hip function and quality of life assessments	High risk of bias
		Objective (e.g. mortality, radiography, dislocation)	These outcomes (e.g. dislocation) unlikely to be influenced by the knowledge of the allocation group	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Assessors were unaware of the treatment allocation	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	Assessors were unaware of the treatment allocation	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	Small attrition rate (< 5%)	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	Small attrition rate (< 5%)	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All specified outcomes in the protocol were reported in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Research for the Patient Benefit scheme of the National Institute for Health Research; ITT analysis used; no baseline imbalances present	Low risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): HHS, OHS, Disability Rating Index, health-related quality of life (EQ-5D score)	Low risk of bias
Objective (list of outcomes): complications (implant dislocation, infection, deep-vein thrombosis)	Low risk of bias

Garbuz et al.[131]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Not described	Unclear risk of bias
	Allocation concealment		The assignments were contained in sealed envelopes and were opened the day before surgery by the study co-ordinator to allow for proper set-up in the operating room	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	The patients, nurses and physiotherapists were blinded to their assignment	Low risk of bias
		Objective (e.g. mortality, radiography, dislocation)	NA	NA
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	NA	NA
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	8% of data incomplete	Low risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	NA	NA
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All prespecified outcomes reported in the results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		Industry funded	High risk of bias

ITT, intention to treat; NA, not applicable; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): WOMAC, SF-36 and UCLA activity scores	Unclear risk of bias
Objective (list of outcomes): NR	NA

NA, not applicable; NR, not reported.

Vendittoli et al.,[132,133,136] *Girard* et al.[134] *and Rama* et al.[135]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Bias domain	Source of bias		Support for judgement[a]	Authors' judgement[b]
Selection bias	Random sequence generation		Computer-generated random number table	Low risk of bias
	Allocation concealment		Both surgeons and patients were kept blinded to the randomisation group until the morning of surgery	Low risk of bias
Performance bias	Blinding of participants and personnel	Subjective (e.g. patient reported)	Patients and personnel knew the treatment that they were getting	High risk of bias
		Objective (e.g. mortality, radiography, dislocation)	This knowledge would not influence the results	Low risk of bias
Detection bias	Blinding of outcome assessors	Subjective (e.g. patient reported)	Not described	Unclear risk of bias
		Objective (e.g. mortality, radiography, dislocation)	This knowledge would not influence the results	Low risk of bias
Attrition bias	Incomplete outcome data	Subjective outcomes (e.g. patient reported)	About 30% incomplete data	High risk of bias
		Objective outcomes (e.g. mortality, radiography, dislocation)	No incomplete data	Low risk of bias
Reporting bias	Selective reporting of the outcome, subgroups or analysis		All outcomes prespecified in the methods section are reported in results section	Low risk of bias
Other bias	Funding source, adequacy of statistical methods used, type of analysis (ITT/PP), baseline imbalance in important characteristics		No major concern	Low risk of bias

ITT, intention to treat; PP, per protocol.
a Statement, description or quote supporting the judgement.
b Low, high or unclear risk of bias.

Summary assessment of the risk of bias for an outcome within a study across domains

Outcome measure	Summary risk of bias across all domains within a study
Subjective (list of outcomes): WOMAC, UCLA activity and Merle d'Aubigné scores	Unclear risk of bias
Objective (list of outcomes): revision, infection, implant dislocation, aseptic loosening and fracture rates	Low risk of bias

Quality assessment of included systematic reviews: studies comparing total hip replacement with resurfacing arthroplasty ($n = 3$)

Jiang et al.[142]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☐ No ☒ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model

should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☐ **High quality**

☒ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Name of first reviewer: Alexander Tsertsvadze
Name of second reviewer: Paul Sutcliffe
First author surname year of publication: Smith 2010[140]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☐ No ☒ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model

should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☒ **High quality**

☐ **Medium quality**

☐ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Name of first reviewer: Alexander Tsertsvadze
Name of second reviewer: Paul Sutcliffe
First author surname year of publication: Springer 2009[141]

1. Was an 'a priori' design provided?
The research question and inclusion criteria should be established before the conduct of the review.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

2. Was there duplicate study selection and data extraction?
There should be at least two independent data extractors and a consensus procedure for disagreements should be in place.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

3. Was a comprehensive literature search performed?
At least two electronic sources should be searched. The report must include years and databases used (e.g. Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews, textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

4. Was the status of publication (i.e. grey literature) used as an inclusion criterion?
The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

5. Was a list of studies (included and excluded) provided?
A list of included and excluded studies should be provided.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

6. Were the characteristics of the included studies provided?
In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g. age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported.

☒ Yes ☐ No ☐ Can't answer ☐ Not applicable

7. Was the scientific quality of the included studies assessed and documented?
'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.

☐ Yes ☐ No ☒ Can't answer ☐ Not applicable

8. Was the scientific quality of the included studies used appropriately in formulating conclusions?
The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

9. Were the methods used to combine the findings of studies appropriate?
For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e. Chi-squared test for homogeneity, I^2). If heterogeneity exists a random effects model

should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e. is it sensible to combine?).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

10. Was the likelihood of publication bias assessed?
An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test).

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

11. Was the conflict of interest stated?
Potential sources of support should be clearly acknowledged in both the systematic review and the included studies.

☐ Yes ☒ No ☐ Can't answer ☐ Not applicable

OVERALL RATING (tick one box):

☐ **High quality**

☐ **Medium quality**

☒ **Low quality**

Rating (by criteria fulfilled, i.e. 'yes' response): 9 to 11 high quality; 5 to 8 medium quality; 0 to 4 low quality

Appendix 3 Study details and patient characteristics of included randomised controlled trials and systematic reviews

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Angadi 2012,[112] UK	*Outcome category*: cup fixation *Aim*: to compare the clinical and radiological results of a cemented all-PE Ultima acetabular component with those of a cementless PFC following THR *Length of follow-up*: up to 14 years [mean follow-up 7.52 (range 0.4–15.0) years for patients in the Ultima group and 7.87 (0.5–14.0) years for those in the PFC group]	*Inclusion criteria*: candidates for primary THR with OA or RA who were independently mobile without cognitive impairment *Exclusion criteria*: patients aged < 55 years and those judged to be unsuitable for cementless fixation at surgery at the discretion of the senior operating surgeon were excluded. Patients with cognitive impairment were also excluded	*THR 1*: cemented all-PE Ultima acetabular component. The Ultima acetabular component is an UHMWPE implant with a minimum thickness of 5 mm. It is hemispherical with a circumferential flange *THR 2*: cementless porous-coated acetabular component with a PE liner. The PFC acetabular component is a cobalt–chromium alloy hemispherical shell without holes, which is a porous-coated surface of cobalt–chromium–molybdenum alloy beads with a mean pore size of 290 µm and an UHMWPE liner	Patients randomised, *n*	183	104
				Age (years), mean (range)	71.3 (55–89)	69.8 (56–89)
				Sex, female, *n/N* (%)	110/183 (60)	56/104 (54)
				Weight (kg), mean (range)	73.9 (43–128)	76.1 (49–124)
				BMI (kg/m^2), mean (range)	26.7 (13.3–41.4)	27.4 (18.8–44.1)
				Primary OA, *n/N* (%)	172/183 (94)	99/104 (95)
				Bilateral OA, *n/N* (%)	55/183 (30)	37/104 (36)
				HHS, mean (range)	35.2 (9–76)	35.7 (10–70)
				OHS, mean (range)	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Bjørgul 2010,[110] Bjørgul 2010,[111] Norway	*Outcome category*: cup fixation *Aim*: to compare the effects of cemented (Charnley) and uncemented (Duraloc) cups on long-term follow-up for radiographic and clinical outcomes *Length of follow-up*: 10–14 years	*Inclusion criteria*: patients aged ≤ 75 years with OA, post-traumatic arthritis, psoriatic arthritis, gout, RA, juvenile RA and systemic lupus erythematosus *Exclusion criteria*: previous prosthetic replacement was a contraindication to participation, but not osteotomies and internal fixations	*THR 1*: THR using cemented (Charnley) cup. Cement containing gentamycin and a Charnley stem (DePuy) with a 22.225-mm head diameter *THR 2*: THR using uncemented (Duraloc 1200, DePuy) cup. Hemispherical modular cup consisting of a titanium shell with a porous-coated surface	Patients randomised, *n*	107	108
				Age (years), mean (95% CI)	65 (64 to 66)	66 (65 to 67)
				Sex, female, *n/N* (%)	81/107 (76.0)	76/108 (71.0)
				Weight (kg), mean (95% CI)	NR	NR
				BMI (kg/m^2), mean (95% CI)	27 (27 to 28)	27 (26 to 27)
				Primary OA, *n/N* (%)	93/107 (87.0)	94/108 (87.0)
				Bilateral OA, *n/N* (%)	13/107 (12.1)	12/108 (11.1)
				HHS, mean (95% CI)	47 (45 to 50)	49 (47 to 52)
				OHS, mean (95% CI)	NR	NR
McCalden 2009,[145] Canada	*Outcome category*: cup liner bearing surface *Aim*: to report the clinical and radiographic results, after a minimum of 5 years' follow-up, of a randomised, blinded, controlled trial comparing conventional PE liners with first-generation HXLPE liners *Length of follow-up*: Mean 6.8 years	*Inclusion criteria*: a patient had to have degenerative arthritis of one hip requiring total hip arthroplasty, a designation of A or B according to the Charnley hip classification and an age between 40 and 79 years *Exclusion criteria*: pre-existing bone disease (such as severe osteoporosis or osteomalacia), systemic conditions affecting bone density (such as inflammatory arthritis or renal disease) and a contralateral revision or poorly functioning THR	*THR 1*: HXLPE acetabular cup liners. The HXLPE liners, calcium stearate-free GUR® 1050 resin, was also utilised to create compression-moulded sheets, which were then machined into the final implant geometry *THR 2*: conventional PE acetabular cup liners made of calcium stearate-free GUR 1050 resin machined from compression-moulded sheet PE. The final implant was then sterilised with gamma radiation (25 kGy) in an inert nitrogen environment	Patients randomised, *n*	50	50
				Age (years), mean (range)	72.31 (56–79)	72.58 (56–79)
				Sex, female, *n/N* (%)	33/50 (66)	36/50 (72)
				Weight (kg), mean (range)	NR	NR
				BMI (kg/m^2), mean (range)	29.7 (22–39)	29.71 (18–48)
				Primary OA, *n/N* (%)	NR	NR
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	38.96 (11.35)	35.64 (12.97)
				OHS, mean (SD)	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Engh 2012,[113] Engh 2006,[114] USA	*Outcome category*: cup liner bearing surface *Aim*: to compare the clinical outcome of THR patients randomised to either cross-linked or conventional non-cross-linked PE cup liners *Length of follow-up*: 10 years	*Inclusion criteria*: NR *Exclusion criteria*: NR	*THR 1*: cross-linked PE cup liners (Marathon). Patients were implanted with a Duraloc 100 (DePuy) cup incorporating a 4-mm lateralised liner. The PE liner was secured in the Duraloc shell by a peripheral locking ring that engaged a groove machined into the liner and shell *THR 2*: non-cross-linked (conventional) PE cup liners (Enduron)	Patients randomised, *n*	111	109
				Age (years), mean (SD)	62.5 (10.6)	62.0 (11.1)
				Sex, female, *n/N* (%)	65/111 (58.0)	57/109 (52.3)
				Weight (kg), mean (SD)	84.3 (21.3)	81.6 (18.1)
				BMI (kg/m^2), mean (SD)	28.6 (5.5)	27.9 (5.1)
				Primary OA, *n/N* (%)	99/111 (89.1)	90/109 (82.5)
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	88.0 (14.0)	86.0 (15.0)
				OHS, mean (SD)	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics			
					THR 1	THR 2	THR 3
Capello 2008,[115] D'Antonio 2005,[116] D'Antonio 2003,[117] Mesko 2011,[118] USA	*Outcome category*: cup shell design *Aim*: to compare clinical and radiography outcomes between patients receiving THR with ceramic-on-ceramic bearings and patients receiving THR with metal-on-PE bearings; to compare the results between two groups of patients receiving THR with ceramic-on-ceramic bearings but with different cup designs (porous-coated shell vs. arc-deposited hydroxyapatite-coated shell) *Length of follow-up*: 10 years	*Inclusion criteria*: patients aged 21–75 years; not morbidly obese; clinically qualified for THR; diagnosis of OA, traumatic arthritis, avascular necrosis, slipped capital epiphysis, pelvic fracture, femoral fracture, failed fracture, fixation or diastrophic variant; absence of active infection in the affected hip or no previous THR, no psychiatric disorder, senile dementia, Alzheimer's disease, presence of alcohol or substance abuse, no neuromuscular or neurosensory deficiency, no systemic disorder, not immunologically suppressed nor receiving steroids in excess of physiological dose requirements; skeletally mature; not pregnant; and no plans to relocate to another geographic area before completion of the study *Exclusion criteria*: see above	*THR 1*: THR with ceramic-on-ceramic bearings (titanium porous-coated shell) – system I *THR 2*: THR with ceramic-on-ceramic bearings (titanium arc-deposited hydroxyapatite-coated shell) – system II *THR 3*: THR with metal-on-PE bearings (titanium porous-coated shell) – system III *THR 4*: this arm was added later and was not a randomised arm; therefore not extracted – system IV	Patients randomised, *n*	113	109	106
				Age (years), mean (SD)	53 (11.4)	54 (10.7)	55 (10.7)
				Sex, female, *n/N* (%)	41/113 (35.0)	41/109 (37.0)	42/106 (39.0)
				Weight (kg), mean (SD)	85.0 (17.9)	87.8 (17.3)	85.9 (17.7)
				BMI (kg/m^2), mean (SD)	NR	NR	NR
				Primary OA, *n/N* (%)	95/113 (84.0)	86/109 (79.0)	81/106 (77.0)
				Bilateral OA, *n/N* (%)	NR	NR	NR
				HHS, mean (SD)	NR	NR	NR
				OHS, mean (SD)	NR	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Corten 2011,[119] Laupacis 2002,[120] Bourne 2010,[121] Corten 2011,[122] Canada	*Outcome category*: cup/stem fixation and femoral head bearing-on-cup liner bearing *Aim*: to compare the effects of cemented cup/stem [Mallory Head (Biomet)] and uncemented cup/stem (Mallory Head) prostheses on long-term follow-up for mortality, revision, time to revision, health-related quality of life and radiography signs *Length of follow-up*: 20 years	*Inclusion criteria*: OA of the hip and undergoing a unilateral primary arthroplasty *Exclusion criteria*: age > 75 years, severe symptomatic OA of either knee or the contralateral hip, a previous arthroplasty of the ipsilateral hip, arthroplasty on the contralateral side > 5 years before the most recent arthroplasty or had had infectious arthritis	*THR 1*: THR using cemented femoral and cemented acetabular components. All patients were operated on by either surgeon with use of an identical direct lateral approach and within a vertical laminar airflow enclosure in which the surgical team wore body-exhaust suits *THR 2*: THR using uncemented femoral and uncemented acetabular components. The prosthesis (Mallory Head) was made from a titanium alloy. The implant was a tapered design with a 3° taper and the proximal one-third was plasma sprayed	Patients randomised	124	126
				Age (years), mean (SD)	64 (8.0)	64 (7.0)
				Sex, female, *n/N* (%)	60/124 (48)	60/126 (46)
				Weight (kg), mean (SD)	NR	NR
				BMI (kg/m^2), mean (SD)	NR	NR
				Primary OA, *n/N* (%)	NR	NR
				Bilateral OA, *n/N* (%)	0/124 (0.0)	0/126 (0.0)
				HHS, mean (SD)	44 (11.0)	43 (10.0)
				OHS, mean (SD)	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Howie 2012,[123] Australia	*Outcome category*: femoral head size *Aim*: to compare the incidence of dislocation at 1 year after THR between two groups of patients who had received 36-mm and 28-mm femoral head articulations *Length of follow-up*: 1 year	*Inclusion criteria*: patients aged ≥ 60 years with OA and RA referred for THR *Exclusion criteria*: patients aged < 60 years with diagnoses other than OA, RA, inflammatory arthritis or previous fracture/dislocation/surgery involving the hip, abnormal acetabulum, neuromuscular disorder affecting the hip, tumour of the hip, unable to provide consent, unable to complete follow-up	*THR 1*: 36-mm femoral head. All arthroplasties were performed with use of uncemented acetabular components, which comprised a cluster three-holed acetabular shell (Trilogy) fixed with one or two screws and a 10° elevated 36- or 28-mm inner diameter HXLPE liner (Longevity). A cemented femoral stem was used for all arthroplasties (CPT®). During the trial, the taper of the CPT femoral stem was changed from a 6° taper to a 12/14 taper by the manufacturer *THR 2*: 28-mm femoral head. See above	Patients randomised, *n*	273	284
				Age (years), mean (95% CI)	72.3 (71.5 to 73.0)	72.3 (71.6 to 73.1)
				Sex, female, *n/N* (%)	152/273 (56.0)	175/284 (61.3)
				Weight (kg), mean (95% CI)	NR	NR
				BMI (kg/m^2), mean (95% CI)	28.0 (27.4 to 28.7)	28.4 (27.8 to 29.0)
				Primary OA, *n/N* (%) (95% CI)	96.3 (94.1 to 98.6)	95.4 (93.0 to 97.9)
				Bilateral OA, *n/N* (%)	0	0
				HHS, mean (SD)	NR	NR
				OHS, mean (SD)	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Lewis 2008,[124] Canada	*Outcome category*: femoral head bearing *Aim*: to compare clinical outcomes in patients who received THR with oxinium vs. cobalt–chromium femoral heads *Length of follow-up*: 2 years	*Inclusion criteria*: NR *Exclusion criteria*: NR	*THR 1*: THR with oxinium femoral heads. In total, 46 patients received an Echelon stem; the remaining four received a Synergy stem. The acetabular components were press-fit, uncemented Reflection cups paired with either standard PE (22 cases) or HXLPE (28 cases) *THR 2*: THR with cobalt–chromium femoral heads. In total, 30 patients received an Echelon stem and the remaining 20 patients received a Synergy stem. The acetabular components were press-fit, uncemented Reflection cups paired with either standard PE (31 eases) or HXLPE (19 cases)	Patients randomised, *n*	50	50
				Age (years), mean (SD)	51 (10.8)	51 (11.0)
				Sex, female, *n/N* (%)	24/50 (48.0)	24/50 (48.0)
				Weight (kg), mean (SD)	NR	NR
				BMI (kg/m^2), mean (SD)	NR	NR
				Primary OA, *n/N* (%)	NR	NR
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	NR	NR
				OHS, mean (SD)	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Amanatullah 2011,[125] USA	*Outcome category*: femoral head bearing-on-cup liner bearing *Aim*: to compare the clinical performance and evaluate the wear rate of ceramic-on-ceramic vs. ceramic-on-PE bearing surfaces *Length of follow-up*: > 5 years	*Inclusion criteria*: patients were included if clinically indicated for a THR as a result of OA or RA and were aged 21–80 years with a HHS ≤ 60, available for ≥ 2 years of clinical follow-up, able to meet acceptable preoperative medical clearance and without the presence or history of treatment for cardiac, pulmonary, haematological or any other medical condition that would pose an excessive operative risk *Exclusion criteria*: NR	*THR 1*: Ceramic-on-ceramic. Each ceramic–ceramic articulation was implanted with a 28- or 32-mm alumina ceramic femoral head and an alumina ceramic acetabular cup liner *THR 2*: ceramic-on-PE. Alumina ceramic components were also sterilised with ethylene oxide gas. Metal components were sterilised with a minimum of 25 kGy of gamma irradiation	Patients randomised, *n*	166 patients (196 hips)	146 patients (161 hips)
				Age (years), mean (SD)	50.4 (12.8)	54.7 (12.9)
				Sex, female, *n/N* (%)	60/166 (36.1)	62/146 (42.5)
				Weight (kg), mean (SD)	86.9 (20.0)	83.7 (18.5)
				BMI (kg/m^2), mean (SD)	29.6 (12.4)	28.0 (5.1)
				Primary OA, *n/N* (%)	NR	NR
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	NR	NR
				OHS, mean (SD)	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics					
					THR 1	THR 2	THR 3	THR 4	THR 5
Kadar 2011,[126] Norway	*Outcome category*: femoral head bearing-on-cup liner bearing *Aim*: to evaluate wear and migration patterns between cemented highly cross-linked Reflection All-Poly (HXLPE) cup and All-Poly cup articulated with either oxinium or cobalt–chromium femoral heads compared with the Charnley Ogee prostheses	*Inclusion criteria*: primary or secondary OA of the hip *Exclusion criteria*: BMI > 35 kg/m^2, uncompensated cardiopulmonary disease, malignant disease, dementia, RA or other serious systemic diseases	*THR 1*: Charnley Ogee prosthesis. Charnley monoblock stainless steel femoral stem with a 22.2-mm head articulated with a cemented Charnley Ogee UHMWPE (GUR 1050) acetabular cup that was gamma sterilised with 2.5 Mrad in nitrogen *THR 2*: cobalt–chromium-on-PE articulation. Spectron EF femoral stem with a 28-mm cobalt–chromium femoral head and a All-Poly UHMWPE (GUR 1050) cup that was sterilised by ethylene oxide	Patients randomised, *n*	30	30	30	30	30
				Age (years), mean (SD)	70 (6.1)	69 (5.9)	69 (6.7)	70 (5.3)	70 (5.4)
				Sex, female, *n/N* (%)	20/30 (66.6)	20/30 (66.6)	23/30 (76.6)	20/30 (66.6)	22/30 (73.3)
				Weight (kg), mean (SD)	76 (14.9)	76 (11.1)	72 (13.9)	80 (14.8)	76 (14.6)
				BMI (kg/m^2), mean (SD)	NR	NR	NR	NR	NR
				Primary OA, *n/N* (%)	28/30 (93.3)	26/30 (86.6)	26/30 (86.6)	22/30 (73.3)	27/30 (90.0)
				Bilateral OA, *n/N* (%)	NR	NR	NR	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics					
					THR 1	THR 2	THR 3	THR 4	THR 5
	Length of follow-up: 2 years		*THR 3*: Oxinium-on-PE articulation. Spectron EF femoral stem with a 28-mm oxinium femoral head and a All-Poly UHMWPE (GUR 1050) cup that was sterilised by ethylene oxide *THR 4*: Cobalt–chromium-on-HXLPE articulation. Spectron EF femoral stem with a 28-mm cobalt–chromium femoral head and a Reflection All-Poly HXLPE (GUR 1050) cup irradiated with 10 Mrad, melted at 135°C, and ethylene oxide sterilised *THR 5*: Oxinium-on-HXLPE articulation. Spectron EF femoral stem with a 28-mm oxinium femoral head and a Reflection All-Poly HXLPE (GUR 1050) cup irradiated with 10 Mrad, melted at 135°C, and ethylene oxide sterilised	HHS, mean (SD)	45 (NR)	41 (NR)	47 (NR)	47 (NR)	40 (NR)
				OHS, mean (SD)	NR	NR	NR	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Healy 2009,[127] USA	*Outcome category*: stem composition *Aim*: to compare the effects of cobalt–chromium vs. titanium femoral stems in terms of post-THR clinical and radiographic measures *Length of follow-up*: 4.7 (range 2.0–8.9) years	*Inclusion criteria*: NR *Exclusion criteria*: NR	*THR 1*: THR with cobalt–chromium femoral stem. Trilock femoral stem made of cobalt–chromium or titanium in 11 sizes is a straight, collarless, modular, cementless, porous-coated femoral implant with a flat tapered wedge design of the intraosseous body of the stem *THR 2*: THR with titanium femoral stem. See full data extraction sheet for more details (see *Appendix 4*)	Patients randomised, *n*	199	191
				Age (years), mean (range)	66 (25–100)	64 (35–102)
				Sex, female, *n/N* (%)	96/199 (48.2)	92/191 (48.1)
				Weight (kg), mean (range)	81.7 (44.9–136.0)	81.7 (36.3–149.6)
				BMI (kg/m^2), mean (SD)	NR	NR
				Primary OA, *n/N* (%)	182/199 (91.4)	168/191 (88.0)
				Bilateral OA, *n/N* (%)	33/199 (16.6)	24/191 (12.6)
				HHS, mean (range)	50.3 (16.2–70.7)	50.8 (23.4–75.4)
				OHS, mean (range)	NR	NR

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Kim 2011,[128] the Republic of Korea	*Outcome category*: stem design *Aim*: to compare a short metaphyseal-fitting femoral stem with a conventional metaphyseal- and diaphyseal-fitting stem with respect to postoperative clinical and radiographic parameters *Length of follow up*: 3.35 years	*Inclusion criteria*: NR *Exclusion criteria*: Severe osteoporosis of the proximal femur	*THR 1*: short anatomical metaphyseal-fitting cementless stem (Proxima). Cementless Pinnacle acetabular component made of titanium alloy, Proxima stem and 28-mm-internal-diameter Biolox forte ceramic liner were used in all hips *THR 2*: Conventional anatomical metaphyseal- and diaphyseal-fitting cementless stem (Profile). The cementless Profile femoral component is made of titanium alloy	Patients randomised, *n*	50	50
				Age (years), mean	54.3 (12.97)	51.8 (12.3)
				Sex, female, *n/N* (%)	28/50 (56)	26/50 (52)
				Weight (kg), mean (SD)	66.5 (9.51)	64.8 (10.6)
				BMI (kg/m^2), mean (SD)	25.6 (2.82)	24.7 (3.6)
				Primary OA, *n/N* (%)	24/50 (48)	24/50 (48)
				Bilateral OA, *n/N* (%)	5/50 (10)	5/50 (10)
				HHS, mean (SD)	44.6 (NR)	48.4 (NR)
				OHS, mean (SD)	NR	NR

continued

TABLE 130 Study details and patient characteristics of included RCTs: studies comparing different types of THR (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR 1	THR 2
Kim 2011,[129] the Republic of Korea	*Outcome category*: stem fixation *Aim*: to compare the clinical and radiological results, rates of revision and survival of implants after THR with cemented (hybrid) vs. cementless femoral components performed in patients < 50 years of age at a minimum 16 years' follow-up *Length of follow-up*: 20 years	*Inclusion criteria*: NR *Exclusion criteria*: NR	*THR 1*: THR with cemented femoral stem. The Charnley Elite or Elite-plus stem (Ortron 90™) (DePuy) was used in the cemented (hybrid) group and the Profile stem (DePuy) was used in the cementless group *THR 2*: THR with cementless femoral stem A cementless Duraloc 100 or 1200 series acetabular component (DePuy) was used in all hips in both groups. Of the 62 Duraloc 1200 acetabular components used in both groups, 28 were fixed with one or two screws and the remaining 34 were press-fitted without using an additional screw	Patients randomised, *n*	83	83
				Age (years), mean (range)	43.4 (21–50)	46.8 (21–49)
				Sex, female, *n/N* (%)	16/78 (21)	21/79 (27)
				Weight (kg), mean (range)	59 (45–82)	60.5 (48–87)
				BMI (kg/m^2), mean (range)	22.2 (22.1–24.8)	22.2 (21.9–24.4)
				Primary OA, *n/N* (%)	10/78 (12.8)	12/79 (15.2)
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (range)	44 (5–66)	48.8 (6–55)
				OHS, mean (range)	NR	NR

HXLPE, highly cross-linked PE; PE, polyethylene; PFC, porous-coated acetabular component; UHMWPE, ultra-high molecular weight polyethylene.

TABLE 131 Study details of included systematic reviews: studies comparing different types of THR

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Voigt 2012,[137] USA	*Outcome category*: cup fixation *Aim*: to compare uncemented metal-backed acetabular components with PE inserts and cemented all-PE acetabular components (using the same type of femoral component and method of femoral fixation in both arms of the trial) in terms of revision rates, function, complications and costs in patients with OA	*Databases searched*: PubMed-MEDLINE, The Cochrane Library *Last date of search*: 13 June 2011	*Participants*: patients with OA or RA *Interventions*: primary total hip implant with uncemented metal-backed acetabular components with PE inserts *Comparators*: primary total hip implant with cemented all-PE acetabular components *Outcome measures*: revision rate, function (HHS, OHS), complications (infection or wound, deep-vein thrombosis, pulmonary embolism, dislocations, over-reaming, fractures and costs of treatment) *Types of studies*: RCTs	*Quality assessment tool used*: Cochrane risk of bias tool *Risk of bias assessment criteria*: sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other issues *Summary of risk of bias of included studies*: random sequence generation low (4/6 trials), allocation concealment high (3/6 trials) or low (2/6 trials), caregiver blinding high (6/6 trials), patient blinding unknown (4/6 trials) or high (1/6 trials), assessor blinding high (5/6 trials), incomplete outcome data low (5/6 trials), selective reporting low (6/6 trials), other bias low (3/6 trials) or high (2/6 trials)	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: yes *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: yes, (b) publication bias: yes, (c) overall quality/strength of evidence (GRADE): no

continued

TABLE 131 Study details of included systematic reviews: studies comparing different types of THR (*continued*)

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Pakvis 2011,[138] the Netherlands	*Outcome category*: cup fixation *Aim*: to identify all relevant RCTs and comparative cohort studies in which cemented and cementless sockets were compared	*Databases searched*: MEDLINE and EMBASE (1980–December 2009) *Last date of search*: December 2009	*Participants*: indication for performing THR had to be primary or secondary OA *Interventions*: cemented acetabular components *Comparators*: cementless acetabular components *Outcome measures*: minimal follow-up had to be 12 months; data presented had to be clinical (complications, HHS and survival) and radiological (wear, migration and osteolysis) outcome measurements *Types of studies*: non-randomised studies, RCTs	*Quality assessment tool used*: van Tulder checklist (for RCTs); Newcastle–Ottawa quality assessment scale (for non-RCTs) *Risk of bias assessment criteria*: van Tulder checklist: randomisation, allocation concealment, prognostic factors, patient blinding, surgeon blinding, outcome assessor blinding, co-interventions, compliance, drop-out, timing of the outcome assessments, intention to treat and homogeneity; Newcastle–Ottawa quality assessment scale: representativeness of the exposed cohort, selection of the non-exposed cohort, ascertainment of exposure, demonstration that outcome of interest was not present at start of study, comparability of cohorts on the basis of the design or analysis, assessment of outcome and adequacy of follow-up of cohorts) *Summary of risk of bias of included studies*: three RCTs scored 'yes' on > 50% of the van Tulder criteria. In orthopaedic surgery, surgeon blinding is not feasible. Therefore, when re-evaluating the results of the van Tulder questionnaire we could select seven articles that scored 'yes' on > 50% of the van Tulder items	*Direct comparison*: (a) non-quantitative: yes, (b) quantitative: no *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: no, (b) publication bias: no, (c) overall quality/strength of evidence (GRADE): no

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Clement 2012,[139] France	*Outcome category*: cup fixation *Aim*: to perform a critical analysis of the current evidence from a systemic literature review of comparative studies, long-term case series, previous literature reviews, meta-analyses and national arthroplasty registry data on cemented and uncemented acetabular components	*Database searched*: MEDLINE *Last date of search*: 2011	*Participants*: young patients and patients with dysplastic hip disease *Intervention*: cemented THR *Comparator*: uncemented THR *Outcome measures*: aseptic loosening, radiographic loosening, overall survival, wear rates per year, dislocation, osteolysis, acetabular revision, liner exchange, quality of life *Types of studies*: (1) all published review articles and meta-analyses; (2) all studies comparing cemented with uncemented acetabular components with a minimal follow-up of 5 years, quoting survival and wear or complications; (3) all arthroplasty registers reporting a comparison between cemented and uncemented acetabular components; (4) single-centre outcome studies with > 13 years mean follow-up; (5) single-centre series reporting survivorship of cemented or uncemented cups in young patients and patients with dysplastic hip disease	*Quality assessment tool used*: none *Risk of bias assessment criteria*: none *Summary of risk of bias of included studies*: NA	*Direct comparison*: (a) non-quantitative: yes, (b) quantitative: no *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: no, (b) publication bias: no, (c) overall quality/strength of evidence (GRADE): no

continued

TABLE 131 Study details of included systematic reviews: studies comparing different types of THR (*continued*)

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Sedrakyan 2011,[140] USA	*Outcome category*: femoral head bearing-on-cup liner bearing *Aim*: to determine comparative safety and effectiveness of combinations of bearing surfaces of hip implants	*Databases searched*: MEDLINE, EMBASE and The Cochrane Central Register of Controlled Trials from January 1995 *Last date of search*: June 2011	*Participants*: adults, reporting any one of the clinical outcomes of interest (any functional outcomes or revisions or both) *Interventions*: conventional hip replacement *Comparators*: conventional hip replacement *Outcome measures*: any functional outcome (HHS and general quality of life measures such as SF-12) and occurrence of revision *Types of studies*: RCTs, controlled clinical trials, observational comparative controlled studies	*Quality assessment tool used*: selected validity items (RCTs) and STROBE (observational studies) *Risk of bias assessment criteria*: RCTs: methods of random allocation generation, allocation concealment, masking of patients and outcome assessors, and intention-to-treat analysis *Summary of risk of bias of included studies*: four studies were classified as moderate to high quality, five studies as moderate quality and six studies as low quality	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: yes *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: yes, (b) publication bias: yes, (c) overall quality/strength of evidence (GRADE): no
Yoshitomi 2009,[141] Japan	*Outcome category*: femoral head bearing-on-cup liner bearing *Aim*: to compare the survivorship/revision and annual PE wear rates between THRs with zirconia-on-PE and THRs with non-zirconia-on-PE and to explore if manufacturers or fixation method influenced survivorship	*Databases searched*: PubMed (1966–July 2007), EMBASE (1974–July 2007) and The Cochrane Central Register of Controlled Trials (Issue 4, July 2007) *Last date of search*: July 2007	*Participants*: NR *Interventions*: THR using zirconia heads with PE cup liners (regardless of femoral head size, method of fixation) *Comparators*: THR using non-zirconia heads with PE cup liners (regardless of femoral head size, method of fixation) *Outcome measures*: survivorship/revision, PE wear rates *Types of studies*: RCTs, non-RCTs and cohort studies with follow-up of > 5 years	*Quality assessment tool used*: The Cochrane Back Review Group 11-item criteria *Risk of bias assessment criteria*: Generation of random allocation, allocation concealment, blinding, co-interventions, compliance, sample attrition, outcome assessment timing and type of analysis *Summary of risk of bias of included studies*: mean (range) score: cohort studies 4.5 (4–5), RCTs 6.3 (6–7)	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: yes *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: yes, (b) publication bias: yes, (c) overall quality/strength of evidence (GRADE): no

NA, not applicable; NR, not reported; PE, polyethylene; STROBE, STrengthening the Reporting of OBservational studies in Epidemiology.

TABLE 132 Study details and patient characteristics of included RCTs: THR vs. RS

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics	THR	RS
Costa 2012,[130] Achten 2010,[107] UK	*Outcome category*: THR vs. RS *Aim*: to compare the clinical effectiveness of THR with that of RS in patients with severe arthritis of the hip with regard to hip function, quality of life, physical activity and harms *Length of follow-up*: 12 months	*Inclusion criteria*: age > 18 years, medically fit for an operation and suitable for RS *Exclusion criteria*: evidence indicating that patient would be unable to adhere to trial procedures or complete questionnaires; if a recruited patient needed a contralateral hip replacement during the trial period, the second hip was not included in the study	*THR*: NR *RS*: NR	Patients randomised, *n*	60	66
				Age (years), mean (SD)	56.3 (7.3)	56.6 (6.6)
				Sex, female, *n/N* (%)	22/60 (37.0)	30/66 (45.0)
				Weight (kg), mean (SD)	NR	NR
				BMI (kg/m^2), mean (SD)	28.6 (6.3)	28.7 (4.6)
				Primary OA, *n/N* (%)	59/60 (98)	61/66 (93)
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	48.6 (14.2)	50.1 (13.5)
				OHS, mean (SD)	19.1 (8.0)	19.6 (7.8)
Garbuz 2010,[131] Canada	*Outcome category*: THR vs. RS *Aim*: to compare THR (metal-on-metal with large diameter head) with RS (metal-on-metal) in terms of quality-of-life measures and metal ion levels in serum *Length of follow-up*: 2 years	*Inclusion criteria*: patients aged between 19 and 70 years deemed suitable for hip RS as judged by the treating surgeon *Exclusion criteria*: previous fracture of the hip requiring internal fixation, previous femoral or pelvic osteotomy, dysplasia requiring structural graft, presence of osteopenia or osteoporosis, and hepatic or renal insufficiency	*THR*: large-head metal-on-metal Durom acetabular cup, M/L Taper titanium femoral stem, large Metasul head attached via a cobalt–chromium alloy metal sleeve adapter and Morse taper to match 12/14 stem taper *RS*: Durom acetabular cup, Durom femoral RS component, large Metasul head attached via a cobalt–chromium alloy metal sleeve adapter and Morse taper to match 12/14 stem taper Bearing surface in each arm was identical (all Zimmer, Inc.)	Patients randomised, *n*	56	48
				Age (years), mean (SD)	52.0 (NR)	51.5 (NR)
				Sex, female, *n/N* (%)	6/56 (10.7)	5/48 (10.4)
				Weight (kg), mean (SD)	NR	NR
				BMI (kg/m^2), mean (SD)	28.2 (NR)	28.3 (NR)
				Primary OA, *n/N* (%)	NR	NR
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	NR	NR
				OHS, mean (SD)	NR	NR

continued

TABLE 132 Study details and patient characteristics of included RCTs: THR vs. RS (*continued*)

Study and country	Study details	Inclusion/exclusion criteria	Intervention and prostheses used	Patient characteristics		
					THR	RS
Vendittoli 2010,[132] Vendittoli 2006,[133] Girard 2006,[134] Rama 2009,[135] Vendittoli 2006,[136] Canada	*Outcome category*: THR vs. RS *Aim*: to compare THR (metal-on-metal) with RS (metal-on-metal) with respect to postoperative clinical and radiographic outcomes and complications *Length of follow-up*: 6 years	*Inclusion criteria*: patients aged 18–65 years with degenerative hip joint disease who were candidates for both metal–metal THR and metal–metal RS *Exclusion criteria*: implantation, hip arthrodesis, renal insufficiency, known or suspected metal allergy, osteopenia or osteoporosis of the hip	*THR*: titanium, cementless CLS® Spotorno® Hip Stem (Zimmer, Inc.) femoral stem, Allofit™ acetabular cup, 28-mm Metasul cobalt–chromium insert and femoral head (Zimmer, Inc.). Implant options included 135° or 145° neck-shaft angle and neck lengths from −4 mm to +8 mm *RS*: hybrid Durom™ (Zimmer, Inc.), cemented femoral component, cementless acetabular component. Femoral head preparation included drilling of any sclerotic area and routine pulse lavage. Cementing with low-viscosity cement with tobramycin (Simplex, Stryker) at approximately 4 minutes	Patients randomised, *n*	100	109
				Age (years), mean (SD)	51.0 (8.6)	49.2 (9.0)
				Sex, female, *n/N* (%)	32/100 (32.0)	40/109 (37.0)
				Weight (kg), mean (SD)	NR	NR
				BMI (kg/m^2), mean (SD)	30.0 (6.8)	27.0 (5.3)
				Primary OA, *n/N* (%)	39/100 (39.0)	34/109 (31.2)
				Bilateral OA, *n/N* (%)	NR	NR
				HHS, mean (SD)	NR	NR
				OHS, mean (SD)	NR	NR

NR, not reported.

TABLE 133 Study details of included systematic reviews: THR vs. RS

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Jiang 2011,[142] China	*Outcome category*: THR vs. RS *Aim*: to compare the clinical results of metal-on-metal RS with those of standard THR for the treatment of hip disease in active young patients	*Databases searched*: Cochrane Bone, Joint and Muscle Trauma Group Specialised Register (June 2009), Cochrane Central Register of Controlled Trials (The Cochrane Library, Issue 2, 2009), PubMed (January 1990–June 2009), Ovid (January 1990–June 2009), Science Direct Online (January 1990–June 2009) *Last date of search*: June 2009	*Participants*: age < 65 years, skeletally mature, end-stage hip disease, follow-up > 12 months *Interventions*: modern metal-on-metal RS *Comparators*: THR *Outcome measures*: rate of revision, mortality, femoral neck fracture, component loosening, dislocation and deep hip joint infection as well as hip function and range of motion *Types of studies*: RCTs and controlled clinical trials	*Quality assessment tool used*: Cochrane risk of bias assessment tool *Risk of bias assessment criteria*: sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other issues *Summary of risk of bias of included studies*: three studies were adequately randomised and the fourth study was randomised only by the patients' date of admission. Two studies included patients who were lost before follow-up could be completed, and the allocation concealment of all four eligible studies was unclear. None of the studies included adequate blinding procedures	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: yes *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: yes, (b) publication bias: no, (c) overall quality/strength of evidence (GRADE): no

continued

TABLE 133 Study details of included systematic reviews: THR vs. RS (*continued*)

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Smith 2010,[143] UK	*Outcome category*: THR vs. RS *Aim*: to compare THR and RS for clinical and radiological outcomes and complication rates	Databases searched: MEDLINE (1950–January 2010), CINAHL (1982 to January 2010), AMED (1985 to January 2010) and EMBASE (1974 to January 2010) *Last date of search*: January 10, 2010	*Participants*: patients with hip pathology *Interventions*: RS *Comparators*: THR *Outcome measures*: incidence of revision, mortality, dislocation, aseptic loosening, avascular necrosis, infection and fracture; incision length, last acetabular reamer size, duration of operation, blood loss and frequency of blood transfusion requirement, length of hospital stay, pain, functional and quality of life outcomes, and hip range of motion; femoral/acetabular offset, incidence of femoral/acetabular radiolucency, leg length, cup height and heterotopic ossification; incidence of complications (venous thromboembolic events, acetabular malposition, trochanteric malunion or non-union, nerve palsy and presence of Trendelenburg sign) *Types of studies*: RCTs and non-RCTs	*Quality assessment tool used*: modified 17-item appraisal tool (CASP) *Risk of bias assessment criteria*: subject identification, randomisation, blinding and dropout rates; presentation of results using descriptive and inferential statistics; and external validity to clinical practice *Summary of risk of bias of included studies*: CASP score (maximum 17): 0–3 (one study), 4–7 (16 studies), 8–11 (23 studies), 12–15 (five studies), 16–17 (no studies); nine RCTs clearly described the method of randomisation; for 25 studies the groups were comparable at baseline; assessor blinding was used in four studies; patients were blinded in only two studies; in 16 studies the results were analysed by intention-to-treat methods; the results were interpreted appropriately in 35 studies	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: yes *Indirect comparison*: (a) unadjusted: no, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: no, (b) publication bias: yes, (c) overall quality/strength of evidence (GRADE): no

Study and country	Study details	Search strategy	Inclusion criteria	Quality assessment	Methods of synthesis
Springer 2009,[144] Canada	*Outcome category*: THR vs. RS *Aim*: to compare the effects of THR and hip RS on failure rates of cementless femoral components in younger patients	*Databases searched*: MEDLINE, PubMed, and CINAHL searched from their inception *Last date of search*: 31 March 2008	*Participants*: young adults (mean age < 55 years) *Interventions*: THR with modern cementless components *Comparators*: RS *Outcome measures*: femoral failure for any reason, femoral failure because of revision and femoral failure because of mechanical reasons *Types of studies*: RCTs, observational studies including single-arm studies (THR or RS only)	*Quality assessment tool used*: methodological quality assessment was performed by a single reviewer by assigning non-randomised studies a Methodological Index for Non-Randomized Studies score *Risk of bias assessment criteria*: NR *Summary of risk of bias of included studies*: NA	*Direct comparison*: (a) non-quantitative: no, (b) quantitative: no *Indirect comparison*: (a) unadjusted: yes, (b) adjusted: no, (c) mixed treatment comparison: no *Specific methods of assessment*: (a) heterogeneity: no, (b) publication bias: no, (c) overall quality/strength of evidence (GRADE): no

AMED, Allied and Complementary Medicine Database; CASP, Critical Appraisal Skills Programme; CINAHL, Cumulative Index to Nursing and Allied Health Literature; NA, not applicable; NR, not reported.

Appendix 4 Full data extraction of included randomised controlled trials and systematic reviews

Included randomised controlled trials: studies comparing different types of total hip replacement (*n* = 13)

Cup fixation

Angadi *et al.*[112]
Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details
Country: UK
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: NR
Funding (government/private/manufacturer/other – specify): NR
Aim of the study
To compare the clinical and radiological results of a cemented all-PE Ultima acetabular component with those of a cementless PFC following THR
Participants
Recruitment dates: July 1995 and July 2001
Total number of patients screened for inclusion eligibility: 300
Total number of patients randomised: 287
Inclusion criteria: candidates for primary THR with OA or RA who were independently mobile without cognitive impairment
Exclusion criteria: patients aged < 55 years of age and those judged to be unsuitable for cementless fixation at surgery at the discretion of the senior operating surgeon were excluded. Patients with cognitive impairment were also excluded
Characteristics of participants (total study sample): mean (range) age (years): 70 (55–89); women, *n/N* (%): 166/287 (58); race/ethnicity, *n/N* (%): NR; diagnosis, *n/N* (%): OA 271 (94.4), other 15 (5.2)
Intervention (keep the same order as in the paper)
Intervention 1 (e.g. THR 1): cemented all-PE Ultima acetabular component
Intervention 2 (e.g. THR 2): cementless porous-coated acetabular component with a PE liner
Bilateral procedure (yes/no/NR): yes (*n* = 92)
Implant manufacturer: DePuy [patients in both treatment groups were to receive a cemented Ultima straightstem femoral component (Johnson & Johnson) produced in titanium alloy with a 28-mm diameter cobalt–chromium head]
Postprocedural rehabilitation (e.g. weight bearing, exercise): immediate full weight bearing with a frame then onto sticks with physiotherapy

Intervention (keep the same order as in the paper)	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	✓
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)

Primary outcomes (*list*): functional outcome (HHS), revision rates

Secondary outcomes (*list*): adverse outcomes (infection, dislocation) and medical complications (deep-vein thrombosis, pulmonary embolism and death); anteroposterior pelvic and cross-table lateral radiographs; osteolysis

Imaging method used (i.e. conventional radiography, radiostereometry, none): radiography

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 6 months and 1, 5 and 10 years

Total length of follow-up: up to 14 years [mean (range) follow-up was 7.52 (0.4–15.0) years for patients in the Ultima group and 7.87 (0.5–14.0) years for those in the PFC group]

Number of patients

	Total	***Cemented***	***Cementless***
Randomised	287	183	104
Analysed (if more than one follow-up point, choose and specify the last one)	271	170	101
Losses to follow-up/dropout/ sample attrition (if more than one follow-up point, choose and specify the last one)	16	13	3

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Cemented	The Ultima acetabular component is an UHMWPE implant with a minimum thickness of 5 mm. It is hemispherical with a circumferential flange. All operations were performed in a laminar air-flow operating theatre. Patients received routine antibiotic prophylaxis (cefuroxime three doses 1.5 g + 0.75 g × 2). Anaesthetic technique was determined at the discretion of the anaesthetist. Surgical approach was either posterior or anterolateral depending on the usual practice of the operating surgeon. The prostheses were implanted following the standard operative procedure as detailed in the appropriate surgical manuals. The cemented acetabular components in this study were inserted without pressurisation	NR
Cementless	The PFC acetabular component is a cobalt–chromium alloy hemispherical shell without holes, which is a porous-coated surface of cobalt–chromium–molybdenum alloy beads with a mean pore size of 290 μm and an UHMWPE liner. The porous-coated cups were undersized by 2 mm and the cemented cups were sized to give a 2-mm cement mantle without pressurisation	NR

Patient baseline characteristics		
Characteristic	***Cemented***	***Cementless***
Age (years), mean (range)	71.3 (55–89)	69.8 (56–89)
Sex, female, *n* (%)	110 (60)	56 (54)
Weight (kg), mean (range)	73.9 (43–128)	76.1 (49–124)
BMI (kg/m^2), mean (range)	26.7 (13.3–41.4)	27.4 (18.8–44.1)
Primary OA, *n*/*N* (%)	172/183 (94)	99/104 (95)
Bilateral OA, *n*/*N* (%)	55/183 (30)	37/104 (36)
HHS, mean (range)	35.2 (9–76)	35.7 (10–70)
OHS, mean (range)	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Postprocedural follow-up assessment timing (specify): 10 years

Outcome	***Cemented***	***Cementless***	***Between-group difference and p-value (or 95% CI)***[a]
Mortality (all-cause), *n/N* (%)	Not clear	Not clear	NR
HHS, mean (range)	74.5 (25–100)	78.0 (37–100)	$p > 0.05$ (NS)
OHS, mean (range)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (range)	NR	NR	NA
HOOS, mean (range)	NR	NR	NA
LISOH, mean (range)	NR	NR	NA
AAOS Hip and Knee Questionnaire, mean (range)	NR	NR	NA
UCLA activity score, mean (range)	NR	NR	NA
WOMAC score, mean (range)	NR	NR	NA
AIMS score, mean (range)	NR	NR	NA
MACTAR score, mean (range)	NR	NR	NA
SF-36 score, mean (range)	NR	NR	NA
SF-12 score, mean (range)	NR	NR	NA
NHP score, mean (range)	NR	NR	NA
EQ-5D score, mean (range)	NR	NR	NA
Pain score (VAS), mean (range)	NR	NR	NA
Pain score (other than VAS; specify), mean (range)	NR	NR	NA
Revision rate, *n/N* (%)	17/183 (9)	11/104 (11)	$p > 0.05$ (NS)
Time to revision (years), mean (range)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	86.8 (78.4 to 92.1); 10-year survivorship (revision for any reason)	89.2 (78.3 to 94.8); 10-year survivorship (revision for any reason)	Log-rank test $p = 0.938$ (NS)
Femoral head penetration (mm/year), mean (range)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n/N* (%) patients with an event, beyond 5 years' follow-up

Complication	*Time of occurrence (peri-/postoperational)*	*Cemented*	*Cementless*	*Between-group difference and* p-*value (or 95% CI)*[a]
Osteolysis (any or total)	Beyond 5 years' follow-up	0/183 (0.0)	1/104 (0.9)	NR
Aseptic loosening (any or total)	Beyond 5 years' follow-up	11/183 (4.9)	2/104 (1.9)	NR
Infection	Beyond 5 years' follow-up	0/183 (0.0)	2/104 (1.9) (loosening with infection)	NR
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	Beyond 5 years' follow-up	1/183 (0.5)	3/104 (2.9)	NR
Other (subluxing)	Beyond 5 years' follow-up	0/183(0.0)	1/104 (0.9)	NR

a RR or risk difference.

Authors' conclusions

Patients with cemented all-PE and cementless porous-coated PE-lined acetabular components have similar long-term clinical outcomes

Reviewers' conclusions

The purpose of the study was to compare the clinical and radiological results of a cemented all-PE Ultima acetabular component with those of a cementless porous-coated acetabular component following THR. This is the largest study that specifically addresses acetabular fixation, comparing cemented and uncemented acetabular components. A total of 287 patients were randomised. No significant difference was found between treatment groups in mean HHS change from baseline at 6 months, 1 year, 5 years or 10 years postoperatively. There was a significant difference in radiological loosening between the cemented and porous-coated cups ($p = 0.001$). The cemented acetabular components in this study were inserted without pressurisation. The significantly higher rate of radiological loosening in this group may be a reflection of this as pressurisation has been shown to be a factor affecting the radiological outcome of cemented acetabular components. Note the larger number of people who had no postoperative follow-up in the Ultima cup group ($n = 13$) than in the PFC cup group ($n = 3$). Difficult to determine the impact of the manufacturer modifying the design or withdrawing the implant based on the findings – this represents the difficulty of carrying out a long-term trial in this area using a new device. The authors also comment on the findings being more reflective of clinical performance across the NHS for similar acetabular components, but as a wide spectrum of operating surgeons was used this could potentially have its own risk of bias

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; PFC, porous-coated acetabular component; UHMWPE, ultra-high molecular weight PE.

Bjørgul *et al.*[110,111]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: Norway

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): specialty clinic

Number of centres: 1

Funding (government/private/manufacturer/other – specify): Odd Fellows, Norway

Aim of the study

To compare the effects of cemented (Charnley) and uncemented (Duraloc) cups on long-term follow-up for radiographic and clinical outcomes

Participants

Recruitment dates: between April 1994 and June 1997

Total number of patients screened for inclusion eligibility: NR

Total number of patients randomised: 215

Inclusion criteria: patients aged ≤ 75 years with OA, post-traumatic arthritis, psoriatic arthritis, gout, RA, juvenile RA and systemic lupus erythematosus

Exclusion criteria: previous prosthetic replacement was a contraindication to participation, but not osteotomy or internal fixation

Characteristics of participants (total study sample): mean age (years): 65; women, *n* (%): 157 (73); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 187 (87.0), congenital hip dysplasia 42 (19.5), post-traumatic arthritis 6 (2.8), RA 4 (1.9), avascular necrosis 1 (0.5)

Intervention (keep the same order as in the paper)

Intervention 1 (e.g. THR 1): THR using cemented (Charnley) cup

Intervention 2 (e.g. THR 2): THR using uncemented (Duraloc) cup

Bilateral procedure (yes/no/NR): Yes

Implant manufacturer: Charnley cup: DePuy; Duraloc 1200 cup: DePuy

Postprocedural rehabilitation (e.g. weight-bearing, exercise): Postoperatively, patients were allowed restricted weight bearing on the day after surgery. All patients were encouraged to use two crutches for at least 6 weeks

If several types of THR are compared, indicate the basis for comparison (tick all that apply):

Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	✓
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)

Primary outcomes (*list*): HHS, revision, implant survival time, radiographic outcomes, complications (e.g. dislocation, cardiovascular, pulmonary embolism)

Secondary outcomes (*list*): see above

Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 6 months and 2, 5, 10 and 14 years

Total length of follow-up: 14 years

Number of patients

	Total	***THR using cemented (Charnley) cup***	***THR using uncemented (Duraloc) cup***
Randomised	215	107	108
Analysed (if more than one follow-up point, choose and specify the last one)	151	71	80
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	31	19	12

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR using cemented (Charnley) cup	Surgery was performed using a direct lateral approach. The Charnley cup was an all-PE cup with a flange. The surgeon cut the flange to fit the rim of the acetabulum, which provided increased pressure to the cement, augmenting cement penetration into the bone of the acetabulum. Surgery was performed under laminar air flow. The femoral component was cemented using third-generation cementing techniques with vacuum mixing, retrograde filling of the canal and pressurisation before insertion of the femoral component. Cement containing gentamycin and a Charnley stem with a 22.22-mm head diameter was used in all cases	Five orthopaedic surgeons performed the surgeries
THR using uncemented (Duraloc) cup	Surgery was performed using a direct lateral approach. The Duraloc 1200 cup is a hemispherical modular cup consisting of a titanium shell with a porous-coated surface. The surface has a mean pore size of 250 µm. The cup had a minimum PE thickness of 6 mm, dome-loading of the PE and an improved locking mechanism designed not to interfere with liner–shell conformity. The shell had a central hole for the insertion device and 12 holes for screw fixation. An UHMWPE liner with a 10° posterior lip was used in all cases. For the femoral component see above	See above

Patient baseline characteristics		
Characteristic	***THR using cemented (Charnley) cup***	***THR using uncemented (Duraloc) cup***
Age (years), mean (95% CI)	65 (64 to 66)	66 (65 to 67)
Sex, female, *n*/*N* (%)	81/107 (76.0)	76/108 (71.0)
Weight (kg), mean (95% CI)	NR	NR
BMI (kg/m^2), mean (95% CI)	27 (27 to 28)	27 (26 to 27)
Primary OA, *n*/*N* (%)	93/107 (87.0)	94/108 (87.0)
Bilateral OA, *n*/N (%)	13/107 (12.1)	12/108 (11.1)
HHS, mean (95% CI)	47 (45 to 50)	49 (47 to 52)
OHS, mean (95% CI)	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Postprocedural follow-up assessment timing (specify): 6 months

Outcome	*THR using cemented (Charnley) cup*	*THR using uncemented (Duraloc) cup*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (95% CI)	90.2 (87.9 to 92.6)	89.1 (86.9 to 91.3)	$p > 0.05$ (NS)
OHS, mean (95% CI)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (95% CI)	NR	NR	NA
HOOS, mean (95% CI)	NR	NR	NA
LISOH, mean (95% CI)	NR	NR	NA
AAOS Hip and Knee Questionnaire, mean (95% CI)	NR	NR	NA
UCLA activity score, mean (95% CI)	NR	NR	NA
WOMAC score, mean (95% CI)	NR	NR	NA
AIMS score, mean (95% CI)	NR	NR	NA
MACTAR, mean (95% CI)	NR	NR	NA
SF-36 score, mean (95% CI)	NR	NR	NA
SF-12 score, mean (95% CI)	NR	NR	NA
NHP score, mean (95% CI)	NR	NR	NA
EQ-5D score, mean (95% CI)	NR	NR	NA
Pain score (VAS), mean (95% CI)	NR	NR	NA
Pain score (other than VAS; specify), mean (95% CI)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (95% CI)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 2 years

Outcome	*THR using cemented (Charnley) cup*	*THR using uncemented (Duraloc) cup*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (95% CI)	92.7 (89.6 to 95.8)	94.0 (92.4 to 95.7)	$p > 0.05$ (NS)
OHS, mean (95% CI)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (95% CI)	NR	NR	NA
HOOS, mean (95% CI)	NR	NR	NA
LISOH, mean (95% CI)	NR	NR	NA
AAOS Hip and Knee Questionnaire, mean (95% CI)	NR	NR	NA
UCLA activity score, mean (95% CI)	NR	NR	NA
WOMAC score, mean (95% CI)	NR	NR	NA
AIMS score, mean (95% CI)	NR	NR	NA
MACTAR, mean (95% CI)	NR	NR	NA
SF-36 score, mean (95% CI)	NR	NR	NA
SF-12 score, mean (95% CI)	NR	NR	NA
NHP score, mean (95% CI)	NR	NR	NA
EQ-5D score, mean (95% CI)	NR	NR	NA
Pain score (VAS), mean (95% CI)	NR	NR	NA
Pain score (other than VAS; specify), mean (95% CI)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (95% CI)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 5 years

Outcome	*THR using cemented (Charnley) cup*	*THR using uncemented (Duraloc) cup*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (95% CI)	93.9 (91.6 to 96.2)	91.4 (89.3 to 93.5)	$p > 0.05$ (NS)
OHS, mean (95% CI)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (95% CI)	NR	NR	NA
HOOS, mean (95% CI)	NR	NR	NA
LISOH, mean (95% CI)	NR	NR	NA
AAOS Hip and Knee Questionnaire, mean (95% CI)	NR	NR	NA
UCLA activity score, mean (95% CI)	NR	NR	NA
WOMAC score, mean (95% CI)	NR	NR	NA
AIMS score, mean (95% CI)	NR	NR	NA
MACTAR, mean (95% CI)	NR	NR	NA
SF-36 score, mean (95% CI)	NR	NR	NA
SF-12 score, mean (95% CI)	NR	NR	NA
NHP score, mean (95% CI)	NR	NR	NA
EQ-5D score, mean (95% CI)	NR	NR	NA
Pain score (VAS), mean (95% CI)	NR	NR	NA
Pain score (other than VAS; specify), mean (95% CI)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (95% CI)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 10 years

Outcome	*THR using cemented (Charnley) cup*	*THR using uncemented (Duraloc) cup*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	12/107 (11.2)	14/108 (12.9)	NR
HHS, mean (95% CI)	89.8 (87.0 to 92.6)	87.3 (84.1 to 90.6)	$p > 0.05$ (NS)
OHS, mean (95% CI)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (95% CI)	NR	NR	NA
HOOS, mean (95% CI)	NR	NR	NA
LISOH, mean (95% CI)	NR	NR	NA
AAOS Hip and Knee Questionnaire, mean (95% CI)	NR	NR	NA
UCLA activity score, mean (95% CI)	NR	NR	NA
WOMAC score, mean (95% CI)	NR	NR	NA
AIMS score, mean (95% CI)	NR	NR	NA
MACTAR, mean (95% CI)	NR	NR	NA
SF-36 score, mean (95% CI)	NR	NR	NA
SF-12 score, mean (95% CI)	NR	NR	NA
NHP score, mean (95% CI)	NR	NR	NA
EQ-5D score, mean (95% CI)	NR	NR	NA
Pain score (VAS), mean (95% CI)	NR	NR	NA
Pain score (other than VAS; specify), mean (95% CI)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (95% CI)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	$p = 0.09$ (NS)
Femoral head penetration (mm/year), mean (95% CI)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR using cemented (Charnley) cup***	***THR using uncemented (Duraloc) cup***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	NR	NR	NR	NA
Aseptic loosening (any or total)	NR	NR	NR	NA
Infection	NR	NR	NR	$p = 0.06$ (NS)
Femoral neck fracture	NR	NR	NR	NA
Metallosis	NR	NR	NR	NA
Muscle weakness	NR	NR	NR	NA
Nerve palsy	NR	NR	NR	NA
Deep-vein thrombosis	NR	NR	NR	NA
Implant dislocation	NR	4/107 (3.8)	10/108 (9.25)	$p > 0.05$ (NS)
Others (pulmonary embolism)	NR	1/107 (0.93)	3/108 (2.8)	NR

a RR or risk difference.

Authors' conclusions

There is no statistically significant difference in clinical or radiological outcome between the cemented Charnley cup and the uncemented Duraloc cup after 10 years and no difference in implant survival after 12–14 years. The uncemented Duraloc cup is as good as the cemented Charnley cup after 10 years

Reviewers' conclusions

HHS improved in both groups after baseline and there was no significant between-group difference in mean score at any time. Survival of implants, revision rates, dislocation rates and radiographic results were similar between the treatment groups

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; UHMWPE, ultra-high molecular weight PE.

Cup liner bearing surface

McCalden *et al.*[145]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details

Country: Canada

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): specialty clinic

Number of centres: 1

Funding (government/private/manufacturer/other – specify): financial support for this study was provided by Zimmer, Inc. (Warsaw, IN, USA). The funding was used to support the salary of a research nurse who enrolled patients and gathered preoperative and postoperative clinical outcome data. In addition, the funding supported the salary of a research technician who performed the radiographic image analysis

Aim of the study

To report the clinical and radiographic results, after a minimum of 5 years of follow-up, of a randomised, blinded, controlled trial comparing a conventional PE liner with a first-generation highly cross-linked PE liner

Participants

Recruitment dates: November 1999 and October 2001

Total number of patients screened for inclusion eligibility: NR

Total number of patients randomised: 100

Inclusion criteria: a patient had to have degenerative arthritis of one hip requiring total hip arthroplasty, a designation of A or B according to the Charnley hip classification and be aged between 40 and 79 years

Exclusion criteria: Pre-existing bone disease (such as severe osteoporosis or osteomalacia), systemic conditions affecting bone density (such as inflammatory arthritis or renal disease) and a contralateral revision or poorly functioning THR

Characteristics of participants (total study sample): mean (range) age (years): 72 (56–79); women, *n* (%): 69 (69); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): NR

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): highly cross-linked PE acetabular cup liners	
Intervention 2 (e.g. THR 2): conventional PE acetabular cup liners	
Bilateral procedure (yes/no/NR): NR	
Implant manufacturer: Zimmer, Inc.	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	✓
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): PE wear (femoral head penetration rate)
Secondary outcomes (*list*): HHS, WOMAC score, SF-12 score
Imaging method used (i.e. conventional radiography, radiostereometry, none): radiographic image analysis
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 6 weeks, 3 months and yearly thereafter
Total length of follow-up: Mean 6.8 years

Number of patients			
	Total	***Highly cross-linked PE acetabular liners***	***Conventional PE acetabular liners***
Randomised	100	50	50
Analysed (if more than one follow-up point, choose and specify the last one)	98	49	49
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	11	1 loss to follow-up, 2 deaths	1 loss to follow-up, 7 deaths

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Highly cross-linked PE acetabular liners	The highly cross-linked PE liner, calcium stearate-free GUR 1050 resin, was also utilised to create compression-moulded sheets, which were then machined into the final implant geometry	Not clear
Conventional PE acetabular liners	Conventional PE is a PE for which no formal cross-linking or free-radical reduction process has been used, although there may be some degree of cross-linking in the material as a result of the sterilisation process. The conventional PE liners used in this study were made of calcium stearate-free GUR 1050 resin machined from compression-moulded sheet PE. The final implant was then sterilised with gamma radiation (25 kGy) in an inert nitrogen environment	Not clear

Patient baseline characteristics		
Characteristic	***Highly cross-linked PE acetabular liners***	***Conventional PE acetabular liners***
Age (years), mean (range)	72.31 (56–79)	72.58 (56–79)
Sex, female, *n*/*N* (%)	33/50 (66)	36/50 (72)
Weight (kg), mean (range)	NR	NR
BMI (kg/m^2), mean (range)	29.7 (22–39)	29.71 (18–48)
Primary OA, *n*/*N* (%)	NR	NR
Bilateral OA, *n*/*N* (%)	NR	NR
HHS, mean (SD)	38.96 (11.35)	35.64 (12.97)
OHS, mean (SD)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes			
Postprocedural follow-up assessment timing (specify): 1 year			
Outcome	***Highly cross-linked PE acetabular liners***	***Conventional PE acetabular liners***	***Between-group difference and p-value (or 95% CI)***[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	85.07 (10.32)	83.40 (13.10)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS score, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	83.04 (17.19)	81.61 (17.65)	$p > 0.05$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	MCS: 55.79 (7.38)	MCS: 56.01 (8.55)	$p > 0.05$ (NS)
	PCS: 42.20 (11.37)	PCS: 40.86 (11.11)	$p > 0.05$ (NS)
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

Efficacy outcomes

Post-procedural follow-up assessment timing (specify): 5 years

Outcome	*Highly cross-linked PE acetabular liners*	*Conventional PE acetabular liners*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	7/50 (14%)	2/50 (4%)	$p > 0.05$ (NS)
HHS, mean (SD)	86.03 (13.08)	83.11 (15.41)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS score, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	78.00 (19.42)	78.12 (18.20)	$p > 0.05$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	MCS: 55.24 (8.01)	MCS: 53.36 (10.13)	$p > 0.05$ (NS)
	PCS: 37.24 (12.16)	PCS: 40.00 (11.78)	$p > 0.05$ (NS)
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	1-year to 5-year penetration rate: 0.003 (–0.024 to 0.030)	1-year to 5-year penetration rate: 0.051 (0.029 to 0.073)	$p = 0.006$

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***Highly cross-linked PE acetabular liners***	***Conventional PE acetabular liners***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	Post operation 5 years	0	0	NA
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	5.5 years	Unclear	Unclear	NA
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	NA	NR	NR	NA
Others (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions

At a minimum of 5 years postoperatively, the steady-state femoral head penetration rate associated with the highly cross-linked PE liner was significantly lower than that associated with the conventional PE liner

Reviewers' conclusions

The study aimed to compare highly cross-linked PE with conventional PE in terms of the clinical and radiographic results and steady-state femoral head penetration rate (after bedding in). At a mean of 6.8 years postoperatively, there were no differences between the two PE groups with regard to the HHS, WOMAC score or SF-12 score. The mean femoral head penetration rate in years 1–5 was significantly lower in the highly cross-linked PE group. Men treated with a conventional PE liner had a significantly higher femoral head penetration rate than both men and women with a highly cross-linked liner. There was a high rate of clinical follow-up (98%). Concerns noted: single centre design; several manufacturers have now produced second-generation highly cross-linked PE that is in clinical use; a large number of patients demonstrated negative wear values; only approximately 78% of all available radiographs in the analysis

MCS, mental component summary score; NA, not applicable; NR, not reported; NS, not significant; PCS, physical component summary score; PE, polyethylene.

Engh *et al.*[113,114]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details
Country: USA
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 1
Funding (government/private/manufacturer/other – specify): NR
Aim of the study
To compare the clinical outcomes of THR patients randomised to either cross-linked or conventional non-cross-linked PE cup liners
Participants
Recruitment dates: January 1999–July 2000
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 220
Inclusion criteria: NR
Exclusion criteria: NR
Characteristics of participants (total study sample): mean (range) age (years): 62.0–2.5 (range 26–87); women, *n* (%): 122 (55.4); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 189 (86), avascular necrosis 13 (5.8), hip dysplasia 9 (4.0), fracture/trauma 13 (5.8), RA 5 (2.2)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): cross-linked PE cup liners (Marathon)	
Intervention 2 (e.g. THR 2): non-cross-linked (conventional) PE cup liners (Enduron)	
Bilateral procedure (yes/no/NR): yes	
Implant manufacturer: DePuy	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	✓
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): pain severity, function, mobility, radiography (wear rate, head penetration, head displacement, osteolysis), revisions, dislocation
Secondary outcomes (*list*): See above
Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 5 and 10 years
Total length of follow-up: 10 years

No. of patients			
	Total	***Cross-linked PE cup liners (Marathon)***	***Non-cross-linked (conventional) PE cup liners (Enduron)***
Randomised	220	111	109
Analysed	162	90	72
Losses to follow-up/ dropouts/sample attrition	NR	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Cross-linked PE cup liners (Marathon)	Patients were implanted with a Duraloc 100 (DePuy) cup incorporating a 4-mm lateralised liner. The PE liner was secured in the Duraloc shell by a peripheral locking ring that engaged a groove machined into the liner and shell. The surgical technique for the preparation of the acetabulum involved reaming to a diameter that was 1 mm smaller than the final component, allowing the cup to be press-fit without additional screw fixation. The hemispheric Duraloc 100 cup features sintered beads that create a three-dimensional porous coating for bone ingrowth and a single apical hole. On the femoral side, an extensively porous-coated stem (Anatomical Medullary Locking®) (DePuy) was used with a 28-mm cobalt–chromium alloy femoral head	NR
Non-cross-linked (conventional) PE cup liners (Enduron)	See above	NR

Patient baseline characteristics		
Characteristic	***Cross-linked PE cup liners (Marathon)***	***Non-cross linked (conventional) PE cup liners (Enduron)***
Age (years), mean (SD)	62.5 (10.6)	62.0 (11.1)
Sex, female, *n/N* (%)	65/111 (58.0)	57/109 (52.3)
Weight (kg), mean (SD)	84.3 (21.3)	81.6 (18.1)
BMI (kg/m^2), mean (SD)	28.6 (5.5)	27.9 (5.1)
Primary OA, *n/N* (%)	99/111 (89.1)	90/109 (82.5)
Bilateral OA, *n/N* (%)	NR	NR
HHS, mean (SD)	88.0 (14.0)	86.0 (15.0)
OHS, mean (SD)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 5 years

Outcome	*Cross-linked PE cup liners (Marathon)*	*Non-cross linked (conventional) PE cup liners (Enduron)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	0.24 (0.42)	1.26 (0.62)	$p < 0.001$

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 10 years

Outcome	*Cross-linked PE cup liners (Marathon)*	*Non-cross linked (conventional) PE cup liners (Enduron)*	*Between-group difference and* p *value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	17/111 (15.0)	15/109 (13.7)	$p > 0.05$
HHS, mean (SD)	88.0 (14.0)	86.0 (15.0)	$p = 0.49$
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	2/111 (2.0)	11/109 (10.0)	$p < 0.05$
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate), mean (SD)	98.1 (2.7)	92.6 (5.3)	$p = 0.02$
Femoral head penetration (mm/year), mean (SD)	0.06 (0.05)	0.22 (0.11)	$p < 0.001$

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***Cross-linked PE cup liners (Marathon)***	***Non-cross-linked (conventional) PE cup liners (Enduron)***	***Between-group difference and* p*-value (or 95% CI)*[a]
Osteolysis (femoral, acetabular)	Post	None	15 (22.0)	< 0.001
Aseptic loosening (femoral, acetabular)	Post	None	None	NR
Infection	NA	NR	NR	NA
Femoral neck fracture	Post	2 (1.8)	None	NR
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	Post	NR	NR	NA
Others (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions
No difference between the groups in mortality and disability scores; cross-linked PE cup liners showed significantly better 10-year implant survival rates, revision rates, osteolysis rates and femoral head penetration than conventional non-cross-linked PE cup liners
Reviewers' conclusions
Cross-linked PE cup liner recipients experienced significantly improved rates of implant survival, dislocation, revision, osteolysis and femoral head penetration at 10 years post operation than conventional non-cross-linked PE cup liner recipients; there were no between-group differences in mortality and clinical scores

NA, not applicable; NR, not reported; PE, polyethylene.

Cup shell design

Capello *et al.*,[115] D'Antonio *et al.*[116,117] and Mesko *et al.*[118]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: USA

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): specialty clinic

Number of centres: 16

Funding (government/private/manufacturer/other – specify): Stryker Orthopaedics (Mahwah, NJ, USA)

Aim of the study

To compare clinical and radiography outcomes between patients receiving THRs with ceramic-on-ceramic bearings and patients receiving THRs with metal-on-PE bearing, and to compare the results between two groups of patients receiving THRs with ceramic-on-ceramic bearings but with different cup designs (porous-coated shell vs. arc-deposited HA-coated shell)

Participants

Recruitment dates: October 1996 and October 1998

Total no. of patients screened for inclusion eligibility: NR

Total no. of patients randomised: 328

Inclusion criteria: patients aged 21–75 years, not morbidly obese, clinically qualified for THR, diagnosis of OA, traumatic arthritis, avascular necrosis, slipped capital epiphysis, pelvic fracture, femoral fracture, failed fracture, fixation or diastrophic variant, absence of active infection in the affected hip or no previous THR, no psychiatric disorder, senile dementia, Alzheimer's disease, presence of alcohol, or substance abuse, no neuromuscular or neurosensory deficiency, no systemic disorder, not immunologically suppressed, not receiving steroids in excess of physiological dose requirements, skeletally mature, not pregnant and no plans to relocate to another geographical area before completion of the study

Exclusion criteria: see above

Characteristics of participants (total study sample): mean (SD) age (years): 53–55 (10.0); women, *n* (%): 124 (38); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 262 (80), avascular necrosis 49 (15), other 17 (5)

Intervention (keep the same order as in the paper)

Intervention 1 (e.g. THR 1): THR with ceramic-on-ceramic bearings (titanium porous-coated shell) – system I

Intervention 2 (e.g. THR 2): THR with ceramic-on-ceramic bearings (titanium arc-deposited HA-coated shell) – system II

Intervention 3 (e.g. THR 3): THR with metal-on-PE bearings (titanium porous-coated shell) – system III

Intervention 4 (e.g. THR 4): this arm was added later and was not a randomised arm; therefore not extracted – system IV

Bilateral procedure (yes/no/NR): yes

Implant manufacturer: Stryker Orthopaedics

Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR

If several types of THR are compared, indicate the basis for comparison (tick all that apply):

Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	✓
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	✓
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	✓
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)

Primary outcomes (*list*): Implant survival rate, revision, HHS, satisfaction, radiographic measures (stem subsidence, periosteal cortical hypertrophy, cancellous condensation, head penetration, osteolysis, radiolucency, cup migration, stem stability, scalloping and wear rate) and complications (e.g. fractures, squeaking, dislocation)

Secondary outcomes (*list*): See above

Imaging method used (i.e. conventional radiography, radiostereometry, none): Conventional radiography

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): Preoperatively, 6–8 weeks postoperatively, 6 months, years 1–5 and 7 and 10 years

Total length of follow-up: 10 years

No. of patients

	Total	***System I***	***System II***	***System III***
Randomised	328	113	109	106
Analysed (if more than one follow-up point, choose and specify the last one)	305	101	100	104
Losses to follow-up/dropouts/ sample attrition (if more than one follow-up point, choose and specify the last one)	NR	NR	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
System I	THR with alumina ceramic femoral head-on-alumina ceramic cup liner/ insert with titanium porous-coated acetabular shell; HA-coated titanium femoral component; the head size was dictated by the outer diameter of the shell. Only 32-mm inside diameter acetabular components and the corresponding alumina ceramic femoral heads were available for shells with an outer diameter of ≥ 52 mm	NR
System II	THR with alumina ceramic femoral head-on-alumina ceramic cup liner/ insert with titanium arc-deposited HA-coated shell; HA-coated titanium femoral component; the head size was dictated by the outer diameter of the shell. Only 32-mm inside diameter acetabular components and the corresponding alumina ceramic femoral heads were available for shells with an outer diameter of ≥ 52 mm	NR
System III	THR with metal (cobalt–chromium) femoral head-on-PE cup liner/insert with titanium porous-coated shell; HA-coated titanium femoral component; surgeons were permitted to use a 26-mm, 28-mm or 32-mm head as they saw fit	NR

Patient baseline characteristics			
Characteristic	***System I***	***System II***	***System III***
Age (years), mean (SD)	53 (11.4)	54 (10.7)	55 (10.7)
Sex, female, *n*/*N* (%)	41/113 (35.0)	41/109 (37.0)	42/106 (39.0)
Weight (kg), mean (SD)	85.0 (17.9)	87.8 (17.3)	85.9 (17.7)
BMI (kg/m^2), mean (SD)	NR	NR	NR
Primary OA, *n*/*N* (%)	95/113 (84.0)	86/109 (79.0)	81/106 (77.0)
Bilateral OA, *n*/*N* (%)	NR	NR	NR
HHS, mean (SD)	NR	NR	NR
OHS, mean (SD)	NR	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Postprocedural follow-up assessment timing (specify): 5 years

Outcome	***System I***	***System II***	***System III***	***Between-group difference and p-value (or 95% CI)***[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NR	NA
HHS, mean (SD)	97.0 (NR)	96.4 (NR)	97.0 (NR)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NR	NA
HOOS, mean (SD)	NR	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NR	NA
NHP score, mean (SD)	NR	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NR	NA
Revision rate, *n/N* (%)	2/113 (1.8)	4/109 (3.7)	8/106 (7.5)	$p = 0.045$ (SS), systems I and II vs. system III
Time to revision (years), mean (SD)	NR	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	94.1 (87.5 to 100.0)	94.1 (87.5 to 100.0)	92.2 (87.0 to 97.4)	$p = 0.03$ (SS), systems I and II vs. system III
Femoral head penetration (mm/year), mean (SD)	NR	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 8.5–10 years

Outcome	*System I*	*System II*	*System III*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NR	NA
HHS, mean (SD)	96.0 (NR)	96.7 (NR)	96.4 (NR)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NR	NA
HOOS, mean (SD)	NR	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NR	NA
NHP score, mean (SD)	NR	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NR	NA
Revision rate, *n/N* (%)	2/113 (1.8)	2/109 (1.8)	5/106 (4.7)	Calculated RR 0.38 (0.10 to 1.39), $p = 0.08$ (NS), systems I and II vs. system III
Time to revision (years), mean (SD)	NR	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	95.9 (91.9 to 97.9)	95.9 (91.9 to 97.9)	91.3 (83.4 to 95.6)	$p = 0.01$ (SS), systems I and II vs. system III
Femoral head penetration (mm/year), mean (SD)	NR	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)					
Complication	***Time of occurrence (peri-/post-operational)***	***System I***	***System II***	***System III***	***Between-group difference and* p*-value (or 95% CI)*[a]
Osteolysis (any or total)	Post	1/113 (0.9)	2/109 (1.8)	15/106 (14.0)	Calculated RR 0.10 (0.02 to 0.32), $p < 0.001$ (SS), systems I and II vs. system III
Aseptic loosening (any or total)	NA	NR	NR	NR	NA
Infection	NA	NR	NR	NR	NA
Femoral neck fracture	Post	0/113 (0.0), ceramic fracture	0/109 (0.0), ceramic fracture	NA	NR
Metallosis	NA	NR	NR	NR	NA
Muscle weakness	NA	NR	NR	NR	NA
Nerve palsy	NA	NR	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NR	NA
Implant dislocation	Post	2/113 (1.8)	3/109 (2.8)	5/106 (4.7)	Calculated RR 0.47 (0.14 to 1.61), $p = 0.25$ (NS), systems I and II vs. system III
Others (specify)	NA	NR	NR	NR	NA

a RR or risk difference.

Authors' conclusions

Alumina ceramic-on-ceramic bearings performed better than metal-on-PE bearings in terms of implant survival (until any revision), osteolysis and revision rate. Dislocation rate and HHS were similar between the treatment groups

Reviewers' conclusions

Alumina ceramic-on-ceramic bearings performed better than metal-on-PE bearings in terms of implant survival (until any revision) and osteolysis. The results regarding revision rates and dislocation rates are inconclusive because of low counts and wide CIs compatible with both directions of the effect. The mean HHS was similar between the treatment groups

HA, hydroxyapatite; NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; SS, statistically significant.

Cup/stem fixation

Corten *et al.*,[119,122] Bourne and Corten[121] and Laupacis *et al.*[120]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details
Country: Canada
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): NR
Number of centres: NR
Funding (government/private/manufacturer/other – specify): Medical Research Council of Canada (currently Canadian Institutes of Health Research)
Aim of the study
To compare the effects of cemented cup/stem (Mallory Head) and uncemented cup/stem (Mallory Head) prostheses on long-term follow-up for mortality, revision, time to revision, health-related quality of life and radiography signs
Participants
Recruitment dates: October 1987 and January 1992
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 250
Inclusion criteria: OA of the hip and undergoing a unilateral primary arthroplasty
Exclusion criteria: age > 75 years, severe symptomatic OA of either knee or the contralateral hip, a previous arthroplasty of the ipsilateral hip, arthroplasty on the contralateral side > 5 years before the most recent arthroplasty or had infectious arthritis
Characteristics of participants (total study sample): mean (SD) age (years): 64 (7.0); women, *n* (%): 120 (48); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): NR

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): THR using cemented femoral and cemented acetabular components	
Intervention 2 (e.g. THR 2): THR using uncemented femoral and uncemented acetabular components	
Bilateral procedure (yes/no/NR): no	
Implant manufacturer: Biomet (Warsaw, IN, USA)	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	✓
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	✓

Outcomes (study based)
Primary outcomes (*list*): HHS, revision-free survival rate, mortality, health-related quality of life (WOMAC, MACTAR, Merle d'Aubigné and Postel scores), radiography signs
Secondary outcomes (*list*): see above
Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3 and 6 months and 1, 3, 5, 7, 15 and 20 years
Total length of follow-up: 20 years

Number of patients			
	Total	***THR using cemented cup and stem***	***THR using uncemented cup and stem***
Randomised	250	124	126
Analysed (if more than one follow-up point, choose and specify the last one)	93	41	52
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	30	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR using cemented cup and stem	All patients were operated on by either surgeon with use of an identical direct lateral approach and within a vertical laminar airflow enclosure in which the surgical team wore body-exhaust suits. The prosthesis (Mallory Head, Biomet), including the modular head, was made from a titanium alloy and was available for insertion with cement. Both components (stem and cup) were inserted with cement. The cup was one piece, with the PE moulded directly into the titanium shell	NR
THR using uncemented cup and stem	The prosthesis (Mallory Head, Biomet), including the modular head, was made from a titanium alloy and was available for insertion without cement. The implant was a tapered design with a 3° taper and the proximal one-third was plasma sprayed. The acetabular component consisted of a finned titanium shell that was plasma sprayed and had a hexlock mechanism to capture the PE insert	NR

Patient baseline characteristics		
Characteristic	***THR using cemented cup and stem***	***THR using uncemented cup and stem***
Age (years), mean (SD)	64 (8.0)	64 (7.0)
Sex, female, *n/N* (%)	60/124 (48)	60/126 (46)
Weight (kg), mean (SD)	NR	NR
BMI (kg/m^2), mean (SD)	NR	NR
Primary OA, *n/N* (%)	NR	NR
Bilateral OA, *n/N* (%)	0/124 (0.0)	0/126 (0.0)
HHS, mean (SD)	44 (11.0)	43 (10.0)
OHS, mean (SD)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 3 months

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	41 (12.0)	41 (11.0)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	5.8 (1.9)	5.6 (2.2)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR (pain, stiffness, physical fitness domains reported separately)	NR (pain, stiffness, physical fitness domains reported separately)	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–5.3 (2.5)	–5.2 (2.2)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 6 months

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	47 (12)	50 (13)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	6.7 (2.1)	7.0 (2.2)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–6.6 (1.9)	–6.4 (2.1)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 1 year

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	52 (10.0)	53 (11.0)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	7.5 (1.8)	7.4 (2.1)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–7.0 (1.8)	–6.9 (2.0)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 3 years

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	50 (14.0)	52 (11.0)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	7.1 (2.2)	6.9 (2.1)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NR
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–6.6 (2.3)	–6.4 (2.3)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 5 years

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	47 (14.0)	48 (13.0)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	6.5 (2.3)	6.6 (2.4)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–6.0 (2.8)	–6.2 (2.4)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	97.0 (95.0 to 99.0)	100 (NR)	$p < 0.05$ (SS)
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Post-procedural follow-up assessment timing (specify): 7 years

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	18/124 (14.5)	17/126 (13.5)	NR
HHS, mean (SD)	44 (15)	46 (14)	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	6.1 (2.6)	6.5 (2.8)	NR
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	–6.2 (2.8)	–6.0 (2.6)	$p > 0.05$ (NS)
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	13/124 (10.5)	6/126 (4.7)	$p = 0.11$ (NS)
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	83.0 (79.0 to 87.0)	94.0 (92.0 to 96.0)	$p < 0.05$ (SS)
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 15 years

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	66.0 (61.0 to 71.0)	0.80% (76.0 to 84.0)	$p < 0.05$ (SS)
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 20 years

Outcome	*THR using cemented cup and stem*	*THR using uncemented cup and stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR; domain specific scores reported separately	NR; domain specific scores reported separately	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	48/124 (38.7)	31/126 (24.6)	$p = 0.02$ (S)
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	48.0 (41.0 to 55.0)	69.0% (64.0 to 74.0)	$p < 0.05$ (SS)
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR using cemented cup and stem***	***THR using uncemented cup and stem***	***Between-group difference and p-value (or 95% CI)[a]***
Osteolysis (any or total)	NA	NR	NR	NR
Aseptic loosening (any or total)	NR	9/124 (7.2)	4/126 (3.2)	NR
Infection	NA	NR	NR	NR
Femoral neck fracture	NA	NR	NR	NR
Metallosis	NA	NR	NR	NR
Muscle weakness	NA	NR	NR	NR
Nerve palsy	NA	NR	NR	NR
Deep-vein thrombosis	NA	NR	NR	NR
Implant dislocation	NA	NR	NR	NR
Others (specify)	NA	NR	NR	NR

a RR or risk difference.

Authors' conclusions

After 7 years of follow-up,[120] post-surgery implant (all components) revision-free survival was similar in the cemented and uncemented THR groups ($p = 0.11$); however, the uncemented group experienced less femoral component revisions than the cemented group ($p = 0.002$). Both treatment groups improved postoperatively in hip scores (WOMAC, MACTAR, HHS) but there was no significant between-group difference. After 15–20 years of follow-up,[119,121] the uncemented implant group had a significantly better overall implant (all components), stem or cup revision-free survival than the cemented implant group ($p = 0.01$). The revision rate was higher in the cemented group (38.7% vs. 24.6%, $p = 0.02$)

Reviewers' conclusions

The long-term overall and component-specific revision-free survival postoperatively was significantly different and favoured the uncemented group over the cemented group. Both treatment groups experienced improved hip functioning and symptoms with no significant differences between the groups. The revision rate was lower in the uncemented group than in the cemented group

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; SS, statistically significant.

Femoral head size

Howie *et al.*[123]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details
Country: Australia
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic/tertiary referral
Number of centres: 14
Funding (government/private/manufacturer/other – specify): the National Health and Medical Research Council of Australia and Zimmer, Inc.
Aim of the study
To compare the incidence of dislocation at 1 year after THR between two groups of patients who had received 36-mm and 28-mm femoral head articulations
Participants
Recruitment dates: September 2001–June 2007
Total number of patients screened for inclusion eligibility: 2319
Total number of patients randomised: 557
Inclusion criteria: patients aged $\geq$ 60 years with OA and RA referred for THR
Exclusion criteria: patients aged $<$ 60 years with diagnoses other than OA, RA, inflammatory arthritis or previous fracture/dislocation/surgery involving the hip, abnormal acetabulum, neuromuscular disorder affecting the hip, tumour of the hip, unable to provide consent, unable to complete follow-up
Characteristics of participants (total study sample): mean (range) age (years): 72.3 (59–93); women, *n* (%): 327 (58.7); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 534 (96.0)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): 36-mm femoral head	
Intervention 2 (e.g. THR 2): 28-mm femoral head	
Bilateral procedure (yes/no/NR): no	
Implant manufacturer: Zimmer, Inc.	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	✓
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): incidence of dislocation, revision
Secondary outcomes (*list*): the position of the acetabular component and femoral head
Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 6 weeks, 3 months and 1 year
Total length of follow-up: 1 year

No. of patients			
	Total	***36-mm femoral head***	***28-mm femoral head***
Randomised	557	273	284
Analysed (if more than one follow-up point, choose and specify the last one)	533	258	275
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	7	6	1

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
36-mm femoral head	All arthroplasties were performed with use of uncemented acetabular components, which comprised a cluster three-holed acetabular shell (Trilogy) fixed with one or two screws and a 10° elevated 36- or 28-mm inner diameter highly cross-linked PE liner (Longevity). A cemented femoral stem was used for all arthroplasties (CPT). During the trial, the taper of the CPT femoral stem was changed from a 6° taper to a 12/14 taper by the manufacturer. All arthroplasties were performed through a posterior surgical approach. The operative technique for insertion of the acetabular component through a posterior approach included reliance mainly on the alignment guide and confirmation using the surgeon's judgement that the component was reasonably positioned	NR
28-mm femoral head	See above	NR

Patient baseline characteristics		
Characteristic	***36-mm femoral head***	***28-mm femoral head***
Age (years), mean (95% CI)	72.3 (71.5 to 73.0)	72.3 (71.6 to 73.1)
Sex, female, *n/N* (%)	152/273 (56.0)	175/284 (61.3)
Weight (kg), mean (95% CI)	NR	NR
BMI (kg/m^2), mean (95% CI)	28.0 (27.4 to 28.7)	28.4 (27.8 to 29.0)
Primary OA, mean (95% CI), %	96.3 (95% CI 94.1 to 98.6)	95.4 (95% CI 93.0 to 97.9)
Bilateral OA, mean (95% CI), %	0	0
HHS, mean (95% CI)	NR	NR
OHS, mean (95% CI)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 1 year

Outcome	*36-mm femoral head*	*28-mm femoral head*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	5/273 (1.8)	2/284 (0.7)	RR 2.58 (0.53 to 13.20)
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	4/273 (1.46)	6/284 (2.11)	RR 0.69 (0.19 to 2.43)
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***36-mm femoral head***	***28-mm femoral head***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	NA	NR	NR	NA
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	Post	2/258 (0.8)	12/275 (4.4)	RR 0.17 (0.04 to 0.78)
Others (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions

Patients who received 36-mm femoral head implants had a significantly lower dislocation risk than those who received 28-mm femoral head implants 1 year after the operation. The between-group differences in the occurrence of revision and death were not statistically significant

Reviewers' conclusions

The authors demonstrated that 36-mm femoral head implants conferred a benefit in terms of a significantly reduced risk of dislocation compared with smaller 28-mm femoral head implants; the results for revision are inconclusive because of the wide 95% CIs that were compatible with benefit as well as harm associated with the use of 36-mm femoral head implants relative to 28-mm femoral head implants

NA, not applicable; NR, not reported; PE, polyethylene.

Femoral head bearing

Lewis *et al.*[124]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Study details
Country: Canada
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 1
Funding (government/private/manufacturer/other – specify): Smith & Nephew, Inc.
Aim of the study
To compare clinical outcomes in patients who received a THR with oxinium vs. cobalt–chromium femoral heads
Participants
Recruitment dates: NR
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 100
Inclusion criteria: NR
Exclusion criteria: NR
Characteristics of participants (total study sample): mean (SD) age (years): 51 (11.0); women, *n* (%): 48 (48.0); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 76 (76.0), advanced avascular necrosis 14 (14.0), developmental hip dysplasia 7 (7.0) and RA 3 (3.0)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): THR with oxinium femoral heads	
Intervention 2 (e.g. THR 2): THR with cobalt–chromium femoral heads	
Bilateral procedure (yes/no/NR): NR	
Implant manufacturer: Smith & Nephew	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	✓
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): stem survival rate (before dislocation), dislocation, loosening, HHS, WOMAC score and SF-12 (physical and mental components) score
Secondary outcomes (*list*): see above
Imaging method used (i.e. conventional radiography, radiostereometry, none): NR
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 2 years
Total length of follow-up: 2 years

Number of patients			
	Total	***THR with oxinium femoral heads***	***THR with cobalt–chromium femoral heads***
Randomised	100	50	50
Analysed (if more than one follow-up point, choose and specify the last one)	NR	NR	NR
Losses to follow-up/dropouts/ sample attrition (if more than one follow-up point, choose and specify the last one)	NR	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR with oxinium femoral heads	The anterolateral approach to the hip was used in 17 patients and the posterior approach was used in 33 patients. In total, 46 patients received an Echelon stem and the remaining four received a Synergy stem. The acetabular components were press-fit uncemented Reflection cups paired with either standard PE (22 cases) or highly cross-linked PE (28 cases)	NR
THR with cobalt–chromium femoral heads	The anterolateral approach to the hip was used in five patients and the posterior approach was used in 45 patients. In total, 30 patients received an Echelon stem whereas the remaining 20 received a Synergy stem. The acetabular components were press-fit uncemented Reflection cups paired with either standard PE (31 cases) or highly cross-linked PE (19 cases)	NR

Patient baseline characteristics		
Characteristic	***THR with oxinium femoral heads***	***THR with cobalt–chromium femoral heads***
Age (years), mean (SD)	51 (10.8)	51 (11.0)
Sex, female, *n*/*N* (%)	24/50 (48.0)	24/50 (48.0)
Weight (kg), mean (SD)	NR	NR
BMI (kg/m^2), mean (SD)	NR	NR
Primary OA, *n*/*N* (%)	NR	NR
Bilateral OA, *n*/*N* (%)	NR	NR
HHS, mean (SD)	NR	NR
OHS, mean (SD)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 2 years

Outcome	*THR with oxinium femoral heads*	*THR with cobalt–chromium femoral heads*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (range)	92 (65–100)	92.5 (60–100)	$p > 0.159$ (NS)
OHS, mean (range)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (range)	NR	NR	NA
HOOS, mean (range)	NR	NR	NA
LISOH score, mean (range)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (range)	NR	NR	NA
UCLA activity score, mean (range)	NR	NR	NA
WOMAC score, mean (range)	84.9 (NR)	87.0 (NR)	$p > 0.159$ (NS)
AIMS score, mean (range)	NR	NR	NA
MACTAR score, mean (range)	NR	NR	NA
SF-36 score, mean (range)	NR	NR	NA
SF-12 score, mean (range)	PCS: 45.2 (27.1–56.7)	PCS: 49.2 (26.3–61.8)	$p > 0.05$ (NS)
	MCS: 53.8 (39.2–65.5)	MCS: 52.57 (34.3–64.0)	
NHP score, mean (range)	NR	NR	NA
EQ-5D score, mean (range)	NR	NR	NA
Pain score (VAS), mean (range)	NR	NR	NA
Pain score (other than VAS; specify), mean (range)	NR	NR	NA
Revision rate, *n/N* (%)	1/50 (2.0)	1/50 (2.0)	NR
Time to revision (years), mean (range)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	98.0 (NR)	98.0 (NR)	NR
Femoral head penetration (mm/year), mean (range)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR with oxinium femoral heads***	***THR with cobalt–chromium femoral heads***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	Post	0/50 (0.0)	1/50 (2.0)	NR
Infection	Post	1/50 (2.0)	1/50 (2.0)	NR
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	Post	2/50 (4.0)	1/50 (2.0)	NR
Others (stroke)	Post	0 (0.0)	1/50 (2.0)	NR

a RR or risk difference.

Authors' conclusions

For both groups, there was a significant improvement in mean score for all clinical measures after baseline (HHS, WOMAC, SF-12). The mean scores for the clinical measures were not significantly different between the groups

Reviewers' conclusions

The statistical data are insufficient to verify the findings; the small sample size may have led to inconclusive results

MCS, mental component summary score; NA, not applicable; NR, not reported; NS, not significant; PCS, physical component summary score; PE, polyethylene.

Femoral head bearing-on-cup liner bearing

Amanatullah *et al.*[125]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details
Country: USA
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 9
Funding (government/private/manufacturer/other – specify): NR
Aim of the study
To compare the clinical performance and evaluate the wear rate of ceramic-on-ceramic and ceramic-on-PE bearing surfaces
Participants
Recruitment dates: 1999–2001
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 312 patients (357 hips)
Inclusion criteria: patients were included if clinically indicated for a THR as a result of OA or RA and were aged 21–80 years with a HHS ≤ 60, were available for ≥ 2 years of clinical follow-up, were able to meet acceptable preoperative medical clearance and had no history of treatment for cardiac, pulmonary, haematological or any other medical condition that would pose an excessive operative risk
Exclusion criteria: NR
Characteristics of participants (total study sample): mean (range) age (years): approx. 52 (50–54); women, *n* (%): 122 (39); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): NR

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): ceramic-on-ceramic	
Intervention 2 (e.g. THR 2): ceramic-on-PE	
Bilateral procedure (yes/no/NR): yes [45 (21%)]	
Implant manufacturer: Smith & Nephew	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	✓
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	✓
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): clinical function (HHS, SF-12), wear rate, revision rate and complications (e.g. dislocation, loosening, migration, ossification, osteolysis)
Secondary outcomes (*list*): see above
Imaging method used (i.e. conventional radiography, radiostereometry, none): radiographs were analysed to determine linear and volumetric wear using the Avenger Digital Calliper (Avenger Products, Boulder City, NV, USA)
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3, 6, 12, 24 and 48 months as well as after 60 months
Total length of follow-up: > 60 months

No. of patients			
	Total	***Ceramic-on-ceramic***	***Ceramic-on-PE***
Randomised	312 patients (357 hips)	166 patients (196 hips)	146 patients (161 hips)
Analysed (if more than one follow-up point, choose and specify the last one)	312 patients (357 hips)	166 patients (196 hips)	146 patients (161 hips)
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	92	41	51

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Ceramic-on-ceramic	Each ceramic–ceramic articulation was implanted with a 28- or 32-mm alumina ceramic femoral head and an alumina ceramic acetabular cup liner. Alumina ceramic components were also sterilised with ethylene oxide gas. Metal components were sterilised with a minimum of 25 kGy of gamma irradiation. All implant components were supplied sterile to a sterility assurance level of 10 to 6 in protective packaging and trays. All implant component packages were inspected for puncture or damage before surgery	NR
Ceramic-on-PE	Each ceramic–PE articulation was implanted with a 28-mm alumina ceramic femoral head and an uncross-linked ultra-high molecular-weight PE acetabular cup liner sterilised with ethylene oxide gas and used within the expiration date	NR

Patient baseline characteristics		
Characteristic	***Ceramic-on-ceramic***	***Ceramic-on-PE***
Age (years), mean (SD)	50.4 (12.8)	54.7 (12.9)
Sex, female, *n* (%)	60 (36.1)	62 (42.5)
Weight (kg), mean (SD)	86.9 (20.0)	83.7 (18.5)
BMI (kg/m^2), mean (SD)	29.6 (12.4)	28.0 (5.1)
Primary OA, *n*/*N* (%)	NR	NR
Bilateral OA, *n*/*N* (%)	NR	NR
HHS, mean (SD)	NR	NR
OHS, mean (SD)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 60 months

Outcome	*Ceramic-on-ceramic*	*Ceramic-on-PE*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	$p > 0.05$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	$p > 0.05$ (NS)
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	11/196 (5.6)	3/161 (1.9)	$p = 0.059$ (NS)
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***Ceramic-on-ceramic***	***Ceramic-on-PE***	***Between-group difference and p-value (or 95% CI)[a]***
Osteolysis (any or total)	Postoperative	1/166 (0.6)	1/146 (0.6)	$p = 0.797$(NS)
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	Superficial	6/166 (3.6)	3/146 (2.0)	$p = 0.357$ (NS)
	Deep	1/166 (0.6)	2/146 (1.3)	$p = 0.909$ (NS)
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	Postoperative	3/166 (1.8)	2/146 (1.3)	$p = 0.909$ (NS)
Implant dislocation	Postoperative	10/166 (6.0)	9/146 (6.1)	$p = 0.672$ (NS)
Other (specify)				
Pulmonary embolus	Postoperative	2/166 (1.0)	1/146 (0.6)	$p = 0.573$

a RR or risk difference.

Authors' conclusions

In both groups the mean HHS, SF-12 mental component summary score and SF-12 physical component summary score were higher postoperatively, but there was no statistically significant differences between the groups at any follow-up time (data not reported). However, ceramic–PE couples did not offer sufficiently low linear wear rates to theoretically prevent osteolysis on longer-term follow-up

Reviewers' conclusions

The study aimed to compare the clinical performance and evaluate the wear rate of bearing surfaces using a prospective randomised study design. The study compared ceramic–ceramic and ceramic–PE articulations in THR. There was no statistically significant difference in clinical outcomes, dislocation or revision rates between the groups at any time interval, including at the last clinical follow-up point. Both ceramic–ceramic- and ceramic–PE-bearing surfaces appeared to have excellent short-term to midterm clinical results. The study does not have statistical power to make reliable conclusions about the dislocation rate of either group; caution is therefore needed when interpreting these findings. It is also noted that there was asymmetry in the number of patients in each of the study groups (e.g. mean age). There was also a change in method of randomisation during the course of the study: 'Once the Reflection Ceramic–Ceramic Hip System was approved by the FDA, our randomisation ended, and the total number of patients in each group was set regardless of prior power analyses or group symmetry' (p. 76). The majority of the conclusions appear justified although there was incomplete reporting of data for several outcome measures at different follow-up periods

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene.

Kadar *et al.*[126]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: Norway

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): specialty clinic

Number of centres: 4

Funding (government/private/manufacturer/other – specify): OrtoMedic AS, Smith & Nephew Norway AS and the Regional Health Board of Western Norway

Aim of the study

To evaluate wear and migration patterns between cemented highly cross-linked Reflection All-Poly (HXLPE) cups and All-Poly cups articulated with either oxinium or cobalt–chromium femoral heads compared with the Charnley Ogee prosthesis

Participants

Recruitment dates: November 2004–June 2007

Total number of patients screened for inclusion eligibility: NR

Total number of patients randomised: 150

Inclusion criteria: primary or secondary OA of the hip

Exclusion criteria: BMI > 35 kg/m^2, uncompensated cardiopulmonary disease, malignant disease, dementia, RA or other serious systemic diseases

Characteristics of participants (total study sample): mean (SD) age (years): 70 (6.0); women, *n* (%): 105 (70); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): primary arthrosis 129 (86.0), secondary arthrosis 21 (14.0)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): Charnley Ogee prosthesis	
Intervention 2 (e.g. THR 2): cobalt–chromium-on-PE articulation	
Intervention 3 (e.g. THR 3): oxinium-on-PE articulation	
Intervention 4 (e.g. THR 4): cobalt–chromium-on-HXLPE articulation	
Intervention 5 (e.g. THR 5): oxinium-on-HXLPE articulation	
Bilateral procedure (yes/no/NR): yes	
Implant manufacturer: Smith & Nephew (Memphis, TN, USA) (HXLPE articulations); DePuy (Charnley Ogee)	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): patients were allowed partial weight bearing with crutches from the first postoperative day. Restrictions were discontinued 6 weeks postoperatively	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	✓
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	✓
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): HHS, head penetration and cup migration/rotation
Secondary outcomes (*list*): see above
Imaging method used (i.e. conventional radiography, radiostereometry, none): radiostereometric analysis
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): preoperative, after operation at 3, 12 and 24 months (HHS), after operation at 3, 6, 12 and 24 months (radiostereometric examinations)
Total length of follow-up: 2 years

Number of patients						
	Total	***Charnley Ogee***	***All-Poly cobalt–chromium***	***All-Poly oxinium***	***All-Poly HXLPE cobalt–chromium***	***All-Poly HXLPE oxinium***
Randomised	150	30	30	30	30	30
Analysed (if more than one follow-up point, choose and specify the last one)	128	27	27	21	29	24
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	4	0	1	2	0	1

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Charnley Ogee	Charnley monoblock stainless steel femoral stem with a 22.2-mm head articulated with a cemented Charnley Ogee UHMWPE (GUR 1050) acetabular cup that was gamma sterilised with 2.5 Mrad in nitrogen. The smaller acetabuli received a 40-mm diameter cup	NR
All-Poly cobalt–chromium	Spectron EF femoral stem with a 28-mm cobalt–chromium femoral head and a Reflection All-Poly UHMWPE (GUR 1050) cup that was sterilised with ethylene oxide. The cup diameter corresponded with the largest reamer used. In total, 6–12 anchorage holes were drilled. The components were inserted with Palacos R with gentamicin cement using a third-generation cementing technique. Femoral stem insertion was performed at 5 minutes after cement mixing; cup insertion was performed at 6 minutes after cement mixing	NR
All-Poly oxinium	Spectron EF femoral stem with a 28-mm oxinium femoral head and a Reflection All-Poly UHMWPE (GUR 1050) cup that was sterilised with ethylene oxide. The cup diameter corresponded with the largest reamer used. In total, 6–12 anchorage holes were drilled. The components were inserted with Palacos R with gentamicin cement using a third-generation cementing technique. Femoral stem insertion was performed at 5 minutes after cement mixing; cup insertion was performed at 6 minutes after cement mixing	NR
All-Poly HXLPE cobalt–chromium	Spectron EF femoral stem with a 28-mm cobalt–chromium femoral head and a Reflection All-Poly HXLPE (GUR 1050) cup irradiated with 10 Mrad, melted at 135°C, and ethylene oxide sterilised. The cup diameter corresponded with the largest reamer used. In total, 6–12 anchorage holes were drilled. The components were inserted with Palacos R with gentamicin cement using a third-generation cementing technique. Femoral stem insertion was performed at 5 minutes after cement mixing; cup insertion was performed at 6 minutes after mixing	NR
All-Poly HXLPE oxinium	Spectron EF femoral stem with a 28-mm oxinium femoral head and a Reflection All-Poly HXLPE (GUR 1050) cup irradiated with 10 Mrad, melted at 135°C, and ethylene oxide sterilised. The cup diameter corresponded with the largest reamer used. In total, 6–12 anchorage holes were drilled. The components were inserted with Palacos R with gentamicin cement using a third-generation cementing technique. Femoral stem insertion was performed at 5 minutes after cement mixing; cup insertion was performed at 6 minutes after mixing	NR

Patient baseline characteristics

Characteristic	*Charnley Ogee*	*All-Poly cobalt–chromium*	*All-Poly oxinium*	*All-Poly HXLPE cobalt–chromium*	*All-Poly HXLPE oxinium*
Age (years), mean (SD)	70 (6.1)	69 (5.9)	69 (6.7)	70 (5.3)	70 (5.4)
Sex, female, *n/N* (%)	20/30 (66.6)	20/30 (66.6)	23/30 (76.6)	20/30 (66.6)	22/30 (73.3)
Weight (kg), mean (SD)	76 (14.9)	76 (11.1)	72 (13.9)	80 (14.8)	76 (14.6)
BMI (kg/m^2), mean (SD)	NR	NR	NR	NR	NR
Primary OA, *n/N* (%)	28/30 (93.3)	26/30 (86.6)	26/30 (86.6)	22/30 (73.3)	27/30 (90.0)
Bilateral OA, *n/N* (%)	NR	NR	NR	NR	NR
HHS, mean (SD)	45 (NR)	41 (NR)	47 (NR)	47 (NR)	40 (NR)
OHS, mean (SD)	NR	NR	NR	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 2 years

Outcome	*Charnley Ogee*	*All-Poly cobalt–chromium*	*All-Poly oxinium*	*All-Poly HXLPE cobalt–chromium*	*All-Poly HXLPE oxinium*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NR	NR	NR	NA
HHS, mean (SD)	91 (10.8)	91 (8.5)	91 (11.1)	93 (11.3)	88 (9.5)	$p = 0.7$ (NS)
OHS, mean (SD)	NR	NR	NR	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NR	NR	NR	NA
HOOS, mean (SD)	NR	NR	NR	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NR	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NR	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NR	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NR	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NR	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NR	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NR	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NR	NR	NR	NA
NHP score, mean (SD)	NR	NR	NR	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NR	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NR	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NR	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NR	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NR	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NR	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	0.19 (0.16 to 0.23)	0.40 (0.33 to 0.46)	0.44 (0.37 to 0.51)	0.19 (0.15 to 0.23)	0.18 (0.13 to 0.22)	$p < 0.001$ [All-Poly/CoCr and All-Poly/oxinium vs. Charnley Ogee (favoured)]; $p = 0.94$ (All-Poly HXLPE/CoCr vs. Charnley Ogee); $p = 0.57$ (All-Poly HXLPE/oxinium vs. Charnley Ogee)

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)

Complication	*Time of occurrence (peri-/post-operational)*	*Charnley Ogee*	*All-Poly cobalt–chromium*	*All-Poly oxinium*	*All-Poly HXLPE cobalt–chromium*	*All-Poly HXLPE oxinium*	*Between-group difference and* p-*value (or 95% CI)*[a]
Osteolysis (any or total)	NA	NR	NR	NR	NR	NR	NA
Aseptic loosening (any or total)	NA	NR	NR	NR	NR	NR	NA
Infection	NA	NR	NR	NR	NR	NR	NA
Femoral neck fracture	NA	NR	NR	NR	NR	NR	NA
Metallosis	NA	NR	NR	NR	NR	NR	NA
Muscle weakness	NA	NR	NR	NR	NR	NR	NA
Nerve palsy	NA	NR	NR	NR	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NR	NR	NR	NA
Implant dislocation	NA	NR	NR	NR	NR	NR	NA
Other (specify)	NA	NR	NR	NR	NR	NR	NA

a RR or risk difference.

Authors' conclusions

At 2 years of follow-up, HHS improved after baseline for all five groups but there was no significant difference between the groups. Femoral head penetration was significantly improved (reduced) in the All-Poly HXLPE/cobalt–chromium, All-Poly HXLPE/oxinium and Charnley Ogee groups compared with the All-Poly/cobalt–chromium or All-Poly/oxinium group. The low penetration rate of the Charnley Ogee prosthesis may be explained by the smaller femoral head than those used in the All-Poly HXLPE/cobalt–chromium and All-Poly HXLPE/oxinium implants

Reviewers' conclusions

All-Poly HXLPE cup inserts performed in a similar manner to the Charnley implant but better than All-Poly cup inserts regardless of the femoral head bearing (cobalt–chromium or oxinium) in terms of head penetration. There was no significant difference in mean HHS across the five study groups

HXLPE, highly cross-linked PE; NA, not applicable; NR, not reported; PE, polyethylene; UHMWPE, ultra-high molecular weight PE.

Stem composition

Healy *et al.*[127]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details
Country: USA
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 1
Funding (government/private/manufacturer/other – specify): none received in support of this study
Aim of the study
To compare the effects of cobalt–chromium compared with titanium femoral stems in terms of post-THR clinical and radiographic measures
Participants
Recruitment dates: NR
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 390
Inclusion criteria: NR
Exclusion criteria: NR
Characteristics of participants (total study sample): mean (range) age (years): 65.2 (25–102); women, *n* (%): 188 (48.2); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 350 (89.7), osteonecrosis 16 (4.1), trauma 5 (1.3), RA 8 (2.0), other 11 (2.8)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): THR with cobalt–chromium femoral stem	
Intervention 2 (e.g. THR 2): THR with titanium femoral stem	
Bilateral procedure (yes/no/NR): yes	
Implant manufacturer: DePuy Orthopedics	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	✓
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): pain scale (VAS), HHS, SF-36 score, WOMAC score, revisions, femoral stem survival rate
Secondary outcomes (*list*): radiography outcomes (radiolucency, osteolysis, loosening, migration, subsidence and osteointegration)
Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): NR
Total length of follow-up: mean 4.7 (range 2.0–8.9) years

No. of patients

	Total	***THR with cobalt–chromium femoral stem***	***THR with titanium femoral stem***
Randomised	390	199	191
Analysed (if more than one follow-up point, choose and specify the last one)	358	NR	NR
Losses to follow-up/dropouts/ sample attrition (if more than one follow-up point, choose and specify the last one)	15	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR with cobalt–chromium femoral stem	Trilock femoral stem made of cobalt–chromium or titanium in 11 sizes is a straight, collarless, modular, cementless, porous-coated femoral implant with a flat tapered wedge design of the intraosseous body of the stem. Initial fixation was achieved by wedging the stem into the medial and lateral endosteal cortices of the proximal femur. Long-term fixation was achieved by bony ingrowth into the circumferential proximal porous coating. Modular femoral heads were made of cobalt–chromium with a highly polished bearing surface. Femoral heads were fixed to the femoral neck with a 12/14 Morse taper. In total, 387 hips used a 28-mm femoral head and 36 hips used a 32-mm femoral head. In 284 consecutive hips, a Duraloc acetabular cup with a conventional UHMWPE Duracon acetabular liner was used. In 139 subsequent consecutive hips, a Pinnacle cup with a highly cross-UHMWPE Marathon acetabular liner was implanted. Acetabular fixation was achieved with a press-fit technique with under-reaming by 1 mm. Acetabular screws were used at the surgeon's discretion	All operations were performed by orthopaedic surgeons who specialised in adult reconstruction
THR with titanium femoral stem	See above	See above

Patient baseline characteristics		
Characteristic	**THR with cobalt–chromium femoral stem**	**THR with titanium femoral stem**
Age (years), mean (range)	66 (25–100)	64 (35–102)
Sex, female, *n*/*N* (%)	96/199 (48.2)	92/191 (48.1)
Weight (kg), mean (range)	81.7 (44.9–136.0)	81.7 (36.3–149.6)
BMI (kg/m^2), mean (SD)	NR	NR
Primary OA, *n*/*N* (%)	182/199 (91.4)	168/191 (88.0)
Bilateral OA, *n*/*N* (%)	33/199 (16.6)	24/191 (12.6)
HHS, mean (range)	50.3 (16.2–70.7)	50.8 (23.4–75.4)
OHS, mean (range)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Post-procedural follow-up assessment timing (specify): 4.7 years

Outcome	*THR with cobalt–chromium femoral stem*	*THR with titanium femoral stem*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR (15 patients in both arms)	NR	NA
HHS, mean (range)	83 (34–100)	87 (55–100)	$p = 0.029$ (SS)
OHS, mean (range)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (range)	NR	NR	NA
HOOS, mean (range)	NR	NR	NA
LISOH score, mean (range)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (range)	NR	NR	NA
UCLA activity score, mean (range)	NR	NR	NA
WOMAC score, mean (range)	NR	NR	NA
AIMS score, mean (range)	NR	NR	NA
MACTAR score, mean (range)	NR	NR	NA
SF-36 score, mean (range)	NR	NR	NA
SF-12 score, mean (range)	NR	NR	NA
NHP score, mean (range)	NR	NR	NA
EQ-5D score, mean (range)	NR	NR	NA
Pain score (VAS), mean (range)	1.1 (0–10)	1.0 (0–10)	$p = 0.191$ (NS)
Pain score (other than VAS; specify), mean (range)	NR	NR	NA
Revision rate, *n/N* (%)	2/199 (1.0)	0/191 (0.0)	$p = 0.165$ (NS)
Time to revision (years), mean (range)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	98.9 (NR)	100 (NR)	$p = 0.169$ (NS)
Femoral head penetration (mm/year), mean (range)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR with cobalt–chromium femoral stem***	***THR with titanium femoral stem***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	NA	0/199	0/191	NR
Aseptic loosening (any or total)	NA	1/199 (0.5)	0/191 (0.0)	$p = 0.324$ (NS)
Infection	Peri	1/199 (0.5)	0/191 (0.0)	$p = 0.324$ (NS)
Femoral neck fracture	Peri	0/199 (0.0)	1/191 (0.5)	$p = 0.309$ (NS)
Metallosis	NA	NR	NR	NR
Muscle weakness	NA	NR	NR	NR
Nerve palsy	NA	NR	NR	NR
Deep-vein thrombosis	NA	NR	NR	NR
Implant dislocation	NR	3/199 (1.5)	0/191 (0.0)	$p = 0.678$ (NS)
Other (haematoma)	Peri	1/199 (0.5)	0/191 (0.0)	$p = 0.324$ (NS)

a RR or risk difference.

Authors' conclusions

Both post-treatment pain and HHS improved after baseline regardless of the treatment group, with a significantly better mean HHS for the titanium femoral stem group. At an average of 4.7 years of follow-up there was no significant difference between the groups in post-treatment pain, revision rate, stem survival and complications (dislocation, fracture, haematoma, infection)

Reviewers' conclusions

The allocation method was not truly randomised but was rather quasi-randomised. The results for complications and revisions may have been inconclusive rather than statistically non-significant because of the small counts

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; SS, statistically significant; UHMWPE, ultra-high molecular weight PE.

Stem design

Kim *et al.*[128]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Study details
Country: the Republic of Korea
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 1
Funding (government/private/manufacturer/other – specify): no benefits or funds received to support the study
Aim of the study
To compare a short metaphyseal-fitting femoral stem with a conventional metaphyseal- and diaphyseal-fitting stem with respect to postoperative clinical and radiographic parameters
Participants
Recruitment dates: October 2005–October 2007
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 100
Inclusion criteria: NR
Exclusion criteria: severe osteoporosis of the proximal femur
Characteristics of participants (total study sample): mean (range) age (years): 53 (21–77); women, *n* (%): 54 (54); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): OA 48 (48), osteonecrosis 40 (40), traumatic arthritis 8 (8), femoral neck fracture 4 (4)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): short anatomical metaphyseal-fitting cementless stem (Proxima)	
Intervention 2 (e.g. THR 2): conventional anatomical metaphyseal- and diaphyseal-fitting cementless stem (Profile)	
Bilateral procedure (yes/no/NR): yes	
Implant manufacturer: DePuy	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	✓
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	

Outcomes (study based)
Primary outcomes (*list*): HHS, pain VAS, Tegner and Lysholm activity score
Secondary outcomes (*list*): loosening, femoral component position, femoral offset, abductor moment arm, centre of rotation, femoral neck length, limb-length discrepancy, femoral component migration, radiolucency, bone mineral density
Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3 months, 1 year and yearly thereafter
Total length of follow-up: 3.35 years

Number of patients

	Total	***Short metaphyseal-fitting cementless stem (Proxima)***	***Conventional metaphyseal- and diaphyseal-fitting cementless stem (Profile)***
Randomised	100	50	50
Analysed (if more than one follow-up point, choose and specify the last one)	100	50	50
Losses to follow-up/dropouts/ sample attrition (if more than one follow-up point, choose and specify the last one)	0	0	0

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
Short metaphyseal-fitting cementless stem (Proxima)	A cementless Pinnacle acetabular component made of titanium alloy, Proxima stem and 28-mm-internal-diameter Biolox forte ceramic liner were used in all hips. The Proxima stem is manufactured using titanium alloy and is entirely porous coated with sintered titanium beads. The Proxima stem design has a medial metaphyseal area, which is proximally longer than a conventional stem. This implant has a highly pronounced lateral flare and allows full load transfer onto the proximal femur to be obtained	NR
Conventional metaphyseal- and diaphyseal-fitting cementless stem (Profile)	A cementless Pinnacle acetabular component made of titanium alloy, 28-mm-internal-diameter Biolox forte ceramic liner and Profile femoral component were used in all hips. The cementless Profile femoral component is made of titanium alloy and is an anatomical metaphyseal- and diaphyseal-fitting stem. The proximal metaphyseal portion of the stem (approximately one-third of the stem) is porous coated with sintered beads. The pore size was 250 µm. A 28-mm-diameter Biolox forte ceramic femoral head was used in all hips. All operations were performed using a posterolateral approach	NR

Patient baseline characteristics		
Characteristic	***Short metaphyseal-fitting cementless stem (Proxima)***	***Conventional metaphyseal- and diaphyseal-fitting cementless stem (Profile)***
Age (years), mean (SD)	54.3 (12.97)	51.8 (12.3)
Sex, female, *n/N* (%)	28/50 (56)	26/50 (52)
Weight (kg), mean (SD)	66.5 (9.51)	64.8 (10.6)
BMI (kg/m^2), mean (SD)	25.6 (2.82)	24.7 (3.6)
Primary OA, *n/N* (%)	24/50 (48)	24/50 (48)
Bilateral OA, *n/N* (%)	5/50 (10)	5/50 (10)
HHS, mean (SD)	44.6 (NR)	48.4 (NR)
OHS, mean (SD)	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 3 years

Outcome	*Short metaphyseal-fitting cementless stem (Proxima)*	*Conventional metaphyseal- and diaphyseal-fitting cementless stem (Profile)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	0	0	NR
HHS, mean (SD)	97.0 (NR)	96.0 (NR)	$p = 0.79$ (NS)
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	0	0	NR
Time to revision (years), mean (SD)	NA	NA	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***Short metaphyseal-fitting cementless stem (Proxima)***	***Conventional metaphyseal- and diaphyseal-fitting cementless stem (Profile)***	***Between-group difference and p-value (or 95% CI)[a]***
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	NA	NR	NR	NA
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	NA	NR	NR	NA
Other (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions
In both study groups the 3-year postoperative mean HHS improved; however, there were no significant differences in mean HHS and radiographic outcomes between the two treatment groups at follow-up
Reviewers' conclusions
Although the authors report no significant differences in clinical outcomes between the study groups, it is difficult to confirm their findings as no adequate statistical measures are reported. It is not clear what were the pain and function scores reported in the results section

NA, not applicable; NR, not reported; PE, polyethylene.

Stem fixation

Kim *et al.*[129]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details
Country: the Republic of Korea
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): Specialty clinic
Number of centres: NR
Funding (government/private/manufacturer/other – specify): NR – no benefits in any form were received or will be received from a commercial party related directly or indirectly to the subject of this article
Aim of the study
To compare the clinical and radiological results, rates of revision and survival of implants after THR with cemented (hybrid) compared with cementless femoral components performed in patients aged < 50 years, at a minimum of 16 years' follow-up
Participants
Recruitment dates: January 1991 and February 1993
Total number of patients screened for inclusion eligibility: NR
Total number of patients randomised: 166 patients
Inclusion criteria: NR
Exclusion criteria: NR
Characteristics of participants (total study sample): mean (range) age (years): approximately 45 (21–50); women, *n* (%): 37 (24); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): osteonecrosis 104 (62), OA 22 (14), childhood pyogenic arthritis 18 (11), ankylosing spondylitis 5 (3), multiple epiphyseal dysplasia 4 (2.5), developmental dysplasia 3 (2), RA 1 (1)

Intervention (keep the same order as in the paper)	
Intervention 1 (e.g. THR 1): THR with cemented femoral stem	
Intervention 2 (e.g. THR 2): THR with cementless femoral stem	
Bilateral procedure (yes/no/NR): NR	
Implant manufacturer: the Charnley Elite or Elite-plus stem (Ortron 90; DePuy) was used in the cemented (hybrid) group and the Profile stem (DePuy) in the cementless group. A cementless Duraloc 100 or 1200 series acetabular component (DePuy) was used in all hips in both groups	
Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR	
If several types of THR are compared, indicate the basis for comparison (tick all that apply):	
Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)	
Cup fixation (e.g. cemented, cementless, other)	
Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)	
Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)	
Cup size (mm)	
Femoral head type (e.g. modular, custom, other)	
Femoral head composition (e.g. ceramic, metal, other)	
Femoral head size (mm)	
Stem type (e.g. monolithic, modular/tapered, other)	
Stem composition (e.g. titanium, cobalt–chromium and stainless steel)	
Stem fixation (e.g. cemented, cementless, other)	✓

Outcomes (study based)
Primary outcomes (*list*): HHS, WOMAC score, pain (VAS), ranges of movement, UCLA activity score, revision, survival rate
Secondary outcomes (*list*): radiolucency, stability of the cementless femoral component, loosening of the cemented femoral component, PE liner wear rate, osteolysis
Imaging method used (i.e. conventional radiography, radiostereometry, none): Radiographs of all hips using the isthmus ratio of Dorr
Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3 months and 1 year after the operation and yearly thereafter (minimum of 16 years' follow-up)
Total length of follow-up: 20 years

Number of patients			
	Total	***THR with cemented femoral stem***	***THR with cementless femoral stem***
Randomised	166	83	83
Analysed (if more than one follow-up point, choose and specify the last one)	157	78	79
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	9	5	4

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR with cemented femoral stem	The Charnley Elite or Elite-plus stem was used in the cemented (hybrid) group and the Profile stem in the cementless group. When cement was used it was applied using an intramedullary plug, pulsatile lavage, vacuum mixing, injection with a gun, a proximal rubber seal and a distal centraliser on the femoral component	NR
THR with cementless femoral stem	A cementless Duraloc 100 or 1200 series acetabular component was used in all hips in both groups. Of the 62 Duraloc 1200 acetabular components used in both groups, 28 were fixed with one or two screws and the remaining 34 were press-fitted without using an additional screw. The cementless femoral components were inserted with a press-fit as determined by the preoperative use of templates. At the time of the operation, an attempt was made to fill the femoral canal with the broach, leaving little cancellous bone remaining	NR

Patient baseline characteristics		
Characteristic	***THR with cemented femoral stem***	***THR with cementless femoral stem***
Age (years), mean (range)	43.4 (21–50)	46.8 (21–49)
Sex, female, *n/N* (%)	16/78 (21)	21/79 (27)
Weight (kg), mean (range)	59 (45–82)	60.5 (48–87)
BMI (kg/m^2), mean (range)	22.2 (22.1–24.8)	22.2 (21.9–24.4)
Primary OA, *n/N* (%)	10/78 (12.8)	12/79 (15.2)
Bilateral OA, *n/N* (%)	NR	NR
HHS, mean (range)	44 (5–66)	48.8 (6–55)
OHS, mean (range)	NR	NR

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): Unclear (mean follow-up for both groups 18.4 years)

Outcome	*THR with cemented femoral stem*	*THR with cementless femoral stem*	*Between-group difference and p-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (range)	91 (75–100)	90 (71–100)	$p = 0.71$ (NS)
OHS, mean (range)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (range)	NR	NR	NA
HOOS, mean (range)	NR	NR	NA
LISOH score, mean (range)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (range)	NR	NR	NA
UCLA activity score, mean (range)	7.6 (6–10)	7.8 (5–10)	$p = 0.814$ (NS)
WOMAC score, mean (range)	11 (4–35)	13 (4–33)	$p = 0.927$ (NS)
AIMS score, mean (range)	NR	NR	NA
MACTAR score, mean (range)	NR	NR	NA
SF-36 score, mean (range)	NR	NR	NA
SF-12 score, mean (range)	NR	NR	NA
NHP score, mean (range)	NR	NR	NA
EQ-5D score, mean (range)	NR	NR	NA
Pain score (VAS), mean (range)	NR	NR	
Pain score (other than VAS; specify), mean (range)	NR	NR	NA
Revision rate, *n/N* (%)	Acetabular component: 14/109 (13)	Acetabular component: 18/110 (16)	$p = 0.673$ (NS)
	Femoral component: 3/109 (3)	Femoral component: 4/110 (4)	$p = 0.912$ (NS)
Time to revision (years), mean (range)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	Acetabular component: 87.0 (80.0 to 93.0)	Acetabular component : 84.0 (78.0 to 92.0)	$p = 0.776$ (NS)
	Femoral component: 97.0 (91.0 to 100.0)	Femoral component: 96.0 (93.0 to 100.0)	$p = 0.794$ (NS)
Femoral head penetration (mm/year), mean (range)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (timing unclear – mean follow-up for both groups 18.4 years)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR with cemented femoral stem***	***THR with cementless femoral stem***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	Post	Acetabulum: 35/109 (32)	Acetabulum: 40/110 (36)	$p = 0.168$ (NS)
		Femur: 31/109 (28)	Femur: 35/110 (32)	$p = 0.159$ (NS)
Aseptic loosening (any or total)	Post	NR; revision for aseptic loosening was required for two femoral components at 9 and 16 years, respectively	NR; revision for aseptic loosening was carried out for three femoral components at 14, 16 and 18 years, respectively	NR
Infection	Post	NR	NR	NR
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	Post	NR	NR; one revision at 1 year for recurrent dislocation	NR
Other (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions

In patients aged < 50 years cemented and cementless femoral components appear to provide outstanding long-term fixation and significant pain relief well into the second decade, but wear and periacetabular osteolysis constitute the major challenges in hybrid and cementless THR in these young patients. Similar functional scores and radiological results across cemented and cementless THR recipient patients were observed

Reviewers' conclusions

Study investigated the incidence of osteolysis and the survival of hybrid and cementless THRs in patients aged < 50 years. Similar functional scores and radiological results across cemented and cementless THR recipient patients were observed. Inclusion criteria are unclear making it difficult to generalise the results to patients currently receiving treatment

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene.

Included systematic reviews: studies comparing different types of total hip replacement ($n = 5$)

Cup fixation

Voigt and Mosier[137]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: USA

Funding: NR

Aim of the study

To compare uncemented metal-backed acetabular components with PE inserts with cemented all-PE acetabular components (using the same type of femoral component and method of femoral fixation in both arms of the trial) in terms of revision rates, function, complications and costs in patients with OA

Methods

Search strategy

Databases searched: PubMed-MEDLINE, The Cochrane Library

Last date of search: 13 June 2011

Other methods of identifying literature: websites (technology assessment, meetings), hand search of journals and article references

Inclusion criteria

Participants: patients with OA or RA

Interventions: primary total hip implant with uncemented metal-backed acetabular components with PE inserts

Comparators: primary total hip implant with cemented all-PE acetabular components

Outcome measures: revision rate, function (HHS, OHS), complications (infection or wound, deep-vein thrombosis, pulmonary embolism, dislocations, over-reaming, fractures and costs of treatment)

Types of studies included (i.e. study design): RCTs

Study quality assessment methods

Quality assessment tool used: Cochrane risk of bias tool

Risk of bias assessment criteria applied to included studies: sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other issues

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis)	☐
(b) Quantitative (meta-analysis)	☒

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used)	☐
(b) Adjusted (anchored indirect treatment comparison)	☐
(c) Mixed treatment comparison	☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression)	☒
(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests)	☒
(c) Overall quality/strength of evidence (GRADE):	☐

Results

Number of included studies: 6

Overall risk of bias of included studies: random sequence generation: low 4/6 trials; allocation concealment: high 3/6 trials, low 2/6 trials; caregiver blinding: high 6/6 trials; patient blinding: unknown 4/6 trials, high 1/6 trials; assessor blinding: high 5/6 trials; incomplete outcome data: low 5/6 trials; selective reporting: low 6/6 trials; other bias: low 3/6 trials, high 2/6 trials

Treatment effect per outcome (narrative statement and numerical data for pooled estimate, if available): revision rate at 4–8 years: no significant difference (pooled RR 0.15, 95% CI 0.02 to 1.18); revision rate at 10 years: no significant difference (pooled RR 1.36, 95% CI 0.81 to 1.29); function according to HHS and OHS: no significant difference; total complications: no significant difference (pooled RR 0.61, 95% CI 0.32 to 1.18); costs of treatment: no studies reported this outcome

Strength of evidence per outcome (GRADE): NA

Conclusions: there is no clear evidence favouring the use of either an all-PE or an uncemented acetabular component. Although both appear to have similar outcomes up to 10 years, there was a non-significant trend towards late failure in the all-PE group compared with other forms of acetabular implant fixation

Reviewers' conclusions and additional comments/concerns

No apparent differences were seen between the performance of all-PE acetabular implant fixation and the performance of other forms of acetabular implant fixation. As the evidence was limited, more studies are needed to replicate or refute the findings to reach more definitive conclusions

NA, not applicable; NR, not reported; PE, polyethylene.

Pakvis *et al.*[138]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details

Country: the Netherlands

Funding: NR

Aim of the study

To identify all relevant RCTs and comparative cohort studies in which cemented and cementless sockets were compared

Methods

Search strategy

Databases searched: MEDLINE and EMBASE (1980–December 2009)

Last date of search: December 2009

Other methods of identifying literature: References of the retrieved articles were examined for additional relevant publications

Inclusion criteria

Participants: indication for performing THR had to be primary or secondary OA

Interventions: cemented acetabular components

Comparators: cementless acetabular components

Outcome measures: minimal follow-up had to be 12 months; data presented had to be clinical (complications, HHS and survival) and radiological (wear, migration and osteolysis) outcome measurements

Types of studies included (i.e. study design): non-randomised studies, RCTs

Study quality assessment methods

Quality assessment tool used: van Tulder checklist (for RCTs), Newcastle–Ottawa quality assessment scale (for non-RCTs)

Risk of bias assessment criteria applied to included studies: van Tulder checklist: randomisation, allocation concealment, prognostic factors, patient blinding, surgeon blinding, outcome assessor blinding, co-interventions, compliance, dropout, timing of the outcome assessments, intention to treat and homogeneity; Newcastle–Ottawa quality assessment scale: representativeness of the exposed cohort, selection of the non-exposed cohort, ascertainment of exposure, demonstration that outcome of interest was not present at start of study, comparability of cohorts on the basis of the design or analysis, assessment of outcome and adequacy of follow-up of cohorts

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis) ☒

(b) Quantitative (meta-analysis) ☐

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used) ☐

(b) Adjusted (anchored indirect treatment comparison) ☐

(c) Mixed treatment comparison ☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression) ☐

(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests) ☐

(c) Overall quality/strength of evidence (GRADE): ☐

Results

Number of included studies: 35

Overall risk of bias of included studies: three RCTs scored 'yes' on > 50% of the van Tulder criteria. In orthopaedic surgery, surgeon blinding is not feasible. Therefore, when re-evaluating the results of the van Tulder questionnaire we could select seven articles that scored 'yes' on > 50% of the van Tulder items. Twelve non-randomised studies scored > 67% on the Newcastle–Ottawa quality assessment scale

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available): no overall effect reported. The best analysis of the evidence for complications, wear, osteolysis, migration and clinical scores showed no superiority of either cemented or cementless sockets in the RCTs

Strength of evidence per outcome (GRADE): NR

Conclusions: recommend that an orthopaedic surgeon should choose an established cemented or cementless socket for hip replacement based on patient characteristics, knowledge, experience and preference. Complications are not well described in the RCTs and provided no conclusive answers. The RCTs provided only short- to medium-term follow-up. The non-RCTs provided various long-term comparisons with ambiguous results in terms of determining the superior fixation method

Reviewers' conclusions and additional comments/concerns

The aim of this paper was to undertake a systematic literature review to evaluate studies that compared cemented and cementless sockets to find evidence for the superior method of acetabular fixation. In total, 16 RCTs and 19 non-RCTs were identified in which cemented and cementless acetabular components were compared. No statistically significant differences were found for osteolysis, migration and cup survival. The cemented socket was superior for long-term revision and PE wear. Although the authors performed a form of quality assessment of the included papers, little consideration was given to the quality of the papers when reporting the overall findings. The review is of moderate methodological quality

NR, not reported; PE, polyethylene.

Clement *et al.*[139]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details

Country: France

Funding: NR

Aim of the study

To perform a critical analysis of the current evidence from a systemic literature review of comparative studies, long-term case series, previous literature reviews, meta-analyses and national arthroplasty registry data on cemented and uncemented acetabular components to determine the respective survivorship rates, overall risk of reoperation, dislocation rates and rates of wear-related complications

Methods

Search strategy

Databases searched: MEDLINE

Last date of search: 2011

Other methods of identifying literature: searched the national register annual reports

Inclusion criteria

Participants: young patients and patients with dysplastic hip disease

Interventions: cemented THR

Comparators: uncemented THR

Outcome measures: aseptic loosening, radiographic loosening, overall survival, wear rates per year, dislocation, osteolysis, acetabular revision, liner exchange, quality of life

Types of studies included (i.e. study design): (1) all published review articles and meta-analyses; (2) all studies comparing cemented with uncemented acetabular components with a minimal follow-up of 5 years, quoting survival and wear or complications; (3) all arthroplasty registers reporting a comparison between cemented and uncemented acetabular components; (4) single-centre outcome studies with > 13 years' follow-up; (5) single-centre series reporting survivorship of cemented or uncemented cups in young patients and patients with dysplastic hip disease

Study quality assessment methods

Quality assessment tool used: none

Risk of bias assessment criteria applied to included studies: none

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis)	☒
(b) Quantitative (meta-analysis)	☐

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used)	☐
(b) Adjusted (anchored indirect treatment comparison)	☐
(c) Mixed treatment comparison	☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression)	☐
(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests)	☐
(c) Overall quality/strength of evidence (GRADE):	☐

Results

Number of included studies: 55 studies (four reviews, one meta-analysis, 16 comparative studies, eight reports from arthroplasty registers, 16 long-term case series, five studies reporting young hips and five studies reporting the outcome of dysplastic hip); 11 comparative studies of various cemented all-PE and uncemented press-fit sockets with (non-cross-linked) PE liners

Overall risk of bias of included studies: NA

Treatment effect per outcome (narrative statement and numerical data for pooled estimate, if available): pooled dislocation rates (cemented cup vs. cementless cup): 1.3% (12/914) vs. 4.1% (28/696) ($p = 0.001$) (nine comparative studies)

Strength of evidence per outcome (GRADE): NA

Conclusions: (1) lower or equal risk of failure from aseptic loosening as end point for uncemented sockets; (2) higher risk for dislocation and revision for dislocation with uncemented sockets; (3) higher risk of osteolysis with uncemented sockets; (4) increased overall risk of revision and reoperation with uncemented sockets (and conventional PE); (5) no evidence to support superior uncemented fixation in young patients; (6) potential future improvement in the long-term survival of uncemented socket with enhanced bearing surface technology; (7) improved survival (aseptic and hence overall) of cemented sockets implanted with modern cementing techniques; 8) cemented socket fixation remains the gold standard in all age groups until other methods may/will prove to be superior

Reviewers' conclusions and additional comments/concerns

The paper aimed to undertake a systemic literature review of comparative studies, long-term case series, previous literature reviews, meta-analyses and national arthroplasty registry data on cemented and uncemented acetabular components to determine the respective survivorship rates, overall risk of reoperation, dislocation rates and rates of wear-related complications. A limitation of the review is that no long-term series with cross-linked PE liners could be identified. The review highlighted the difficulty of analysing the evidence because of missing data across studies. The review refers to concerns that author-reported all-cause survival rates often do not reflect the true reoperation/revision rates when subject to more critical analysis. The overall survival of cemented fixation of the socket appears superior to that of the uncemented alternatives. A cemented cup remains the optimal method of fixation for older patients because of the predictable outcome and lower cost. No quality assessment or assessment of risk of bias was carried out. The review attempts to evaluate a broad range of study designs. A limited search of electronic databases was carried out

NA, not applicable; NR, not reported; PE, polyethylene.

Femoral head bearing-on-cup liner bearing

Sedrakyan *et al.*[140]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: USA

Funding: US Food and Drug Administration Center for Devices and Radiological Health

Aim of the study

To determine the comparative safety and effectiveness of combinations of bearing surfaces of hip implants

Methods

Search strategy

Databases searched: MEDLINE, EMBASE and The Cochrane Central Register of Controlled Trials from January 1995

Last date of search: June 2011

Other methods of identifying literature: reference lists of trials and reviews for additional studies, online annual reports of all registries that report information from the registry

Inclusion criteria

Participants: adults, reporting any one of the clinical outcomes of interest (any functional outcomes or revisions or both)

Interventions: conventional hip replacement

Comparators: conventional hip replacement

Outcome measures: any functional outcome (HHS and general quality of life measures such as the SF-12) and occurrence of revision

Types of studies included (i.e. study design): RCTs, controlled clinical trials, observational comparative controlled studies

Study quality assessment methods

Quality assessment tool used: selected validity items (RCTs) and STROBE (observational studies)

Risk of bias assessment criteria applied to included studies: RCTs: methods of random allocation generation, allocation concealment, masking of patients and outcome assessors, intention-to-treat analysis

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis)	☐
(b) Quantitative (meta-analysis)	☒

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used)	☐
(b) Adjusted (anchored indirect treatment comparison)	☐
(c) Mixed treatment comparison	☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression)	☒
(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests)	☒
(c) Overall quality/strength of evidence (GRADE)	☐

Results

Number of included studies: 18

Overall risk of bias of included studies: four studies were classified as being of moderate to high quality, five studies were classified as being of moderate quality and six studies were classified as being of low quality

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available):

HHS (metal-on-metal vs. metal-on-PE): at 2 years (four studies) – pooled MD –2.40, 95% CI –4.47 to –0.33 (in favour of metal-on-PE); beyond 2 years (two studies) – pooled MD 1.21, 95% CI –2.41 to 4.83 (neither favoured)

HHS (ceramic-on-ceramic vs. ceramic-on-PE): no significant difference (five studies)

HHS (ceramic-on-PE vs. metal-on-PE): no significant difference (two studies)

HHS (metal-on-metal vs. ceramic-on-ceramic): no significant difference (one study)

Quality of life measures such as SF-12 (metal-on-metal vs. metal-on-PE): better physical functioning in metal-on-PE group

Quality of life measures such as SF-12 (ceramic-on-ceramic vs. ceramic-on-PE): no significant difference (one study)

Revisions (metal-on-metal vs. metal-on-PE): no significant difference (two studies)

Revisions (ceramic-on-ceramic vs. metal-on-PE): lower occurrence in ceramic-on-ceramic group (one study)

Revisions (ceramic-on-ceramic vs. ceramic-on-PE): no significant difference (five studies)

Revisions (ceramic-on-PE vs. metal-on-PE): no significant difference (one study)

Dislocations (metal-on-metal vs. metal-on-PE): no significant difference (three studies)

Strength of evidence per outcome (GRADE): NA

Conclusions: there is limited evidence regarding the comparative effectiveness of various hip implant bearings and the results do not indicate any advantage for metal-on-metal or ceramic-on-ceramic implants compared with traditional bearings

Reviewers' conclusion and additional comments/concerns

Limited evidence shows that there is no difference in either functional scores or revision rates between different combinations of implant bearings

NA, not applicable; PE, polyethylene; STROBE, STrengthening the Reporting of OBservational studies in Epidemiology.

Yoshitomi *et al.*[141]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: Japan

Funding: Health and Labour Sciences Research Grants

Aim of the study

To compare the survivorship/revision and annual PE wear rates between THR with zirconia-on-PE and THR with non-zirconia-on-PE and to explore if manufacturers or fixation method influenced survivorship

Methods

Search strategy

Databases searched: PubMed (1966–July 2007), EMBASE (1974–July 2007) and the Cochrane Central Register of Controlled Trials (Issue 4, July 2007)

Last date of search: July 2007

Other methods of identifying literature: The Japana Centra Revuo Medicina was also searched for articles written in Japanese (1983–July 2007)

Inclusion criteria

Participants: NR

Interventions: THR using zirconia heads with PE cup liners (regardless of femoral head size, method of fixation)

Comparators: THR using non-zirconia heads with PE cup liners (regardless of femoral head size, method of fixation)

Outcome measures: survivorship/revision, PE wear rates

Types of studies included (i.e. study design): RCTs, non-RCTs and cohort studies with follow-up of > 5 years

Study quality assessment methods

Quality assessment tool used: the Cochrane Back Review Group 11-item criteria

Risk of bias assessment criteria applied to included studies: generation of random allocation, allocation concealment, blinding, co-interventions, compliance, sample attrition, outcome assessment timing and type of analysis

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis)	☐
(b) Quantitative (meta-analysis)	☒

Indirect comparison(tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used)	☐
(b) Adjusted (anchored indirect treatment comparison)	☐
(c) Mixed treatment comparison	☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression)	☒
(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests)	☒
(c) Overall quality/strength of evidence (GRADE)	☐

Results

Number of included studies: seven (three RCTs and four cohort studies); non-zirconia heads were made of alumina ceramic, cobalt–chromium or stainless steel

Overall risk of bias of included studies: The mean (range) score for the cohort studies was 4.5 (4–5) and that for the RCTs was 6.3 (6–7)

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available):

Revision rate: at 89 months, THRs with zirconia heads had more revisions than THRs with non-zirconia heads (seven studies; pooled RD 0.05, 95% CI 0.02 to 0.08). For implants made by Ceraver, THRs with zirconia heads had more revisions than THRs with non-zirconia heads (two studies; pooled RD 0.08, 95% CI 0.03 to 0.14). For implants made by DePuy, THRs with zirconia heads had a similar risk of revisions to THRs with non-zirconia heads (three studies; pooled RD 0.02, 95% CI –0.01 to 0.06). For cemented implants, THRs with zirconia heads had more revisions than THRs with non-zirconia heads (five studies; pooled RD 0.06, 95% CI 0.02 to 0.10)

PE wear rate in mm (annual): THRs with zirconia heads had a similar PE wear rate to THRs with non-zirconia heads (five studies; pooled RD 0.026, 95% CI –0.06 to 0.11)

Strength of evidence per outcome (GRADE): NA

Conclusions: THRs with DePuy zirconia heads demonstrated similar prosthesis revision and PE wear rates to THRs with non-zirconia heads, but THRs with DePuy zirconia heads had a higher incidence of revision than THRs with non-zirconia heads; the method of implant fixation did not modify the observed main difference between the two types of THR

Reviewers' conclusion and additional comments/concerns

There was heterogeneity in the effect of THR (zirconia heads) vs. THR (non-zirconia heads) on the revision rate as demonstrated by the manufacturer of the femoral heads; the method of implant fixation did not modify the effect in terms of revision rates. THRs with zirconia heads and THRs with non-zirconia heads demonstrated similar PE wear rates

NA, not applicable; PE, polyethylene; RD, risk difference.

Included randomised controlled trials: total hip replacement compared with resurfacing arthroplasty (*n* = 3)

Costa et al.[130] *and Achten* et al.[107]

Name of first reviewer: Paul Sutcliffe

Name of second reviewer: Alexander Tsertsvadze

Study details

Country: UK

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): speciality clinic

Number of centres: 1

Funding (government/private/manufacturer/other – specify): This study was funded by the Research for Patient Benefit scheme of the National Institute for Health Research; the University of Warwick; and University Hospitals Coventry and Warwickshire NHS Trust

Aim of the study

To compare the clinical effectiveness of THR with that of RS in patients with severe arthritis of the hip with regard to hip function, quality of life, physical activity and harms

Participants

Recruitment dates: May 2007 and February 2010

Total number of patients screened for inclusion eligibility: 175

Total number of patients randomised: 126

Inclusion criteria: age > 18 years, medically fit for an operation and suitable for RS

Exclusion criteria: evidence indicating that the patient would be unable to adhere to trial procedures or complete questionnaires. If a recruited patient needed a contralateral hip replacement during the trial period, the second hip was not included in the study

Characteristics of participants (total study sample): mean age (years): 56; women, *n* (%): 52 (41); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): primary OA 120 (95); NR 6 (5)

Intervention (keep the same order as in the paper)

Intervention 1 (e.g. THR 1): RS

Intervention 2 (e.g. THR 2): THR

Bilateral procedure (yes/no/NR): no

Implant manufacturer: NR

Postprocedural rehabilitation (e.g. weight-bearing, exercise): Standardised rehabilitation plans for early exercises, precautions to be followed for the first 3 months, functional activity and later-stage exercises

If several types of THR are compared, indicate the basis for comparison (tick all that apply): NA

Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)

Cup fixation (e.g. cemented, cementless, other)

Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)

Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)

Cup size (mm)

Femoral head type (e.g. modular, custom, other)

Femoral head composition (e.g. ceramic, metal, other)

Femoral head size (mm)

Stem type (e.g. monolithic, modular/tapered, other)

Stem composition (e.g. titanium, cobalt–chromium and stainless steel)

Stem fixation (e.g. cemented, cementless, other)

Outcomes (study based)

Primary outcomes (*list*): hip function as assessed by the OHS and HHS

Secondary outcomes (*list*): EQ-5D score, Disability Rating Index, Paffenbarger Physical Activity Questionnaire, complications

Imaging method used (i.e. conventional radiography, radiostereometry, none): NR

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3 weeks and 3, 6 and 12 months

Total length of follow-up: 12 months

Number of patients

	Total	***RS***	***THR***
Randomised	126	60	66
Analysed (if more than one follow-up point, choose and specify the last one)	126	60	66
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	6	3	3

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
RS	The articular surfaces of the femoral head were removed but the neck was left in situ. The femoral component (cap) was then impacted onto the patient's own femoral neck. All used metal-on-metal bearing surfaces but the choice of surgical approach, implant size and positioning was left to the discretion of the operating surgeon. In both forms of arthroplasty, the acetabulum was prepared and the acetabular component inserted into the socket	A total of 20 surgeons performed the 122 arthroplasties in this trial. An orthopaedic consultant was recorded as the lead operating surgeon in 88 (76%) operations, with the remaining 28 (24%) being led by senior orthopaedic trainees. The proportion of consultant vs. trainee surgeons was similar in the two treatment groups
THR	The femoral head was removed along with most of the femoral neck. The femoral shaft was exposed to open up the femoral canal. The femoral component was then inserted into the canal and the articulating femoral head was placed onto the neck of the femoral component. The choice of components (cemented vs. uncemented) and bearing surfaces was left to the discretion of the operating surgeon	See above

Patient baseline characteristics		
Characteristic	***RS***	***THR***
Age (years), mean (SD)	56.3 (7.3)	56.6 (6.6)
Sex, female, *n/N* (%)	22/60 (37.0)	30/66 (45.0)
Weight (kg), mean (SD)	NR	NR
BMI (kg/m^2), mean (SD)	28.6 (6.3)	28.7 (4.6)
Primary OA, *n/N* (%)	59/60 (98)	61/66 (93)
Bilateral OA, *n/N* (%)	NR	NR
HHS, mean (SD)	48.6 (14.2)	50.1 (13.5)
OHS, mean (SD)	19.1 (8.0)	19.6 (7.8)

Efficacy outcomes
For each timing of assessment please provide a separate table
For scores, extract only total scores

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 12 months

Outcome	*RS*	*THR*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (95% CI)	88.4 (84.4 to 92.4)	82.3 (77.2 to 87.5)	6.04 (–0.51 to 12.58)
OHS, mean (95% CI)	40.4 (37.9 to 42.9)	38.2 (35.3 to 41.0)	2.23 (–1.52 to 5.98)
Merle d'Aubigné and Postel score (specify if modified or original)	NR	NR	NA
Mean (SD), mean (95% CI)			
HOOS, mean (95% CI)	NR	NR	NA
LISOH score, mean (95% CI)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (95% CI)	NR	NR	NA
UCLA activity score, mean (95% CI)	NR	NR	NA
WOMAC score, mean (95% CI)	NR	NR	NA
AIMS score, mean (95% CI)	NR	NR	NA
MACTAR score, mean (95% CI)	NR	NR	NA
SF-36 score, mean (95% CI)	NR	NR	NA
SF-12 score, mean (95% CI)	NR	NR	NA
NHP score, mean (95% CI)	NR	NR	NA
EQ-5D score, mean (95% CI)	0.796 (0.721 to 0.870)	0.719 (0.636 to 0.802)	0.077 (–0.034 to 0.188)
Pain score (VAS), mean (95% CI)	NR	NR	NA
Pain score (other than VAS; specify), mean (95% CI)	NR	NR	NA
Revision rate, *n*/*N* (%)	NR	NR	NA
Time to revision (years), mean (95% CI)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (95% CI)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***RS***	***THR***	***Between-group difference and* p-*value (or 95% CI)*[a]
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	NR	0/60 (0.0)	2/66 (3.0)	$p = 0.497$ (NS)
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NR	4/60 (6.7)	0/66 (0.0)	$p = 0.049$ (NS)
Implant dislocation	NR	1/60 (1.7)	1/66 (1.5)	$p = 1.000$ (NS)
Other (superficial wound complication)	NR	2/60 (3.3)	9/66 (13.6)	$p = 0.057$ (NS)

a RR or risk difference.

Authors' conclusions

No evidence of treatment effect on hip function; inconclusive results because of wide CIs. The long-term effects of these interventions remain uncertain

Reviewers' conclusions

The study aimed to compare the clinical effectiveness and cost-effectiveness of THR with that of RS in patients with severe arthritis of the hip. THR involved replacement of entire femoral head and neck; hip RS involved replacement of the articular surface of the femoral head only, with the femoral neck remaining intact). In both treatment groups there was a significant improvement in quality of life, HHS and OHS. For the between-group differences the results were inconclusive because the trial was underpowered and because of the wide CIs. The overall quality of the study was good and the conclusions appear appropriate. The following concerns were noted: broad eligibility criteria; the pragmatic nature of the trial with a relatively large number of surgeons using different hip implants and their own preferred surgical technique could bring its own potential bias to the findings; single-centre recruitment; and patients were not blinded. The main conclusion is that there was no evidence that RS provides improved hip function or increased activity levels compared with THR

NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene.

Garbuz et al.[131]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Study details

Country: Canada

Study design: RCT

Study setting (primary care/specialty clinic/other – specify): specialty clinic, community hospital

No. of centres: 3

Funding (government/private/manufacturer/other – specify): the institution of one or more of the authors (DSG, MT, NVG, BAM, CPD) has received funding from Zimmer, Inc. (Warsaw, IN, USA)

Aim of the study

To compare THR (metal-on-metal with large diameter head) with RS (metal-on-metal) in terms of quality of life measures and metal ion levels in serum

Participants

Recruitment dates: June 2005 and August 2008

Total number of patients screened for inclusion eligibility: NR

Total number of patients randomised: 104

Inclusion criteria: patients aged between 19 and 70 years deemed suitable for RS as judged by the treating surgeon

Exclusion criteria: previous fracture of the hip requiring internal fixation, previous femoral or pelvic osteotomy, dysplasia requiring structural graft, presence of osteopenia or osteoporosis, and hepatic or renal insufficiency

Characteristics of participants (total study sample): mean (range or SD) age (years): 52 (NR); women, *n* (%): 11 (10.5); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): NR

Intervention (keep the same order as in the paper)

Intervention 1 (e.g. THR 1): THR (metal-on-metal with large head)

Intervention 2 (e.g. RS): RS (metal-on-metal)

Bilateral procedure (yes/no/NR): NR

Implant manufacturer: Zimmer, Inc.

Postprocedural rehabilitation (e.g. weight-bearing, exercise): NR

If several types of THR are compared, indicate the basis for comparison (tick all that apply): NA

Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)

Cup fixation (e.g. cemented, cementless, other)

Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)

Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)

Cup size (mm)

Femoral head type (e.g. modular, custom, other)

Femoral head composition (e.g. ceramic, metal, other)

Femoral head size (mm)

Stem type (e.g. monolithic, modular/tapered, other)

Stem composition (e.g. titanium, cobalt–chromium and stainless steel)

Stem fixation (e.g. cemented, cementless, other)

Outcomes (study based)

Primary outcomes (*list*): clinical/quality of life scores [Paper Adaptive Test (PAT)-5D index, WOMAC, SF-36 and UCLA activity scores]

Secondary outcomes (*list*): serum cobalt and chromium levels

Imaging method used (i.e. conventional radiography, radiostereometry, none): none

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 8 and 12 weeks and 1 year

Total length of follow-up: 2 years

No. of patients

Number	***Total***	***THR (metal-on-metal with large head)***	***RS (metal-on-metal)***
Randomised	104	56	48
Analysed (if more than one follow-up point, choose and specify the last one)	96	NR	NR
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	8	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR (metal-on-metal with large head)	All surgeries were performed through the posterior approach. The implants were from one manufacturer (Zimmer, Inc.). The acetabular component in the two groups was identical (Durom cup). In the RS group the femoral component was the Durom femoral resurfacing component and in the THR group the femoral component was the M/L Taper stem made of titanium. Onto this was placed a large Metasul head via a cobalt–chromium alloy metal sleeve adapter and Morse taper in order to match the 12/14 taper of the stem. The bearing surface in each arm was identical. The femoral and acetabular components are made of wrought-forged, high-carbon content cobalt–chromium alloy (0.20–0.25% carbon). The surface roughness was < 0.005 µm and the radial clearance was 75 µm	NR
RS (metal-on-metal)	See above	NR
Patient baseline characteristics		
Characteristic	***THR (metal-on-metal with large head)***	***RS (metal-on-metal)***
Age (years), mean (SD)	52.0 (NR)	51.5 (NR)
Sex, female, *n*/*N* (%)	6/56 (10.7)	5/48 (10.4)
Weight (kg), mean (SD)	NR	NR
BMI (kg/m^2), mean (SD)	28.2 (NR)	28.3 (NR)
Primary OA, *n*/*N* (%)	NR	NR
Bilateral OA, *n*/*N* (%)	NR	NR
HHS, mean (SD)	NR	NR
OHS, mean (SD)	NR	NR
Efficacy outcomes		
For each timing of assessment please provide a separate table		
For scores, extract only total scores		

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 12 months

Outcome	*THR (metal-on-metal with large head)*	*RS (metal-on-metal)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original)	NR	NR	NA
Mean (SD), mean (SD)			
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	6.3 (NR)	6.8 (NR)	$p = 0.24$ (NS)
WOMAC score, mean (SD)	90.18 (NR)	90.40 (NR)	$p = 0.95$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	MCS: 55.13 (NR)	MCS: 53.87 (NR)	$p = 0.97$
	PCS: 51.28 (NR)	PCS: 51.22 (NR)	$p = 0.55$
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	NR	NR	NA
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR (metal-on-metal with large head)***	***RS (metal-on-metal)***	***Between-group difference and p-value (or 95% CI)***[a]
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	NA	NR	NR	NA
Infection	NA	NR	NR	NA
Femoral neck fracture	NA	NR	NR	NA
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NA	NR	NR	NA
Implant dislocation	NA	NR	NR	NA
Other (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions

After 1 year of follow-up there was no significant difference in mean WOMAC, SF-36 and UCLA activity scores between the two groups

Reviewers' conclusions

Because of the small sample size, the between-group differences in mean WOMAC, SF-36 and UCLA activity scores were not clinically meaningful

MCS, mental component summary score; NA, not applicable; NR, not reported; PCS, physical component summary score; PE, polyethylene.

Vendittoli et al.,[132,133,136] *Girard* et al.[134] *and Rama* et al.[135]

Name of the first reviewer: Alexander Tsertsvadze

Name of the second reviewer: Paul Sutcliffe

Study details
Country: Canada
Study design: RCT
Study setting (primary care/specialty clinic/other – specify): specialty clinic
Number of centres: 1
Funding (government/private/manufacturer/other – specify): NR
Aim of the study
To compare THR (metal-on-metal) with RS (metal-on-metal) with respect to postoperative clinical and radiographic outcomes and complications
Participants
Recruitment dates: August 2003 and January 2006
Total number of patients screened for inclusion eligibility: 219 (hips)
Total number of patients randomised: 192 (209 hips)
Inclusion criteria: patients aged 18–65 years with degenerative hip joint disease who were candidates for both metal–metal THR and metal–metal RS were recruited to the study
Exclusion criteria: implantation, hip arthrodesis, renal insufficiency, known or suspected metal allergy, osteopenia or osteoporosis of the hip
Characteristics of participants (total study sample): mean (range) age (years): 50.0 (23–65); women, *n* (%): 73 (34.7); race/ethnicity, *n* (%): NR; diagnosis, *n* (%): primary OA 70 (33.3), impingement hip 77 (36.6), protrusion 12 (5.7), Perthes' disease 6 (2.8), dysplasia (Crowe I and II) 17 (8.0), osteonecrosis 5 (2.4), trauma 5 (2.4), inflammatory arthritis 16 (7.6), RAs 13 (6.0), ankylosing spondylitis 3 (1.4), postseptic arthritis 2 (1.0)

Intervention (keep the same order as in the paper)

Intervention 1 (e.g. THR 1): THR (metal-on-metal)

Intervention 2 (e.g. RS): RS (metal-on-metal)

Bilateral procedure (yes/no/NR): yes

Implant manufacturer: Zimmer, Inc. (Winterthur, Switzerland)

Postprocedural rehabilitation (e.g. weight-bearing, exercise): postoperatively, weight bearing as tolerated was allowed in the THR group and protected weight bearing for 3–4 weeks was advised for the RS group. All patients were discharged home; discharge was allowed once the patient could walk safely for > 50 m, climb stairs, transfer to and from the bed safely and perform exercise programmes independently

If several types of THR are compared, indicate the basis for comparison (tick all that apply): NA

Cup type (e.g. monoblock, custom, preassembled/modular, standard, other)

Cup fixation (e.g. cemented, cementless, other)

Cup composition (e.g. metal, metal–ceramic, metal–PE, PE, other)

Cup liner composition (e.g. metal, ceramic, PE, polyurethane, cross-linked, other)

Cup size (mm)

Femoral head type (e.g. modular, custom, other)

Femoral head composition (e.g. ceramic, metal, other)

Femoral head size (mm)

Stem type (e.g. monolithic, modular/tapered, other)

Stem composition (e.g. titanium, cobalt–chromium and stainless steel)

Stem fixation (e.g. cemented, cementless, other)

Outcomes (study based)

Primary outcomes (*list*): WOMAC, Merle d'Aubigné and Postel and UCLA activity scores, incision length, surgical time, surgical blood loss, complications (e.g. dislocation, fracture, loosening), patient satisfaction and length of hospital stay

Secondary outcomes (*list*): radiographic measures (e.g. acetabular vertical angle, RS femoral component CCD angle, RS CCD angle modification from preoperative value, pre-/postoperative leg length discrepancy, femoral offset)

Imaging method used (i.e. conventional radiography, radiostereometry, none): conventional radiography

Postprocedural timings of primary outcome assessment (e.g. 6 months, 12 months, post operation): 3, 6, 12 and 24 months and 5 years

Total length of follow-up: 6 years

Number of patients

	Total	***THR (metal-on-metal)***	***SR (metal-on-metal)***
Randomised	209	100	109
Analysed (if more than one follow-up point, choose and specify the last one)	157	72	85
Losses to follow-up/dropouts/sample attrition (if more than one follow-up point, choose and specify the last one)	NR	NR	NR

Interventions		
	Description (e.g. intervention type, composition/bearing materials, fixation)	***Operator characteristics***
THR (metal-on-metal)	A posterior approach was used for both groups. Incision length was left to the surgeon's discretion. Standard instruments were used for all operations. The external rotators were completely released while exposing and were reattached with transosseous sutures when closing. The THR group received the CLS Spotorno femoral stem, the Allofit acetabular cup, a wrought high-carbon cobalt–chromium PE sandwich acetabular insert and a 28-mm femoral head. Using the 135° or 145° neck shaft angle stem with different head neck length (–4 mm to +8 mm), surgeons took care to reproduce as closely as possible the patient's leg length and offset	NR
RS (metal-on-metal)	A posterior approach was used for both groups. Incision length was left to the surgeon's discretion. Standard instruments were used for all operations. The external rotators were completely released while exposing and were reattached with transosseous sutures when closing. In the SR group the posterior approach included a circumferential capsulotomy and a partial elevation of the gluteus minimus from the supra-acetabular bone and, when necessary to increase femoral mobilisation, a complete release of the gluteus maximus femoral tendon was performed. The SRA group received the hybrid Durom resurfacing system with a wrought high-carbon cobalt–chromium femoral head and acetabular cup	NR

Patient baseline characteristics		
Characteristic	***THR (metal-on-metal)***	***RS (metal-on-metal)***
Age (years), mean (SD)	51.0 (8.6)	49.2 (9.0)
Sex, female, *n*/*N* (%)	32/100 (32.0)	40/109 (37.0)
Weight (kg), mean (SD)	NR	NR
BMI (kg/m^2), mean (SD)	30.0 (6.8)	27.0 (5.3)
Primary OA, *n*/*N* (%)	39/100 (39.0)	34/109 (31.2)
Bilateral OA, *n*/*N* (%)	NR	NR
HHS, mean (SD)	NR	NR
OHS, mean (SD)	NR	NR

Efficacy outcomes

For each timing of assessment please provide a separate table

For scores, extract only total scores

Postprocedural follow-up assessment timing (specify): 3 months

Outcome	*THR (metal-on-metal)*	*RS (metal-on-metal)*	*Between-group difference and p value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	15.8 (NR)	16.2 (NR)	$p = 0.599$ (NS)
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	19.2 (NR)	19.9 (NR)	$p = 0.767$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	1/102 (1.0)	0/103 (0.0)	NR
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 6 months

Outcome	*THR (metal-on-metal)*	*RS (metal-on-metal)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	17.1 (NR)	17.2 (NR)	$p = 0.721$ (NS)
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	11.3 (NR)	13.9 (NR)	$p = 0.205$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	1/102 (1.0)	1/103 (1.0)	NR
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 12 months

Outcome	*THR (metal-on-metal)*	*RS (metal-on-metal)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	16.6 (NR)	16.7 (NR)	$p = 0.942$ (NS)
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	6.3 (NR)	7.1 (NR)	$p = 0.037$ (SS)
WOMAC score, mean (SD)	11.7 (NR)	9.2 (NR)	$p = 0.363$ (NS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	1/102 (1.0)	2/103 (2.0)	NR
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 24 months

Outcome	*THR (metal-on-metal)*	*RS (metal-on-metal)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n/N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	17.5 (1.3)	17.5 (1.3)	$p = 0.942$ (NS)
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR (NR)	NR (NR)	$p = 0.094$ (NS)
WOMAC score, mean (SD)	9.0 (11.9)	5.7 (8.6)	$p = 0.007$ (SS)
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n/N* (%)	1/102 (1.0)	2/103 (2.0)	NR
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Efficacy outcomes

Postprocedural follow-up assessment timing (specify): 56 months (5 years)

Outcome	*THR (metal-on-metal)*	*RS (metal-on-metal)*	*Between-group difference and* p*-value (or 95% CI)*[a]
Mortality (all-cause), *n*/*N* (%)	NR	NR	NA
HHS, mean (SD)	NR	NR	NA
OHS, mean (SD)	NR	NR	NA
Merle d'Aubigné and Postel score (specify if modified or original), mean (SD)	NR	NR	NA
HOOS, mean (SD)	NR	NR	NA
LISOH score, mean (SD)	NR	NR	NA
AAOS Hip and Knee Questionnaire score, mean (SD)	NR	NR	NA
UCLA activity score, mean (SD)	NR	NR	NA
WOMAC score, mean (SD)	NR	NR	NA
AIMS score, mean (SD)	NR	NR	NA
MACTAR score, mean (SD)	NR	NR	NA
SF-36 score, mean (SD)	NR	NR	NA
SF-12 score, mean (SD)	NR	NR	NA
NHP score, mean (SD)	NR	NR	NA
EQ-5D score, mean (SD)	NR	NR	NA
Pain score (VAS), mean (SD)	NR	NR	NA
Pain score (other than VAS; specify), mean (SD)	NR	NR	NA
Revision rate, *n*/*N* (%)	2/100 (2.0)	4/109 (4.0)	$p = 0.470$ (NS)
Time to revision (years), mean (SD)	NR	NR	NA
Implant survival rate (Kaplan–Meier estimate and 95% CI), %	NR	NR	NA
Femoral head penetration (mm/year), mean (SD)	NR	NR	NA

a RR, risk difference or MD (specify if it is between mean change values from baseline or between mean final end-point values).

Complications – *n*/*N* (%) patients with an event (if more than one follow-up point, choose and specify the last follow-up)				
Complication	***Time of occurrence (peri-/post-operational)***	***THR (metal-on-metal)***	***RS (metal-on-metal)***	***Between-group difference and p-value(or 95% CI)[a]***
Osteolysis (any or total)	NA	NR	NR	NA
Aseptic loosening (any or total)	Post	0/100 (0.0)	6/109 (6.0)	$p = 0.017$ (SS)
Infection	Post	5/100 (5.0)	0/109 (0.0)	$p = 0.02$ (SS)
Femoral neck fracture	Peri	4/100 (4.0)	0/109 (0.0)	$p = 0.038$ (SS)
Metallosis	NA	NR	NR	NA
Muscle weakness	NA	NR	NR	NA
Nerve palsy	NA	NR	NR	NA
Deep-vein thrombosis	NR	3/100 (3.0)	1/109 (1.0)	Calculated RR 3.27 (0.3 to 30.9) (NS)
Implant dislocation	Post	4/100 (4.0)	0/109 (0.0)	$p = 0.038$ (SS)
Other (specify)	NA	NR	NR	NA

a RR or risk difference.

Authors' conclusions
After 2 years of follow-up, the mean WOMAC score (but not UCLA activity and Merle d'Aubigne scores) was significantly better (lower) in the RS group than in the THR group. After 5 years there was no significant difference between the groups in revision rates. The rates of infection, implant dislocation and fracture were significantly lower in the RS group than in the THR group. The RS group, however, experienced a significantly higher incidence of aseptic loosening
Reviewers' conclusions
In general, on long-term follow-up, RS showed a slightly better efficacy and safety profile than THR

CCD, caput collum diaphysis; NA, not applicable; NR, not reported; NS, not significant; PE, polyethylene; SS, statistically significant.

Included systematic reviews: total hip replacement compared with resurfacing arthroplasty ($n = 3$)

Jiang et al.[142]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: China

Funding: NR

Aim of the study

To compare the clinical results of metal-on-metal RS with those of standard THR for the treatment of hip disease in active young patients

Methods

Search strategy

Databases searched: The Cochrane Bone, Joint and Muscle Trauma Group Specialised Register (June 2009), Cochrane Central Register of Controlled Trials (The Cochrane Library, Issue 2, 2009), PubMed (January 1990–June 2009), Ovid (January 1990–June 2009), Science Direct Online (January 1990–June 2009)

Last date of search: June 2009

Other methods of identifying literature: Several orthopaedic journals, conference proceedings and article reference lists were searched. The authors of the studies were contacted for further information

Inclusion criteria

Participants: age < 65 years, skeletally mature, with end-stage hip disease, followed up for > 12 months

Interventions: modern metal-on-metal RS

Comparators: standard THR

Outcome measures: rate of revision, mortality, femoral neck fracture, component loosening, dislocation and deep hip joint infection, as well as hip function and range of motion

Types of studies included (i.e. study design): RCTs and controlled clinical trials

Study quality assessment methods

Quality assessment tool used: Cochrane risk of bias assessment tool

Risk of bias assessment criteria applied to included studies: sequence generation, allocation concealment, blinding, incomplete outcome data, selective outcome reporting and other issues

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis) ☐

(b) Quantitative (meta-analysis) ☒

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used) ☐

(b) Adjusted (anchored indirect treatment comparison) ☐

(c) Mixed treatment comparison ☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression) ☒

(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests) ☐

(c) Overall quality/strength of evidence (GRADE) ☐

Results

Number of included studies: four

Overall risk of bias of included studies: three studies were adequately randomised and the fourth was randomised only by the patients' date of admission. Two studies included patients who were lost before follow-up could be completed, and the allocation concealments of all four eligible studies were unclear. None of the studies included adequate blinding procedures

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available):

Revision rate: the pooled results showed a higher incidence of revision for RS than THR (four studies; pooled RR 2.60, 95% CI 1.31 to 5.15 at 1–10 years' follow-up)

Dislocation rate: the pooled results showed no significant difference between RS and THR (three studies; pooled RR 0.25, 95% CI 0.05 to 1.21 at 1–2 years' follow-up)

Mortality rate: one study showed no significant difference between RS and THR (RR 1.05, 95% CI 0.24 to 4.66 at 3 years' follow-up)

Femoral neck fracture: the incidence of femoral neck fracture may have been higher for RS than THR (three studies)

Component loosening: the pooled results showed a significantly higher incidence of component loosening for RS than THR (four studies; pooled RR 4.96, 95% CI 1.82 to 13.50 at 1–10 years' follow-up)

Deep hip joint infection: the pooled results showed no significant difference between RS and THR (three studies; pooled RR 2.25, 95% CI 0.61 to 8.31 at 1–3 years' follow-up)

Functional scores: no significant difference in mean WOMAC score, HHS and Merle d'Aubigné and Postel score between RS and THR (three studies). The mean UCLA activity score was significantly higher for RS than for THR at 1–2 years' follow-up (two studies)

Strength of evidence per outcome (GRADE): NA

Conclusions: The results indicated increased rates of revision, femoral neck fracture and component loosening among patients who received RS than among patients who received THR. No significant differences in the rates of mortality, dislocation or deep hip joint infection were found between the groups. Hip function scores were similar between the two groups but the RS group showed higher activity levels

Reviewers' conclusions and additional comments/concerns

RS demonstrated improved rates of revision, femoral neck fracture and component loosening compared with THR. The RS and THR groups experienced a similar degree of improvement in function. RS was associated with better activity scores

NA, not applicable; NR, not reported.

Smith et al.[143]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: UK

Funding: NR

Aim of the study

To compare THR and RS for clinical and radiological outcomes and complication rates

Methods

Search strategy

Databases searched: MEDLINE (1950–January 2010), CINAHL (1982 to January 2010), AMED (1985 to January 2010) and EMBASE (1974 to January 2010)

Last date of search: 10 January 2010

Other methods of identifying literature: the following sources were used to search for unpublished literature: SIGLE (System for Information on Grey Literature in Europe), the National Technical Information Service, the National Research Register (UK), the British Library's Integrated Catalogue, Current Controlled Trials, conference proceedings of the British Orthopaedic Association (BOA) Annual Congress, the European Federation of National Associations of Orthopaedics and Traumatology (EFORT) and the British Hip Society. The reference lists of relevant papers were also searched for papers not identified by the initial search and the corresponding authors were contacted for citations not identified from the original searches

Inclusion criteria

Participants: patients with hip pathology

Interventions: RS

Comparators: THR

Outcome measures: incidence of revision, mortality, dislocation, aseptic loosening, avascular necrosis, infection and fracture; incision length, last acetabular reamer size, duration of operation, blood loss and frequency of blood transfusion requirement, length of hospital stay, pain, functional and quality-of-life outcomes and hip range of motion; femoral/acetabular offset, incidence of femoral/acetabular radiolucency, leg length, cup height and heterotopic ossification; incidence of complications (venous thromboembolic events, acetabular malposition, trochanteric malunion or non-union, nerve palsy and presence of Trendelenburg sign)

Types of studies included (i.e. study design): RCTs and non-RCTs

Study quality assessment methods

Quality assessment tool used: modified 17-item appraisal tool (CASP)

Risk of bias assessment criteria applied to included studies: subject identification, randomisation, blinding and dropout rates; presentation of results using descriptive and inferential statistics; and external validity to clinical practice

APPENDIX 4

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis)	☐
(b) Quantitative (meta-analysis)	☒

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used)	☐
(b) Adjusted (anchored indirect treatment comparison)	☐
(c) Mixed treatment comparison	☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression)	☐
(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests)	☒
(c) Overall quality/strength of evidence (GRADE)	☐

Results

Number of included studies: 46 (10 RCTs and 36 observational studies)

Overall risk of bias of included studies: CASP score (maximum 17): 0–3: one study; 4–7: 16 studies; 8–11: 23 studies; 12–15: five studies; 16–17: no studies. Nine RCTs clearly described the method of randomisation; for 25 studies the groups were comparable at baseline; assessor blinding was used in four studies; patients were blinded in only two studies; in 16 studies the results were analysed by intention-to-treat methods; the results were interpreted appropriately in 35 studies

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available):

HHS: higher in RS than in THR group (RS favoured) (no. of studies NR; pooled MD 2.51, 95% CI 1.24 to 3.77)

WOMAC score: lower in RS than in THR group (RS favoured) (no. of studies NR; pooled MD –2.41, 95% CI –3.88 to –0.94)

Revision rate: higher in RS than in THR group (THR favoured) (19 studies; pooled RR 1.7, 95% CI 1.2 to 2.5)

Aseptic loosening: higher in RS than in THR group (THR favoured) (10 studies; pooled RR 3.1, 95% CI 1.1 to 8.5)

Dislocation: lower in RS than in THR group (RS favoured) (number of studies NR; pooled RR 0.2, 95% CI 0.1 to 0.5)

There was no statistically significant difference between the RS and THR groups for mean Merle d'Aubigné and Postel score, UCLA activity score or OHS as well as for incidence of mortality, fracture, deep-vein thrombosis, sciatic nerve palsy and joint infection

Strength of evidence per outcome (GRADE): NA

Conclusions: the findings indicate that functional outcomes following RS are better or the same as those following THR but that there is an increased risk of complications after RS such as aseptic loosening and revision. THR would therefore appear to be superior to RS

Reviewers' conclusion and additional comments/concerns

Although RS shows promising results in terms of improved clinical outcomes, it is associated with a greater risk of complications such as revision and aseptic loosening than THR

AMED, Allied and Complementary Medicine Database; CASP, Critical Appraisal Skills Programme; CINAHL, Cumulative Index to Nursing and Allied Health Literature; NA, not applicable; NR, not reported.

Springer et al.[144]

Name of first reviewer: Alexander Tsertsvadze

Name of second reviewer: Paul Sutcliffe

Study details

Country: Canada

Funding: no benefits or funds were received in support of the study

Aim of the study

To compare the effects of THR and RS on failure rates of cementless femoral components in younger patients

Methods

Search strategy

Databases searched: MEDLINE, PubMed and CINAHL searched from their inception

Last date of search: 31 March 2008

Other methods of identifying literature: reference lists from review articles and potentially relevant studies were hand searched

Inclusion criteria

Participants: young adults (mean age < 55 years)

Interventions: THR with modern cementless components

Comparators: RS

Outcome measures: femoral failure for any reason, femoral failure because of revision and femoral failure for mechanical reasons

Types of studies included (i.e. study design): RCTs, observational studies including single-arm studies (THR or RS only)

Study quality assessment methods

Quality assessment tool used: methodological quality assessment was performed by a single reviewer by assigning non-randomised studies a Methodological Index for Non-Randomized Studies score

Risk of bias assessment criteria applied to included studies: NR

APPENDIX 4

Methods

Methods of synthesis

Direct comparison (tick if applicable):

(a) Non-quantitative (i.e. no meta-analysis) ☐

(b) Quantitative (meta-analysis) ☐

Indirect comparison (tick if applicable):

(a) Unadjusted (naive indirect treatment comparison – no common comparator is used) ☒

(b) Adjusted (anchored indirect treatment comparison) ☐

(c) Mixed treatment comparison ☐

Specific methods of assessment (tick if applicable):

(a) Heterogeneity (e.g. forest plot visual examination, subgroup analysis, meta-regression) ☐

(b) Publication bias (e.g. search of grey literature, funnel plots, statistical tests) ☐

(c) Overall quality/strength of evidence (GRADE) ☐

Results

Number of included studies: 37

Overall risk of bias of included studies: NR

Treatment effect per outcome (narrative statement and numerical data for pooled estimate if available):

Overall pooled failure rate for any reason (revision ± radiography outcome): RS (15 studies): 3.7% (95% CI 2.0% to 6.5%) vs. THR (19 studies): 11.6% (95% CI 7.5% to 17.4%)

Cup failure for any reason (revision ± radiography outcome): RS (15 studies): 1.4% (95% CI 0.5% to 3.4%) vs. THR (21 studies): 10.5% (95% CI 7.0% to 15.4%)

Femoral failure rate for any reason (revision ± radiography outcome): RS (15 studies): 2.8% (95% CI 2.0% to 4.0%) vs. THR (22 studies): 3.2% (95% CI 2.4% to 4.2%)

Femoral failure rate because of revision surgery: RS (15 studies): 2.7% (95% CI 1.8% to 4.0%) vs. THR (22 studies): 2.7% (95% CI 2.1% to 3.5%)

Strength of evidence per outcome (GRADE): NA

Conclusions: implant failure rates were lower in the RS group than in the THR group after 8 years of follow-up

Reviewers' conclusion and additional comments/concerns

The authors conducted an inappropriate synthesis, namely an indirect naive comparison of pooled rates without actually comparing them as in a conventional meta-analysis (no measures of association reported or formal statistical tests)

CINAHL, Cumulative Index to Nursing and Allied Health Literature; NA, not applicable; NR, not reported.

Appendix 5 Summary of manufacturer submissions

DePuy International Ltd

Contents of the submission

DePuy provided an economic model in Excel and a 244-page technology assessment of the clinical effectiveness and cost-effectiveness of THR and RS for the treatment of pain or disability in adult patients with end-stage arthritis of the hip.

DePuy investigated the following comparators:

- different types of primary THR and hip RS compared with each other for people in whom both procedures are suitable
- different types of primary THR compared with each other for people in whom hip RS is not suitable.

The assessment included comprehensive systematic reviews of the effectiveness and cost-effectiveness of the comparisons under review and a cost–utility analysis using a Markov model with probabilistic sensitivity analysis. The report provided details on methodology including inclusion criteria, details of the searches and databases searched for the reviews, and model structure, assumptions and sources of data for the model. The model considered the following hip replacement procedures:

- cemented THR
- cementless THR
- hybrid THR
- reverse hybrid THR
- hip RS.

Data for the model were generally derived from the NJR (revision rates), the literature (utility data) and a microcosting analysis (costs). The national PROMs database[97] and the New Zealand Joint Registry[376] were further data sources.

The overall conclusions were that THR dominated hip RS in patients suitable for both procedures and that DePuy cemented THR was the optimal treatment strategy for patients, both those who were suitable and those who were unsuitable for hip RS. Between different classes of THR, costs and QALYs overlapped considerably in sensitivity analyses for both patient populations.

DePuy recommended that the choice of prosthesis not only should be based on the results of cost–utility analyses but also should take into consideration the operational issues associated with the provision of hip replacement, the impact of training, the variability in costs and results between centres and the preference of different centres for the use of particular implants on the basis of effectiveness, efficiency and costs at a local level.

Literature search considerations

The searches reported in the manufacturer's submission are thorough and accurate. However, there are several concerns:

1. The MEDLINE In-Process & Other Non-Indexed Citations database was searched in the normal MEDLINE database with a strategy that ends by using limits assigned by NLM indexers, which means that all of the In-Process articles that the search initially found would not have been retrieved in the final set.
2. Most of the searches are limited by age group, which is not good practice because not all articles are age specific and NLM's indexing by age can be unreliable. For example, the systematic review by Ethgen *et al.*[192] included in the current report would not have been retrieved because it has not been indexed for age.

The limitations and strengths of the clinical effectiveness review provided by DePuy are listed in *Box 1*.

BOX 1 Limitations and strengths of the clinical effectiveness review by DePuy

Limitations

- Search strategy limitations.
- Omission of grey literature search.
- No standardised quality assessment of the included studies.
- No risk-of-bias assessment.
- Unclear if the extracted data were cross-checked by another reviewer.
- No narrative synthesis of study and baseline population characteristics is provided (only in tables).
- Narrative presentation of results not synthesised (given separately for each study). The authors left the task of synthesising the evidence to a reader.
- No list of excluded studies.
- Tables with study results are not provided.
- Conclusions are vague for both comparisons, with no take-home message. What are the overall findings? Are they conclusive? If not, there should be a statement acknowledging that the findings are inconclusive because of clinical heterogeneity, inconsistent results, etc.
- There is no discussion section in the report; instead, a short paragraph (i.e. conclusion) is presented.
- No information on the validity of the findings, implications, knowledge gaps, future research needs and limitations/advantages of the review, etc. is provided.
- There is no section on equity considerations.

Strengths

- The manufacturer's description of the underlying health problem is appropriate and relevant to the decision problem under consideration.
- The manufacturer's overview of current service provision appears appropriate and relevant to the decision problem under consideration.
- The clinical evidence submitted by the manufacturer appears to reflect the characteristics of the patient population in England and Wales eligible for treatment.
- The interventions described by the manufacturer match the interventions described in the final scope.
- The comparators described by the manufacturer match the comparators described in the final scope.
- The outcomes described by the manufacturer match the outcomes described in the final scope.
- The research question is clearly formulated.
- A comprehensive search was carried out.
- Inclusion/exclusion criteria provided.
- Independent screening and data extraction were carried out.
- Study and baseline population characteristics in tables are well presented.

Cost-effectiveness review: overall quality considerations

The reviews undertaken to identify health state utilities and costs for use in the economic analysis are comprehensive and accurate, using comprehensive searches and inclusion/exclusion criteria that are in line with the research question. A small number of relevant papers were not retrieved by the searches. The limitation of the cost review to studies reporting cost–utility analyses and cost per QALY outcomes might have restricted the review, resulting in studies reporting basic costs and/or resource use for patients undergoing THR or RS being missed. The study selection is transparent; however, no table of excluded studies with reasons is given. The review did not provide a standardised quality assessment of the included studies nor of the key studies that provided data for the economic model. The data extraction tables are detailed but there is no indication if data extraction was cross-checked by another reviewer. The review is lacking a narrative description of the included studies.

Even though the reviews identified a number of relevant studies, only one key study was selected to provide data on utilities and one key study was selected to provide data on costs for the model. These two studies[38,298] investigated THR only and did not provide up-to-date utility and cost data on THR as well as RS, revision and follow-up.

Model structure

A Markov model using a state transition approach was developed in Microsoft Excel. The structure of the model is consistent with previous cost-effectiveness models of THR for the Health Technology Assessment programme.[19,38] The manufacturer considered two cohorts of patients with pain and disability resulting from arthritis of the hip, one for whom hip RS or THR is suitable and one for whom RS is not suitable and who received THR. The population selected and the interventions and comparators are appropriate, as outlined in the NICE scope.[375] The model assumes a quarter-year cycle and a lifetime horizon is adopted. The perspective adopted for the analysis is that of the NHS and PSS. Both costs and benefits were discounted at 3.5%.

Revision rates

Revision rates were modelled using a single Weibull fit that predicted a monotonic decreasing hazard over time. A bathtub hazard was briefly considered following Briggs *et al.*[38] The graphs of observed revision rates that were included in the submission indicate that for most an increasing rate of revision occurred from about 4 years after primary hip replacement and therefore it is likely that a bathtub model could have been used. The submission acknowledges that this is a limitation of the modelling. The manufacturer's probabilistic analysis was described as 'including the use of multivariate distribution for revision model regression parameters'; however, this was difficult to confirm with the model version received.

The submission claims that the Weibull parametric distribution was 'chosen because all previous economic evaluations which assumed parametric distributions assumed Weibull distributions', naming the models of Briggs *et al.*[38] and Higashi and Barendregt.[273] This statement is misleading because these two models used two Weibull fits (one to early and one to late failures) to generate a U-shaped hazard, whereas in direct contrast the manufacturer's single Weibull generates a monotonic decreasing hazard.

Health-related quality of life

The manufacturer had undertaken a systematic review to identify utility scores relating to THR or RS. The EQ-5D scores used in the model were based on Rolfson *et al.*,[298] but this paper reports preoperative and postoperative utility scores following THR only. It should be noted that the majority of patients suitable for hip RS are active males aged < 55 years, as detailed in the manufacturer's submission. The use of the same utility scores for the THR population and the hip RS population is questionable. In the study by Costa *et al.*[130] the EQ-5D scores were collected during a RCT that reports data for RS and THR patients at baseline and 3, 6 and 12 months' follow-up. We acknowledge that the sample size for this study was small (THR 66 patients; RS 60 patients) but the data collected were from a UK-based RCT.

The manufacturer has applied a disutility score of 0.145 following revision and referenced it to Briggs *et al.*[367] It should be noted that the figure for disutility was originally from a regression model output. Dawson *et al.*[295] reported the mean EQ-5D score of 601 revision patients in the UK at 1 year as 0.62. However, applying disutility (0.145) to the postoperative utility score does not reflect the lower quality of life as reported in the original study (0.62 vs. 0.635).

Resources and costs

The cost-effectiveness of DePuy cementless prostheses (i.e. Corail/Pinnacle) and DePuy cemented prostheses was compared with the cost-effectiveness of different types of THR and RS. In the base-case analysis the costs were based on a microcosting analysis and in scenario analysis the costs were based on NHS reference costs. It was assumed that all patients who received primary THR received a metal-on-polyethylene articulation (regardless of whether they received a cemented, cementless or hybrid prosthesis). We agree with the manufacturer that the list price for DePuy products does not reflect the price available to the NHS and in the absence of any further prosthesis cost data from the literature we had to rely on the manufacturer's submission. The sample size used to estimate the costs from the NHS hospital/time and motion study has not been specified in the manufacturer's report and this will increase the uncertainty around the cost data inputs.

The expert opinion of the review team suggests that the cost of the RS prosthesis is higher. It should be noted that Vale *et al.*[19] used a higher cost estimate for RS than for THR. An additional scenario analysis using costs from Vale *et al.*,[19] as indicated in the manufacturer's submission, could not be identified in the report.

In the absence of any further data on surgical resource use, costs and theatre time, the review team had to rely on Appendices E and H of the submission.

The expert opinion of the review team suggests that the cost of revision surgery is greater than the cost of primary THR/RS; however, revisions are carried out for a variety of reasons and to assume that the cost of all revision procedures is the same is not reasonable. In light of this, the manufacturer should have presented a sensitivity analysis around the costs associated with different indications for revision surgery.

Sensitivity analysis

The manufacturer undertook a range of univariate sensitivity analyses and also additional scenario analyses. As mentioned earlier, there is no sensitivity analysis reported around the costs used by Vale *et al.*;[19] given that the cost of revision increased by only 45% and not double, the cost of revision should have been tested using inflated Vale *et al.*[19] costs. The utility score for RS was assumed to be the same as that for THR and sensitivity analysis varying the utility score for RS from published data sources should have been undertaken.

Results

The manufacturer has presented base-case deterministic and probabilistic results. All THRs dominate RS. No attempts were made to identify the cost-effectiveness of the different types of prosthesis based on age and sex. Subgroup analysis of patients based on age and sex should have been undertaken when comparing THR and RS because of the dissimilarities among the patient populations. In the patient population in whom RS was not suitable, DePuy cemented THR was reported as the most cost-effective intervention.

The base-case probabilistic results are similar to the base-case deterministic results. Although the model was probabilistic, the parameters in the model were assumed to be independent and no attempt has been made to check for correlation between the parameters.

Rerunning the model, the base-case deterministic and probabilistic results cross-check with those of the manufacturer.

In the base-case analysis the manufacturer's submission was largely in line with the NICE reference case. However, costs in the base-case analysis were not based on NHS reference costs but on a microcosting study. As mentioned earlier, the sample size for the microcosting study was not mentioned; however, the NHS reference cost estimates were based on a large sample size for both primary and revision surgery ($n = 43,420$ for primary surgery and $n = 26,797$ for revision surgery). Applying the NHS reference costs to both patient cohorts, the optimal strategy at a WTP threshold of £20,000 per QALY was hybrid THR for both patient cohorts. The above suggests that a key uncertainty of the model is the cost inputs that have been used.

Conclusions

The manufacturer's submission is rigorous and complete with regard to relevant clinical studies and relevant data within those studies. The submission contains an unbiased analysis of the literature in terms of treatment effects in relation to relevant populations, interventions, comparators and outcomes. There are uncertainties about the reliability of the clinical effectiveness evidence because of weaknesses highlighted related to transparency, synthesis and lack of quality assessment. The submitted evidence reflects the decision problem defined in the final scope. The main conclusion of the cost-effectiveness analysis was that the DePuy devices are more cost-effective than all other prostheses. In patients suitable for both procedures, hip RS was dominated by cemented THR, cementless THR, DePuy cementless THR (Corail/Pinnacle), DePuy cemented THR, hybrid THR and reverse hybrid THR. It was also noted that DePuy cemented THR was the optimal treatment strategy in both patient populations in the base-case analysis. It should be noted that these conclusions cannot be verified as the cost data were derived from a microcosting analysis of a single centre.

Smith & Nephew, Inc.

Contents of the submission

Smith & Nephew provided a 10-page non-systematic summary of the literature. Evidence was presented on the factors that should be included in the sensitivity analysis of a cost-effectiveness model. No methodology was reported and no economic evaluation was presented. The evidence was drawn from the literature as well as the National Joint Registries of England and Australia. It was concluded that revision rates (and implant prices) drive the cost-effectiveness of THR and that bearing surfaces are known factors that impact revision rates following primary THR and should therefore be considered in sensitivity analyses of economic evaluations.

Literature search considerations

No details of any search methods were reported. The limitations and strengths of the clinical effectiveness review provided by Smith & Nephew, Inc. are listed in *Box 2*.

Cost-effectiveness review: overall quality considerations

Smith & Nephew provided a non-systematic coverage of the cost-effectiveness evidence concerning revision surgery post THR. The research question therefore only partially meets the decision problem under consideration. No methods were reported in terms of the literature search, inclusion criteria, data extraction and synthesis of evidence. No quality assessment of included studies was reported nor was a table of excluded studies provided. The cost-effectiveness review included a number of key papers but the list of included studies was not exhaustive, probably because of the focus on revision.

BOX 2 Limitations and strengths of the clinical effectiveness review by Smith & Nephew, Inc.

Limitations

- A limited non-systematic review of the clinical effectiveness literature with a clear focus on revision surgery is provided.
- Lack of a clearly formulated research question.
- No methodology reported.
- No search strategy reported.
- Inclusion/exclusion criteria not specified.
- Intervention considered does not include RS.
- Outcomes solely consider revision following THR.
- No standardised quality assessment of the included studies.
- No risk-of-bias assessment.
- No details concerning the methods of screening and data extraction.
- No list of excluded studies.
- Narrative presentation of results not synthesised.
- Study and baseline population characteristics are not clearly presented.
- Tables with study results are not presented.
- No section on equity considerations.

Strength

- Revision rates by bearing surface extrapolated to 11 years.

Conclusions

The report includes a subjective summary of the importance of bearing surfaces for revision rates and a justification for considering bearing surfaces in a sensitivity analysis within the cost-effectiveness model of the NICE report.[375] It concludes that known factors that modify revision rates, such as bearing surfaces, should be considered in analyses. It suggests that individual prostheses or design elements should be considered separately in analyses so that their impact on revision rates does not get lost when grouping new technology implants for analysis. Overall, the report lacks objectivity, transparency and methodological rigour.

Stryker

Contents of the submission

Stryker provided a 22-page report that consisted of an executive summary and a review of the literature without any evidence of a systematic review. The report did not include any methodology on how the evidence was collected nor did it report any economic analyses. Stryker considered cemented and cementless THR as well as RS and summarised the complexity of available implants and corresponding revision rates considering evidence from the literature and the National Joint Registries of England and Wales, Sweden, Norway and Australia. It was concluded that the complexity of hip replacement procedures should be taken into consideration in economic evaluations and was reported that Stryker is currently working with a group of researchers at the University of East Anglia and orthopaedic surgeons to develop a cost-effectiveness model to address the above-mentioned issues.

Literature search considerations

No details of the search methods were reported. The limitations and strengths of the clinical effectiveness review provided by Stryker are listed in *Box 3*.

Cost-effectiveness review: overall quality considerations

Stryker provided a limited non-systematic coverage of the cost-effectiveness evidence concerning THRs. A brief statement is made about the complexity of the cost-effectiveness modelling around THR. Stryker state that, 'Few cost-effectiveness studies have been published regarding THR compared to other broadly used surgical interventions'. In contrast, the current report has demonstrated a plethora of evidence on the cost-effectiveness of different types of THR.

BOX 3 Limitations and strengths of clinical effectiveness review by Stryker

Limitations

- A non-systematic review of the clinical effectiveness literature is provided using data referenced to the NJR 2011 annual report.[36]
- Lack of a clearly formulated research question.
- No search strategy.
- Inclusion/exclusion criteria not specified.
- No standardised quality assessment of the included studies.
- No risk-of-bias assessment.
- No details concerning the methods of screening and data extraction.
- No list of excluded studies.
- Narrative presentation of results not synthesised.
- Study and baseline population characteristics are not clearly presented.
- Tables with study results are not presented.
- Conclusions are vague.
- No information on the validity of the findings, implications, knowledge gaps, future research needs and limitations/advantages of the review, etc.
- No section on equity considerations.

Strengths

- PROMs data were reported for THR.
- Revision rates were reported for 3 and 8 years follow-up.
- The manufacturer's description of the underlying health problem is appropriate and relevant to the decision problem under consideration.
- The manufacturer's overview of current service provision appears appropriate and relevant to the decision problem under consideration.
- The clinical evidence submitted by the manufacturer appears to reflect the characteristics of the patient population in England and Wales eligible for treatment.
- The interventions described by the manufacturer match the intervention described in the final scope although the list of included studies is not exhaustive.
- The comparators described by the manufacturer match the comparators described in the final scope.
- The outcomes described by the manufacturer match the outcomes described in the final scope.

Conclusions

Stryker did not answer a clearly formulated question but presented a summary of a selection of the available evidence. Details were provided on the cemented procedure for the Exeter stem (Stryker). Stryker report 'very good' mid-term results for the Exeter V40 stem. Stryker also reported results from the 2011 AOANJRR for the two stems listed above. Various published studies are listed that report positive results for these stems. The information in the report is limited and the report lacks objectivity, transparency, methodological rigour and clear conclusions.

JRI Orthopaedics Ltd

Contents of the submission

JRI Orthopaedics provided a 14-page report detailing a summary of JRI products and a price list of JRI components with limited reference to the literature and data from the National Joint Registries of England and Wales, Sweden and Australia. The submission did not include an economic evaluation in the form of a model. The report compared JRI cementless THR with cemented, hybrid and cementless THR data from the NJR and concluded that revision rates for JRI cementless implants are lower than those for all other cementless THRs, the majority of the hybrid THRs and two of the six categories of cemented THRs. Analysis of risk of revision by liner type and age showed that the risk of revision increased after the age of 70 years when using a poly liner instead of a ceramic liner. Furthermore, a comparison of death rates for cemented compared with cementless JRI implants demonstrated a slightly higher death rate for patients receiving a cemented JRI implant than for patients receiving a cementless JRI implant.

The JRI submission also included detailed clinical evaluation reports on four specific JRI brands including literature reviews and quality appraisals and four technical reports considering the JRI cemented and cementless components, coatings, details of the polyethylene used and specifications of the trunnion design. Finally, JRI submitted statistics from the NJR and complaints data by device collected by JRI.

Literature search considerations

A search strategy was developed for each brand to identify relevant literature over the last 5 years. The authors state that the majority of the literature for the reviews was obtained online. Searches were undertaken in the *Journal of Bone & Joint Surgery*, Entrez PubMed, the NJR and Google Scholar.

The limitations and strengths of the clinical effectiveness review provided by JRI Orthopaedics are listed in *Box 4*.

Cost-effectiveness review: overall quality considerations

The submission provided very limited information on cost-effectiveness.

Conclusions

JRI Orthopaedics presented an overview of its brands. Accompanying reports for each brand were provided as appendices. Average selling prices per component were listed, which were useful. Overall, the report lacks transparency, objectivity and any clear conclusions.

BOX 4 Limitations and strengths of the clinical effectiveness review by JRI Orthopaedics Ltd

Limitations

- Brief scoping reviews of the clinical effectiveness literature for each brand.
- No details concerning the methods of screening and data extraction.
- No section on equity considerations.
- Revision rates for cemented THR and cementless THR from three national joint registries are only briefly discussed.

Strengths

- Inclusion/exclusion criteria specified.
- Clearly formulated research questions.
- Brief search strategy presented.
- Study and baseline population characteristics are presented in tables.
- Quality assessment of the included studies is provided.
- A paragraph on trunnion design is provided.
- List of excluded studies is provided.
- PROMs data reported for THR.
- Revision rates reported for 3 and 8 years' follow-up.
- Information provided on the validity of the findings, implications, knowledge gaps, future research needs and limitations/advantages of the review.
- The manufacturer's description of the underlying health problem is appropriate and relevant to the decision problem under consideration.
- The manufacturer's overview of current service provision appears appropriate and relevant to the decision problem under consideration.
- The clinical evidence submitted by the manufacturer appears to reflect the characteristics of the patient population in England and Wales eligible for treatment.
- The interventions described by the manufacturer match the intervention described in the final scope although the list of included studies is not exhaustive.
- The comparators described by the manufacturer match the comparators described in the final scope.
- The outcomes described by the manufacturer match the outcomes described in the final scope.
- A brief review of the evidence highlighting data from the NJR is provided. This included the number of JRI implants, revision rates for JRI cementless brands with comparative data, survival rates and risk of revision by age group for a Furlong H-A.C THR (JRI), trends in femoral head size, revision rate by liner type with different head size, revision rate by liner type and age group and mortality rates for JRI cemented and cementless implants.

Appendix 6 Excluded papers and reasons for exclusion

Note: excluded papers are those originally excluded at full-text stage or those that were unavailable or excluded based on date/number of patients ($n = 204$).

Paper	Reason for exclusion
Alberta Heritage Foundation for Medical Research. *Metal-on-Metal Hip Resurfacing for Young, Active Adults with Degenerative Hip Disease*. Technote TN 33. Edmonton, AB: Alberta Heritage Foundation for Medical Research (AHFMR); 2002[208]	Exclude – publication date before 2008
Ayers DC, Hays PL, Drew JM, Eskander MS, Osuch D, Bragdon CR. Two-year radiostereometric analysis evaluation of femoral head penetration in a challenging population of young total hip arthroplasty patients. *J Arthroplasty* 2009;**24**(6 Suppl.):9–14[218]	Exclude – total number of patients < 100
Baad-Hansen T, Kold S, Olsen N, Christensen F, Soballe K. Excessive distal migration of fiber-mesh coated femoral stems. *Acta Orthop* 2011;**82**:308–14[381]	Exclude – comparison of different surfaces
Baker RP, Pollard TCB, Eastaugh-Waring SJ, Bannister GC. A medium-term comparison of hybrid hip replacement and Birmingham hip resurfacing in active young patients. *J Bone Joint Surg Br* 2011;**93**:158–63[252]	Exclude – comparison of different coatings
Beckmann J, Stengel D, Tingart M, Gotz J, Grifka J, Luring C. Navigated cup implantation in hip arthroplasty: a meta-analysis. *Acta Orthop* 2009;**80**:538–44[165]	Exclude – non-RCT
Bernath V. *Hip Resurfacing in Patients with Osteoarthritis*. Clayton, Victoria: Centre for Clinical Effectiveness; 2002[382]	Unavailable
Beswick A, Wylde V, Blom A, Gooberman-Hill R, Dieppe PA. Pain after hip or knee joint replacement for osteoarthritis: a systematic review. *Arthritis Rheum* 2011;**63**:S413[383]	Exclude – abstract
Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. *BMJ Open* 2012;**2**:e000435[384]	Exclude – this systematic review excluded RCTs and did not compare THR types
Bhan S, Pankaj A, Malhotra R. One- or two-stage bilateral total hip arthroplasty: a prospective, randomised, controlled study in an Asian population. *J Bone Joint Surg Br* 2006;**88**:298–303[146]	Exclude – comparison of different surgical approaches
Bisseling P, Smolders JMH, Hol A, Van Susante JLC. No clear influence of preference bias on satisfaction and early functional outcome in resurfacing hip arthroplasty. *Acta Orthop* 2011;**82**:161–5[385]	Exclude – observational study comparing satisfaction between randomised (no preference) group and preference (non-randomised) group
Boden H, Adolphson P. No adverse effects of early weight bearing after uncemented total hip arthroplasty: a randomized study of 20 patients. *Acta Orthop Scand* 2004;**75**:21–9[386]	Exclude – early vs. late weight bearing post THR
Boe BG, Rohrl SM, Heier T, Snorrason F, Nordsletten L. A prospective randomized study comparing electrochemically deposited hydroxyapatite and plasma-sprayed hydroxyapatite on titanium stems. *Acta Orthop* 2011;**82**:13–19[387]	Exclude – comparison of different coatings
Brodner W, Bitzan P, Meisinger V, Kaider A, Gottsauner-Wolf F, Kotz R. Serum cobalt levels after metal-on-metal total hip arthroplasty. *J Bone Joint Surg Am* 2003;**85-A**:2168–73[187]	Exclude – publication date before 2008
Butler RA, Rosenzweig S, Myers L, Barrack RL. The Frank Stinchfield Award: the impact of socioeconomic factors on outcome after THA: a prospective, randomized study. *Clin Orthop Relat Res* 2011;**469**:339–47[388]	Exclude – compares two different types of stems
Cai P, Hu Y, Xie J. Large-diameter delta ceramic-on-ceramic versus common-sized ceramic-on-polyethylene bearings in THA. *Orthopedics* 2012;**35**:e1307–13[219]	Exclude – total number of patients < 100

Paper	Reason for exclusion
Carlsson LV, Albrektsson BE, Albrektsson BG, Albrektsson TO, Jacobsson CM, Macdonald W, *et al.* LR. Stepwise introduction of a bone-conserving osseointegrated hip arthroplasty using RSA and a randomized study: I. Preliminary investigations – 52 patients followed for 3 years. *Acta Orthop* 2006;**77**:549–58[188]	Exclude – publication date before 2008
Carlsson LV, Albrektsson T, Albrektsson BE, Jacobsson CM, Macdonald W, Regner L, *et al.* Stepwise introduction of a bone-conserving osseointegrated hip arthroplasty using RSA and a randomized study: II. Clinical proof of concept – 40 patients followed for 2 years. *Acta Orthop* 2006;**77**:559–66[189]	Exclude – publication date before 2008
Cobb J. The functional outcome of hip resurfacing and large-head THA is the same: a randomized, double-blind study. *Clin Orthop Relat Res* 2010;**468**:3134[389]	Exclude – letter to editor
Conroy JL, Chawda M, Kaushal R, Whitehouse SL, Crawford RW, English H. Does use of a 'rim cutter' improve quality of cementation of the acetabular component of cemented Exeter total hip arthroplasty? *J Arthroplasty* 2009;**24**:71–6[147]	Exclude – comparison of different surgical approaches
Corbett KL, Losina E, Nti AA, Prokopetz JJZ, Katz JN. Population-based rates of revision of primary total hip arthroplasty: a systematic review. *PLOS ONE* 2010:**5**:e13520[390]	Exclude – no RCTs included in this systematic review
Coyle D, Coyle K, Vale L, Verteuil R, Imamura M, Glazener C, *et al. Minimally Invasive Arthroplasty in the Management of Hip Arthritic Disease: Systematic Review and Economic Evaluation*. Ottawa, ON: Canadian Agency for Drugs and Technologies in Health (CADTH); 2008[148]	Exclude – comparison of different surgical approaches
Dahlstrand H, Stark A, Anissian L, Hailer NP. Elevated serum concentrations of cobalt, chromium, nickel, and manganese after metal-on-metal alloarthroplasty of the hip: a prospective randomized study. *J Arthroplasty* 2009;**24**:837–45[220]	Exclude – total number of patients < 100
D'Angelo F, Murena L, Zatti G, Cherubino P. The unstable total hip replacement. *Ind J Orthop* 2008;**42**:252–9[391]	Exclude – non-systematic review
D'Arrigo C, Speranza A, Monaco E, Carcangiu A, Ferretti A. Learning curve in tissue sparing total hip replacement: comparison between different approaches. *J Orthop Traumatol* 2009;**10**:47–54[149]	Exclude – comparison of different surgical approaches
de Verteuil R, Imamura M, Zhu S, Glazener C, Fraser C, Munro N, *et al.* A systematic review of the clinical effectiveness and cost-effectiveness and economic modelling of minimal incision total hip replacement approaches in the management of arthritic disease of the hip. *Health Technol Assess* 2008;**12**(26)[11]	Exclude – comparison of different surgical approaches
Digas G, Karrholm J, Thanner J. Addition of fluoride to acrylic bone cement does not improve fixation of a total hip arthroplasty stem. *Clin Orthop Relat Res* 2006;**448**:58–66[392]	Exclude – effect of fluoride; two different cements
Digas G, Karrholm J, Thanner J. Different loss of BMD using uncemented press-fit and whole polyethylene cups fixed with cement: repeated DXA studies in 96 hips randomized to 3 types of fixation. *Acta Orthop* 2006;**77**:218–26[190]	Exclude – publication date before 2008
Digas G, Thanner J, Anderberg C, Karrholm J. Bioactive cement or ceramic/porous coating vs. conventional cement to obtain early stability of the acetabular cup – randomised study of 96 hips followed with radiostereometry. *J Orthop Res* 2004;**22**:1035–43[191]	
Digas G, Karrholm J, Thanner J, Herberts P. 5-year experience of highly cross-linked polyethylene in cemented and uncemented sockets: two randomized studies using radiostereometric analysis. *Acta Orthop* 2007;**78**:746–54[221]	Exclude – total number of patients < 100
Digas G, Karrholm J, Thanner J, Malchau H, Herberts P. Highly cross-linked polyethylene in cemented THA: randomized study of 61 hips. *Clin Orthop Relat Res* 2003;**417**:126–38[218]	
Digas G, Karrholm J, Thanner J, Malchau H, Herberts P. The Otto Aufranc Award. Highly cross-linked polyethylene in total hip arthroplasty: randomized evaluation of penetration rate in cemented and uncemented sockets using radiostereometric analysis. *Clin Orthop Relat Res* 2004;**429**:6–16[223]	
Johanson PE, Digas G, Herberts P, Thanner J, Karrholm J. Highly crosslinked polyethylene does not reduce aseptic loosening in cemented THA: 10-year findings of a randomized study. *Clin Orthop Relat Res* 2012;**470**:3083–93[224]	

Paper	Reason for exclusion
Digas G, Thanner J, Anderberg C, Karrholm J. Fluoride-containing acrylic bone cement in total hip arthroplasty. Randomized evaluation of 97 stems using radiostereometry and dual-energy x-ray absorptiometry. *J Arthroplasty* 2005;**20**:784–92[393]	Exclude – effect of fluoride; two different cements
Dorr LD, Maheshwari AV, Long WT, Wan Z, Sirianni LE. Early pain relief and function after posterior minimally invasive and conventional total hip arthroplasty. A prospective, randomized, blinded study. *J Bone Joint Surg Am* 2007;**89**:1153–60[150]	Exclude – comparison of different surgical approaches
Dutka J, Sosin P, Libura M, Skowronek P. Total hip arthroplasty through a minimally invasive lateral approach – our experience and early results. *Ortop Traumatol Rehabil* 2007;**9** 39–45[151]	Exclude – comparison of different surgical approaches
Edwards SJ, Hamilton V, Nherera L, Arber M. Systematic review of the impact different metal femoral stems (MFSS) have on patient outcomes in total hip replacement (THR) due to osteoarthritis (OA). *Value Health* 2011;**14**:A245[394]	Exclude – abstract
Eingartner C, Piel S, Weise K. Results of a cemented straight titanium alloy femoral stem after mean follow-up of 13 years. *Eur J Orthop Surg Traumatol* 2007;**17**:587–93[395]	Exclude – single-arm cohort
Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster JY. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am* 2004;**86-A**:963–74[192]	Exclude – publication date before 2008
Fenandez-Lopez JC, Gossec L, Dougados M. Magnitude of the symptomatic at 3, 6 and 12 months after total articular replacement in hip and knee osteoarthritis: a systematic review and meta-analysis. *Arthritis Rheum* 2008;**58**:S245–6[396]	Exclude – abstract
Fick DP, Nivbrant B. Minimally invasive surgical approaches for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database Syst Rev* 2004;**2**:CD004798[397]	Exclude – systematic review concerned with the comparison of different surgical approaches
Flivik G. Fixation of the cemented acetabular component in hip arthroplasty. *Acta Orthop* 2005;**76**(Suppl. 76):3–30[153]	Exclude – comparison of different surgical approaches
Flivik G, Kristiansson I, Kesteris U, Ryd L. Is removal of subchondral bone plate advantageous in cemented cup fixation? A randomized RSA study. *Clin Orthop Relat Res* 2006;**448**:164–72[152]	Exclude – comparison of different surgical approaches
Foucher KC, Wimmer MA, Moisio KC, Hildebrand M, Berli MC, Walker MR, *et al.* Time course and extent of functional recovery during the first postoperative year after minimally invasive total hip arthroplasty with two different surgical approaches – a randomized controlled trial. *J Biomech* 2011;**44**:372–8[154]	Exclude – comparison of different surgical approaches
Freund KG, Houshian S, Riegels-Nielsen P. Occlusion and stability of two different femoral canal plugs in cemented hip arthroplasty. A prospective and randomized study, with a two year follow-up. *Hip Int* 2003;**13**:142–7[398]	Exclude – compared two different femoral plugs
Gallart X, Riba J, Garcia S, Combalia A, Esteban PL, Marmolejo C. Time saving during acrylic bone cement setting in femoral stem implantation of hip arthroplasty: a prospective, double-blind, randomised study. *Hip Int* 2005;**15**:143–8[399]	Exclude – comparison is concerned with two types of cement
Geerdink CH, Grimm B, Ramakrishnan R, Rondhuis J, Verburg AJ, Tonino AJ. Crosslinked polyethylene compared to conventional polyethylene in total hip replacement: pre-clinical evaluation, in-vitro testing and prospective clinical follow-up study. *Acta Orthop* 2006;**77**:719–25[193]	Exclude – publication date before 2008
Glyn-Jones S, Isaac S, Hauptfleisch J, McLardy-Smith P, Murray DW, Gill HS. Does highly cross-linked polyethylene wear less than conventional polyethylene in total hip arthroplasty? A double-blind, randomized, and controlled trial using roentgen stereophotogrammetric analysis. *J Arthroplasty* 2008;**23**:337–43[225]	Exclude – total number of patients < 100
Glyn-Jones S, McLardy-Smith P, Gill HS, Murray DW. The creep and wear of highly cross-linked polyethylene: a three-year randomised, controlled trial using radiostereometric analysis. *J Bone Joint Surg Br* 2008;**90**:556–61[83]	
Thomas GER, Simpson DJ, Mehmood S, Taylor A, McLardy-Smith P, Gill HS, *et al.* The seven-year wear of highly cross-linked polyethylene in total hip arthroplasty: a double-blind, randomized controlled trial using radiostereometric analysis. *J Bone Joint Surg Am* 2011;**93**:716–22[226]	

Paper	Reason for exclusion
Goosen JHM, Kollen BJ, Castelein RM, Kuipers BM, Verheyen CC. Minimally invasive versus classic procedures in total hip arthroplasty: a double-blind randomized controlled trial. *Clin Orthop Relat Res* 2011;**469**:200–8[155]	Exclude – comparison of different surgical approaches
Grant P, Aamodt A, Falch JA, Nordsletten L. Differences in stability and bone remodeling between a customized uncemented hydroxyapatite coated and a standard cemented femoral stem. A randomized study with use of radiostereometry and bone densitometry. *J Orthop Res* 2005;**23**:1280–5[194]	Exclude – publication date before 2008
Grubl A, Weissinger M, Brodner W, Gleiss A, Giurea A, Gruber M, *et al.* Serum aluminium and cobalt levels after ceramic-on-ceramic and metal-on-metal total hip replacement. *J Bone Joint Surg Br* 2006;**88**:1003–5[195]	Exclude – publication date before 2008
Hailer NP, Blaheta RA, Dahlstrand H, Stark A. Elevation of circulating HLA DR(+) CD8(+) T-cells and correlation with chromium and cobalt concentrations 6 years after metal-on-metal hip arthroplasty. *Acta Orthop* 2011;**82**:6–12[227]	Exclude – total number of patients < 100
Hallan G, Aamodt A, Furnes O, Skredderstuen A, Haugan K, Havelin LI. Palamed G compared with Palacos R with gentamicin in Charnley total hip replacement. *J Bone Joint Surg Br* 2006;**88**:1143–8[400]	Exclude – comparison of different cements
Hamadouche M, Baque F, Lefevre N, Kerboull M. Minimum 10-year survival of Kerboull cemented stems according to surface finish. *Clin Orthop Relat Res* 2008;**466**:332–9[401]	Exclude – comparison of different finishes
Hartl A, Schillinger M, Wanivenhaus A. Cemented versus cementless total hip arthroplasty for osteoarthrosis and other non-traumatic diseases. *Cochrane Database Syst Rev* 2004;**3**:CD004850[196]	Exclude – publication date before 2008
Haverkamp D, Klinkenbijl MN, Somford MP, Albers GHR, van der Vis HM. Obesity in total hip arthroplasty – does it really matter? A meta-analysis. *Acta Orthop* 2011;**82**:417–22[402]	Exclude – the effect of BMI on THR outcomes
Haverkamp D, Van Den Bekerom MPJ, Harmse I, Schafroth MU. One stage bilateral total hip arthroplasty, is it safe? A meta-analysis. *Hip Int* 2010;**20**:440–6[164]	Exclude – comparison of different surgical approaches, e.g. one- vs. two-stage approaches
HAYES, Inc. *Ganz Trochanteric Flip Osteotomy Approach to Hip Resurfacing for Treatment of Osteoarthritis*. Healthcare Technology Brief Publication. Lansdale, PA: HAYES, Inc.; 2012[403]	Unavailable because of cost
Hoebink E, Struijs Peter AA. Effects of different bearing surface materials on aseptic loosening of total hip arthroplasty in patients with osteoarthritis and other non-traumatic diseases of the hip. *Cochrane Database Syst Rev* 2008;**4**:CD007494[404]	Exclude – systematic review protocol for which no full systematic review is available
Honl M, Dierk O, Gauck C, Carrero V, Lampe F, Dries S, *et al.* Comparison of robotic-assisted and manual implantation of a primary total hip replacement. A prospective study. *J Bone Joint Surg Am* 2003;**85-A**:1470–8[156]	Exclude – comparison of different surgical approaches
Howie DW, McGee MA, Costi K, Graves SE. Metal-on-metal resurfacing versus total hip replacement – the value of a randomized clinical trial. *Orthop Clin North Am* 2005;**36**:195–201[197]	Exclude – publication date before 2008
Husby OS, Haugan K, Benum P, Foss OA. A prospective randomised radiostereometric analysis trial of SmartSet HV and Palacos R bone cements in primary total hip arthroplasty. *J Orthop Traumatol* 2010;**11**:29–35[405]	Exclude – comparison of different cements
Ise K, Kawanabe K, Tamura J, Akiyama H, Goto K, Nakamura T. Clinical results of the wear performance of cross-linked polyethylene in total hip arthroplasty. Prospective randomized trial. *J Arthroplasty* 2009;**24**:1216–20[406]	Exclude – secondary OA
ISPOR 14th Annual European Congress Research Abstracts. *Value Health* 2011;**14**:A233–A574[407]	Exclude – abstract
Jager M, Begg MJ, Ready J, Bittersohl B, Millis M, Krauspe R, Thornhill TS. Primary total hip replacement in childhood, adolescence and young patients: quality and outcome of clinical studies. *Technol Health Care* 2008;**16**:195–214[408]	Exclude – non-systematic review, population not relevant
Jandric S, Jovicic Z, Novakovic S. Differences in pain between women and men in patients with total hip arthroplasty. *Eur J Pain* 2009;**13**(Suppl. 1):S133[409]	Exclude – abstract

Paper	Reason for exclusion
Jandric S, Manojlovic S. Quality of life of men and women with osteoarthritis of the hip and arthroplasty assessment by WOMAC questionnaire. *Am J Phys Med Rehabil* 2009;**88**:328–35[410]	Exclude – non-RCT
Jensen C, Aagaard P, Overgaard S. Recovery in mechanical muscle strength following resurfacing vs standard total hip arthroplasty – a randomised clinical trial. *Osteoarthritis Cartilage* 2011;**19**:1108–16[228]	Exclude – total number of patients < 100
Jensen C, Aagaard P, Overgaard S. Recovery in horizontal gait after hip resurfacing vs. total hip arthroplasty at 6-month follow-up – a randomized clinical trial. *Osteoarthritis Cartilage* 2012;**20**(Suppl. 1):S99[411]	Exclude – abstract
Jolles BM, Bogoch ER. Surgical approach for total hip arthroplasty: direct lateral or posterior? *J Rheumatol* 2004;**31**:1790–6[158]	Exclude – comparison of different surgical approaches
Jolles BM, Bogoch ER. Posterior versus lateral surgical approach for total hip arthroplasty in adults with osteoarthritis. *Cochrane Database Systc Rev* 2006;**3**:CD003828[157]	Exclude – comparison of different surgical approaches
Jolles BM, Grzesiak A, Eudier A, Dejnabadi H, Voracek C, Pichonnaz C, *et al.* A randomised controlled clinical trial and gait analysis of fixed- and mobile-bearing total knee replacements with a five-year follow-up. *J Bone Joint Surg Br* 2012;**94B**:648–55[412]	Exclude – knee
Jolles BM, Michel J, Burnand B, Leyvraz P. Surgical treatment for advanced stage of avascular necrosis of the femoral head in adults. *Cochrane Database Syst Rev* 2006;**3**:CD006079[413]	Exclude – population not relevant
Jones CA, Pohar S. Health-related quality of life after total joint arthroplasty: a scoping review. *Clin Geriatr Med* 2012;**28**:395–429[294]	Exclude – not systematic review; no RCTs included
Kadar T, Hallan G, Aamodt A, Indrekvam K, Badawy M, Havelin LI, *et al.* A randomized study on migration of the Spectron EF and the Charnley flanged 40 cemented femoral components using radiostereometric analysis at 2 years. *Acta Orthop* 2011;**82**:538–44[414]	Exclude – comparison of different finishes
Karachalios T, Tsatsaronis C, Efraimis G, Papadelis P, Lyritis G, Diakoumopoulos G. The long-term clinical relevance of calcar atrophy caused by stress shielding in total hip arthroplasty: a 10-year, prospective, randomized study. *J Arthroplasty* 2004;**19**:469–75[198]	Exclude – publication date before 2008
Karas S. Outcomes of Birmingham hip resurfacing: a systematic review. *Asian J Sports Med* 2012;**3**:1–7[415]	Exclude – this systematic review included no RCTs
Karrholm J, Anderberg C, Snorrason F, Thanner J, Langeland N, Malchau H, *et al.* Evaluation of a femoral stem with reduced stiffness. A randomized study with use of radiostereometry and bone densitometry. *J Bone Joint Surg Am* 2002;**84-A**:1651–8[199]	Exclude – publication date before 2008
Kenanidis EI, Potoupnis ME, Papavasiliou KA, Sayegh FE, Kapetanos GA. Re: Prospective randomized study of two surgical approaches for total hip arthroplasty. *J Arthroplasty* 2011;**26**:821[416]	Exclude – letter to editor
Khan RJ, Maor D, Hofmann M, Haebich S. A comparison of a less invasive piriformis-sparing approach versus the standard posterior approach to the hip: a randomised controlled trial. *J Bone Joint Surg Br* 2012;**94**:43–50[159]	Exclude – comparison of different surgical approaches
Kim S, Losina E, Solomon DH, Wright J, Katz JN. Effectiveness of clinical pathways for total knee and total hip arthroplasty: literature review. *J Arthroplasty* 2003;**18**:69–74[417]	Exclude – irrelevant – effectiveness of clinical pathways
Kim YH. Comparison of polyethylene wear associated with cobalt-chromium and zirconia heads after total hip replacement: a prospective, randomized study. *J Bone Joint Surg Am* 2005;**87**:1769–76[201]	Exclude – publication date before 2008
Kim Y-H. Comparison of primary total hip arthroplasties performed with a minimally invasive technique or a standard technique. A prospective and randomized study. *J Arthroplasty* 2006;**21**:1092–8[160]	Exclude – comparison of different surgical approaches
Kim YH, Kim JS, Joo JH, Park JW. Is hydroxyapatite coating necessary to improve survivorship of porous-coated titanium femoral stem? *J Arthroplasty* 2012;**27**:559–63[418]	Exclude – comparison of different coatings

Paper	Reason for exclusion
Kim YH, Oh JH. A comparison of a conventional versus a short, anatomical metaphyseal-fitting cementless femoral stem in the treatment of patients with a fracture of the femoral neck. *J Bone Joint Surg Br* 2012;**94**:774–81[419]	Exclude – population with fractures
Kim YH, Oh SW, Kim JS. Prevalence of fat embolism following bilateral simultaneous and unilateral total hip arthroplasty performed with or without cement: a prospective, randomized clinical study. *J Bone Joint Surg Am* 2002;**84**:1372–9[420]	Exclude – comparison and no outcome
Kim Y-H, Yoon S-H, Kim J-S. Changes in the bone mineral density in the acetabulum and proximal femur after cementless total hip replacement. *J Bone Joint Surg Br* 2007;**89**:174–9[200]	Exclude – publication date before 2008
Kraay MJ, Thomas RD, Rimnac CM, Fitzgerald SJ, Goldberg VM. Zirconia versus Co-Cr femoral heads in total hip arthroplasty: early assessment of wear. *Clin Orthop Relat Res* 2006;**453**:86–90[202]	Exclude – publication date before 2008
Krych AJ, Pagnano MW, Coleman WK, Meneghini RM, Kaufman K. No strength or gait benefit of two-incision THA: a brief followup at 1 year. *Clin Orthop Relat Res* 2011;**469**:1110–18[185]	Exclude – comparison of different operative techniques
Krych AJ, Pagnano MW, Wood KC, Meneghini RM, Kaufmann K. No benefit of the two-incision THA over mini-posterior THA: a pilot study of strength and gait. *Clin Orthop Relat Res* 2010;**468**:565–70[186]	Exclude – comparison of different operative techniques
Lane NE. Osteoarthritis of the hip. *N Engl J Med* 2007;**357**:1413–21[421]	Exclude – case report – narrative review
Laursen MB, Nielsen PT, Soballe K. Bone remodelling around HA-coated acetabular cups: a DEXA study with a 3-year follow-up in a randomised trial. *Int Orthop* 2007;**31**:199–204[422]	Exclude – comparison of different coatings
Lavigne M, Masse V, Girard J, Roy AG, Vendittoli PA. [Return to sport after hip resurfacing or total hip arthroplasty: a randomized study]. *Rev Chir Orthop Reparatrice Appar Mot* 2008;**94**:361–7[423]	Exclude – abstract only in English
Lavigne M, Therrien M, Nantel J, Roy A, Prince F, Vendittoli PA. The John Charnley Award: the functional outcome of hip resurfacing and large-head THA is the same: a randomized, double-blind study. *Clin Orthop Relat Res* 2010;**468**:326–36[229]	Exclude – total number of patients < 100
Lewis PM, Al-Belooshi A, Olsen M, Schemitch EH, Waddell JP. Prospective randomized trial comparing alumina ceramic-on-ceramic with ceramic-on-conventional polyethylene bearings in total hip arthroplasty. *J Arthroplasty* 2010;**25**:392–7[230]	Exclude – total number of patients < 100
Li N, Deng Y, Chen L. Comparison of complications in single-incision minimally invasive THA and conventional THA. *Orthopedics* 2012;**35**:e1152–8[166]	Exclude – comparison of different operative approaches
Lombardi J, Mallory TH, Cuckler JM, Williams J, Berend KR, Smith TM. Mid-term results of a polyethylene-free metal-on-metal articulation. *J Arthroplasty* 2004;**19**(7 Suppl. 2):42–7[203]	Exclude – publication date before 2008
MacDonald SJ, McCalden RW, Chess DG, Bourne RB, Rorabeck CH, Cleland A, *et al.* Metal-on-metal versus polyethylene in hip arthroplasty: a randomized clinical trial. *Clin Orthop Relat Res* 2003;**406**:282–96[204]	Exclude – publication date before 2008
Mallmin H, Wolf O, Larsson S, Milbrink J, Mattsson P. Body composition and BMD after total hip arthroplasty. A randomised clinical trial of two different postoperative regimes with 5 years of follow-up. *Bone* 2011;**48**(Suppl. 2):S267[424]	Exclude – abstract
Malviya A, Ramaskandhan JR, Bowman R, Hashmi M, Holland JP, Kometa S, *et al.* What advantage is there to be gained using large modular metal-on-metal bearings in routine primary hip replacement? A preliminary report of a prospective randomised controlled trial. *J Bone Joint Surg Br* 2011;**93**:1602–9[231]	Exclude – total number of patients < 100
Markmiller M, Weiss T, Kreuz P, Ruter A, Konrad G. Partial weightbearing is not necessary after cementless total hip arthroplasty: a two-year prospective randomized study on 100 patients. *Int Orthop* 2011;**35**:1139–43[425]	Exclude – effect of weight bearing
Martin R, Clayson PE, Troussel S, Fraser BP, Docquier PL. Anterolateral minimally invasive total hip arthroplasty: a prospective randomized controlled study with a follow-up of 1 year. *J Arthroplasty* 2011;**26**:1362–72[167]	Exclude – comparison of different operative approaches

Paper	Reason for exclusion
Mayr E, Nogler M, Benedetti MG, Kessler O, Reinthaler A, Krismer M, *et al.* A prospective randomized assessment of earlier functional recovery in THA patients treated by minimally invasive direct anterior approach: a gait analysis study. *Clin Biomech* 2009;**24**:812–18[168]	Exclude – comparison of different operative approaches
Mazoochian F, Weber P, Schramm S, Utzschneider S, Fottner A, Jansson V. Minimally invasive total hip arthroplasty: a randomized controlled prospective trial. *Arch Orthop Trauma Surg* 2009;**129**:1633–9[169]	Exclude – comparison of different operative approaches
McCalden RW, Charron KD, Yuan X, Bourne RB, Naudie DD, MacDonald SJ. Randomised controlled trial comparing early migration of two collarless polished cemented stems using radiostereometric analysis. *J Bone Joint Surg Br* 2010;**92B**:935–40[232]	Exclude – total number of patients < 100
McCombe P, Williams SA. A comparison of polyethylene wear rates between cemented and cementless cups. A prospective, randomised trial. *J Bone Joint Surg Br* 2004;**86**:344–9[205]	Exclude – publication date before 2008
Medical Advisory Secretariat. Metal-on-metal total hip resurfacing arthroplasty: an evidence-based analysis. *Ont Health Technol Assess Ser* 2006;**6**:1–57[206]	Exclude – publication date before 2008
Meek RD, Allan DB. Cemented versus cementless surgical approach for total hip arthroplasty revision. *Cochrane Database Syst Rev* 2005;**2**:CD005322[207]	Exclude – publication date before 2008
Montin L, Leino-Kilpi H, Suominen T, Lepisto J. A systematic review of empirical studies between 1966 and 2005 of patient outcomes of total hip arthroplasty and related factors. *J Clin Nurs* 2008;**17**:40–5[426]	Exclude – no RCTs; not a standard systematic review
Morshed S, Bozic KJ, Ries MD, Malchau H, Colford JM Jr. Comparison of cemented and uncemented fixation in total hip replacement: a meta-analysis. *Acta Orthop* 2007;**78**:315–26[39]	Exclude – publication date before 2008
Moskal JT, Capps SG. Acetabular component positioning in total hip arthroplasty: an evidence-based analysis. *J Arthroplasty* 2011;**26**:1432–7[427]	Exclude – comparison of different cup positioning
Mouilhade F, Matsoukis J, Oger P, Mandereau C, Brzakala V, Dujardin F. Component positioning in primary total hip replacement: a prospective comparative study of two anterolateral approaches, minimally invasive versus gluteus medius hemimyotomy. *Orthop Traumatol Surg Res* 2011;**97**:14–21[428]	Exclude – comparison of different approaches
Muller M, Schwachmeyer V, Tohtz S, Tohtz S, Duda GN, Perka C, *et al.* The direct lateral approach: impact on gait patterns, foot progression angle and pain in comparison with a minimally invasive anterolateral approach. *Arch Orthop Trauma Surg* 2012;**132**:725–31[170]	Exclude – comparison of different operative approaches
Muller M, Tohtz S, Springer I, Dewey M, Perka C. Randomized controlled trial of abductor muscle damage in relation to the surgical approach for primary total hip replacement: minimally invasive anterolateral versus modified direct lateral approach. *Arch Orthop Trauma Surg* 2011;**131**:179–89[171]	Exclude – comparison of different operative approaches
Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty? A systematic review. *Clin Orthop Relat Res* 2010;**468**:1891–904[429]	Exclude – no RCTs; not a standard systematic review
Nakamura N, Sugano N, Nishii T, Kakimoto A, Miki H. A comparison between robotic-assisted and manual implantation of cementless total hip arthroplasty. *Clin Orthop Relat Res* 2010;**468**:1072–81[172]	Exclude – comparison of different operative approaches
Nantel J, Termoz N, Vendittoli PA, Lavigne M, Prince F. Gait patterns after total hip arthroplasty and surface replacement arthroplasty. *Arch Phys Med Rehabil* 2009;**90**:463–9[430]	Exclude – non-randomised study
Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RG. Good diagnostic performance of early migration as a predictor of late aseptic loosening of acetabular cups: results from ten years of follow-up with Roentgen stereophotogrammetric analysis (RSA). *J Bone Joint Surg Am* 2012;**94**:874–80[431]	Exclude – single-arm cohort study
Nieuwenhuijse MJ, Valstar ER, Kaptein BL, Nelissen RG. The Exeter femoral stem continues to migrate during its first decade after implantation: 10–12 years of follow-up with radiostereometric analysis (RSA). *Acta Orthop* 2012;**83**:129–34[432]	Exclude – single-arm cohort study

Paper	Reason for exclusion
Nikolaou VS, Edwards MR, Bogoch E, Schemitsch EH, Waddell JP. A prospective randomised controlled trial comparing three alternative bearing surfaces in primary total hip replacement. *J Bone Joint Surg Br* 2012;**94**:459–65[233]	Exclude – total number of patients < 100
Nygaard M, Elling F, Bastholm L, Soballe K, Borgwardt A. No difference in early cellular response of the pseudo-synovial membrane after total hip arthroplasty: comparison of 3 combinations of bearing materials. *Acta Orthop* 2006;**77**:402–12[433]	Exclude – observational study
Nygaard M, Zerahn B, Bruce C, Soballe K, Borgwardt A. Early periprosthetic femoral bone remodelling using different bearing material combinations in total hip arthroplasties: a prospective randomised study. *Eur Cells Mater* 2004;**8**:65–72[434]	Exclude – no outcome
Nysted M, Benum P, Klaksvik J, Foss O, Aamodt A. Periprosthetic bone loss after insertion of an uncemented, customized femoral stem and an uncemented anatomical stem. A randomized DXA study with 5-year follow-up. *Acta Orthop* 2011;**82**:410–16[234]	Exclude – total number of patients < 100
Ogonda L, Wilson R, Archbold P, Lawlor M, Humphreys P, O'Brien S, *et al.* A minimal-incision technique in total hip arthroplasty does not improve early postoperative outcomes. A prospective, randomized, controlled trial. *J Bone Joint Surg Am* 2005;**87**:701–10[173]	Exclude – comparison of different operative approaches
Pabinger C, Kroner A, Lange A, Eyb R. Cemented titanium stems show high migration: transprosthetic drainage system has no advantage over third-generation cementation technique. *Arch Orthop Trauma Surg* 2004;**124**:489–94[435]	Exclude – comparison of different cementation techniques
Pakvis D, Luites J, Hellemondt G, Spruit M. A cementless, elastic press-fit socket with and without screws. *Acta Orthop* 2012;**83**:481–7[174]	Exclude – comparison of different operative approaches
Palm L, Jacobsson S-A, Ivarsson I. Hydroxyapatite coating improves 8- to 10-year performance of the Link RS cementless femoral stem. *J Arthroplasty* 2002;**17**:172–5[436]	Exclude – comparison of different coating techniques
Palm L, Olofsson J, Astrom SE, Ivarsson I. No difference in migration or wear between cemented low-profile cups and standard cups: a randomized radiostereographic study of 53 patients over 3 years. *Acta Orthop* 2007;**78**:479–84[209]	Exclude – publication date before 2008
Parvizi J, Tarity TD, Sheikh E, Sharkey PF, Hozack WJ, Rothman RH. Bilateral total hip arthroplasty: one-stage versus two-stage procedures. *Clin Orthop Relat Res* 2006;**453**:137–41[175]	Exclude – comparison of different operative approaches
Patel D, Parvizi J, Sharkey PF. Alternative bearing surface options for revision total hip arthroplasty. *Instruct Course Lectures* 2011;**60**: 257–67[437]	Exclude – review chapter – does not meet criteria for a systematic review
Petersen MK, Andersen NT, Mogensen P, Voight M, Soballe K. Gait analysis after total hip replacement with hip resurfacing implant or Mallory-head Exeter prosthesis: a randomised controlled trial. *Int Orthop* 2011;**35**:667–74[235]	Exclude – total number of patients < 100
Petersen MK, Andersen NT, Soballe K. Self-reported functional outcome after primary total hip replacement treated with two different perioperative regimes: a follow-up study involving 61 patients. *Acta Orthop* 2008;**79**:160–7[438]	Exclude – effect of preoperative regime
Pitto RP, Blanquaert D, Hohmann D. Alternative bearing surfaces in total hip arthroplasty: zirconia–alumina pairing. Contribution or caveat? *Acta Orthop Belg* 2002;**68**:242–50[210]	Exclude – publication date before 2008
Pitto RP, Hamer H, Fabiani R, Radespiel-Troeger M, Koessler M. Prophylaxis against fat and bone-marrow embolism during total hip arthroplasty reduces the incidence of postoperative deep-vein thrombosis: a controlled, randomized clinical trial. *J Bone Joint Surg Am* 2002;**84**:39–48[439]	Exclude – effect of bone vacuum technique
Pitto RP, Schikora N, Willmann G, Graef B, Schmidt R. Radiostereoanalysis of press-fit cups with alumina liner – a randomized clinical trial. *Bioceramics* 2003;**15**:817–21[211]	Exclude – publication date before 2008
Pivec R, Johnson AJ, Mont MA. Results of total hip arthroplasty in patients who have rapidly progressive hip disease: a systematic review of the literature. *Expert Rev Med Dev* 2012;**9**:257–62[440]	Exclude – no RCT included
Pospischill M, Kranzl A, Attwenger B, Knahr K. Minimally invasive compared with traditional transgluteal approach for total hip arthroplasty: a comparative gait analysis. *J Bone Joint Surg Am* 2010;**92**:328–37[176]	Exclude – comparison of different operative approaches

Paper	Reason for exclusion
Prudhon JL. Dual-mobility cup and cemented femoral component: 6 year follow-up results. *Hip Int* 2011;**21**:713–17[441]	Exclude – non-randomised prospective study
Rasanen P, Paavolainen P, Sintonen H, Koivisto AM, Blom M, Ryynanen OP, *et al.* Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthop* 2007;**78**:108–15[442]	Exclude – non-randomised study
Rasquinha VJ, Ranawat CS, Dua V, Ranawat AS, Rodriguez JA. A prospective, randomized, double-blind study of smooth versus rough stems using cement fixation: minimum 5-year follow-up. *J Arthroplasty* 2004;**19**(7 Suppl. 2):2–9[443]	Exclude – comparison of different surface finishes
Ratko TA, Aronson N, Ziegler KM, Bonnell CJ. *Metal-on-Metal Total Hip Resurfacing*. Chicago, IL: Blue Cross and Blue Shield Association, Technology Evaluation Center. Assessment Program; 2007[444]	Exclude – this systematic review included only one RCT, which is already included in the present review
Reininga IH, Wagenmakers R, van den Akker-Scheek I, Stant AD, Groothoff JW, Bulstra SK, *et al.* Effectiveness of computer-navigated minimally invasive total hip surgery compared to conventional total hip arthroplasty: design of a randomized controlled trial. *BMC Musculoskelet Disord* 2007;**8**:4[177]	Exclude – comparison of different operative approaches
Reininga IHF, Zijlstra W, Wagenmakers R, Boerboom AL, Huijbers BP, Groothoff JW, *et al.* Minimally invasive and computer-navigated total hip arthroplasty: a qualitative and systematic review of the literature. *BMC Musculoskelet Disord* 2010;**11**:92[178]	Exclude – comparison of different operative approaches
Renkawitz T, Santori FS, Grifka J, Valverde C, Morlock MM, Learmonth ID. A new short uncemented, proximally fixed anatomic femoral implant with a prominent lateral flare: design rationals and study design of an international clinical trial. *BMC Musculoskelet Disord* 2008;**9**:147[445]	Exclude – protocol with no publication of full trial
Restrepo C, Parvizi J, Pour AE, Hozack WJ. Prospective randomized study of two surgical approaches for total hip arthroplasty. *J Arthroplasty* 2010;**25**:671–9[179]	Exclude – comparison of different operative approaches
Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: a systematic review. *Arthritis Rheum Arthritis Care Res* 2008;**59**:876–83[446]	Exclude – non-intervention study
Rohrl SM, Nivbrant B, Strom H, Nilsson KG. Effect of augmented cup fixation on stability, wear, and osteolysis: a 5-year follow-up of total hip arthroplasty with RSA. *J Arthroplasty* 2004;**19**:962–71[447]	Exclude – comparison of different types of cup fixation
Saito S, Tokuhashi Y, Ishii T, Mori S, Hosaka K, Taniguchi S. One- versus two-stage bilateral total hip arthroplasty. *Orthopedics* 2010;**33**[180]	Exclude – comparison of different operative approaches
Santaguida PL, Hawker GA, Hudak PL, Glazier R, Mahomed NN, Kreder HJ, *et al.* Patient characteristics affecting the prognosis of total hip and knee joint arthroplasty: a systematic review. *Can J Surg* 2008;**51**:428–36[448]	Exclude – THR/RS not compared; prognostic factors only; no RCTs included
Sariali E, Mauprivez R, Khiami F, Pascal-Mousselard H, Catonne Y. Accuracy of the preoperative planning for cementless total hip arthroplasty. A randomised comparison between three-dimensional computerised planning and conventional templating. *Orthop Traumatol Surg Res* 2012;**98**:151–8[181]	Exclude – comparison of different operative approaches
Schauss SM, Hinz M, Mayr E, Bach CM, Krismer M, Fischer M. Inferior stability of a biodegradable cement plug. 122 total hip replacements randomized to degradable or non-degradable cement restrictor. *Arch Orthop Trauma Surg* 2006;**126**:324–9[449]	Exclude – comparison of different cementing techniques
Schmidutz F, Dull T, Voges O, Grupp T, Muller P, Jansson V. Secondary cement injection technique reduces pulmonary embolism in total hip arthroplasty. *Int Orthop* 2012;**36**:1575–81[450]	Exclude – effect of cement injection
Schouten R, Malone AA, Tiffen C, Frampton CM, Hooper G. A prospective, randomised controlled trial comparing ceramic-on-metal and metal-on-metal bearing surfaces in total hip replacement. *J Bone Joint Surg Br* 2012;**94**:1462–7[236]	Exclude – Total number of patients < 100
Scott D. Osteoarthritis of the hip. *Clin Evid Handbook* December 2009:398–9[451]	Exclude – not a systematic review; no comparative results reported; includes THR and replacement. This is an updated version from 2009, based on 2007 search

Paper	Reason for exclusion
Seyler TM, Bonutti PM, Shen J, Naughton M, Kester M. Use of an alumina-on-alumina bearing system in total hip arthroplasty for osteonecrosis of the hip. *J Bone Joint Surg Am* 2006;**88**:116–25[452]	Exclude – osteonecrosis patients
Sharma V, Morgan PM, Cheng EY. Factors influencing early rehabilitation after THA: a systematic review. *Clin Orthop Relat Res* 2009;**467**:1400–11[453]	Exclude – prognostic factors in postoperative patients
Shetty V, Shitole B, Shetty G, Thakur H, Bhandari M. Optimal bearing surfaces for total hip replacement in the young patient: a meta-analysis. *Int Orthop* 2011;**35**:1281–7[454]	Exclude – no RCTs included
Singh JA. Epidemiology of knee and hip arthroplasty: a systematic review. *Open Orthop J* 2011;**5**:80–5[455]	Exclude – no RCTs included
Singh JA, Kundukulam J, Riddle DL, Strand V, Tugwell P. Early postoperative mortality following joint arthroplasty: a systematic review. *J Rheumatol* 2011;**38**:1507–13[456]	Exclude – no RCTs included
Sluimer JC, Hoefnagels NH, Emans PJ, Kuijer R, Geesink RG. Comparison of two hydroxyapatite-coated femoral stems: clinical, functional, and bone densitometry evaluation of patients randomized to a regular or modified hydroxyapatite-coated stem aimed at proximal fixation. *J Arthroplasty* 2006;**21**:344–52[457]	Exclude – comparison of different cementing techniques
Smolders JM, Hol A, Rijnders T, van Susante JL. Changes in bone mineral density in the proximal femur after hip resurfacing and uncemented total hip replacement: a prospective randomised controlled study. *J Bone Joint Surg Br* 2010;**92**:1509–14[238]	Exclude – total number of patients < 100
Smolders JM, Hol A, Rijnberg WJ, van Susante JL. Metal ion levels and functional results after either resurfacing hip arthroplasty or conventional metal-on-metal hip arthroplasty. *Acta Orthop* 2011;**82**:559–66[237]	Exclude – total number of patients < 100
Sonny BB, Aleto TJ, Garino JP, Toni A, Hendricks KJ. Ceramic-on-ceramic versus ceramic-on-polyethylene bearings in total hip arthroplasty: results of a multicenter prospective randomized study and update of modern ceramic total hip trials in the United States. *Hip Int* 2005;**15**:129–35[212]	Exclude – publication date before 2008
Speranza A, Iorio R, Ferretti M, D'Arrigo C, Ferretti A. A lateral minimal-incision technique in total hip replacement: a prospective, randomized, controlled trial. *Hip Int* 2007;**17**:4–8[161]	Exclude – comparison of different surgical approaches
Stanat SJC, Capozzi JD. Squeaking in third- and fourth-generation ceramic-on-ceramic total hip arthroplasty. Meta-analysis and systematic review. *J Arthroplasty* 2012;**27**:445–53[458]	Exclude – prognostic factors for squeaking
Stilling M, Rahbek O, Soballe K. Inferior survival of hydroxyapatite versus titanium-coated cups at 15 years. *Clin Orthop Relat Res* 2009;**467**:2872–9[459]	Exclude – comparison of different cup coatings
Strom H, Kolstad K, Mallmin H, Sahlstedt B, Milbrink J. Comparison of the uncemented Cone and the cemented Bimetric hip prosthesis in young patients with osteoarthritis: an RSA, clinical and radiographic study. *Acta Orthop* 2006;**77**:71–8[213]	Exclude – publication date before 2008
Suda AJ, Knahr K. Early results with the cementless Variall hip system. *Expert Rev Med Devices* 2009;**6**:21–5[460]	Exclude – lack of statistical comparative data between different THR treatments
Tanavalee A, Jaruwannapong S, Yuktanandana P, Itiravivong P. Early outcomes following minimally invasive total hip arthroplasty using a two-incision approach versus a mini-posterior approach. *Hip Int* 2006;**16**(Suppl. 4):17–22[162]	Exclude – comparison of different surgical approaches
Tang Z. *Minimally Invasive Total Hip Replacement*. No. 4. Ottawa, ON: Canadian Coordinating Office for Health Technology Assessment (CCOHTA); 2004[163]	Exclude – comparison of different surgical approaches
Tarasevicius S, Robertsson O, Wingstrand H. Posterior soft tissue repair in total hip arthroplasty: a randomized controlled trial. *Orthopedics* 2010;**33**:871[461]	Exclude – effect of soft tissue repair after THR
ten Broeke RHM, Hendrickx RPM, Leffers P, Jutten LMC, Geesink RGT. Randomised trial comparing bone remodelling around two uncemented stems using modified Gruen zones. *Hip Int* 2012;**22**:41–9[462]	Exclude – comparison of different uncemented stems
Thien TM, Thanner J, Karrholm J. Randomized comparison between 3 surface treatments of a single anteverted stem design: 84 hips followed for 5 years. *J Arthroplasty* 2010;**25**:437–44[463]	Exclude – comparison of different stem coatings

Paper	Reason for exclusion
Thien TM, Thanner J, Karrholm J. Fixation and bone remodeling around a low-modulus stem. Seven-year follow-up of a randomized study with use of radiostereometry and dual-energy X-ray absorptiometer. *J Arthroplasty* 2012;**27**:134–42[464]	Exclude – comparison of different stem coatings
Thomas W, Tafuro L, Thomas S. Osteoinductive gel in cementless hip joint replacement: a randomized prospective study. *Curr Orthop Pract* 2009;**20**:655–9[465]	Exclude – effect of gel
Timperley AJ, Whitehouse SL, Hourigan PG. The influence of a suction device on fixation of a cemented cup using RSA. *Clin Orthop Relat Res* 2009;**467**:792–8[466]	Exclude – effect of suction device
Ullmark G, Sorensen J, Nilsson O. Analysis of bone formation on porous and calcium phosphate-coated acetabular cups: a randomised clinical [18F]fluoride PET study. *Hip Int* 2012;**22**:172–8[467]	Exclude – comparison of different stem coatings
Vail TP, Goetz D, Tanzer M, Fisher DA, Mohler CG, Callaghan JJ. A prospective randomized trial of cemented femoral components with polished versus grit-blasted surface finish and identical stem geometry. *J Arthroplasty* 2003;**18**:95–102[468]	Exclude – comparison of different stem coatings
Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC. A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease. *Health Technol Assess* 2002;**6**(15)[19]	Exclude – publication date before 2008
Van der Wal BC, Rahmy AI, Grimm B, Blake GM, Heyligers IC, Tonino AJ. The influence of implant design on periprosthetic bone remodelling of two types of uncemented HA-coated hip stems. A two-year follow-up study using DEXA. *Hip Int* 2006;**16**:8–17[469]	Exclude – comparison of different stem coatings
Van Der Weegen W, Hoekstra HJ, Sijbesma T, Bos E, Schemitsch EH, Poolman RW. Survival of metal-on-metal hip resurfacing arthroplasty. A systematic review of the literature. *J Bone Joint Surg Br* 2011;**93**:298–306[470]	Exclude – this systematic review includes only one RCT but no comparative results are reported between RS and THR
van Gerwen, Shaerf DA, Veen RM. Hip resurfacing arthroplasty: a systematic review of functional outcome. *Acta Orthop* 2010;**81**:680–3[471]	Exclude – this systematic review includes only one RCT but no comparative results are reported between RS and THR
Veldstra R, van Dongen A, Kraaneveld EC. Comparing alumina-reduced and conventional surface grit-blasted acetabular cups in primary THA: early results from a randomised clinical trial. *Hip Int* 2012;**22**:296–301[472]	Exclude – comparison of different stem coatings
Vendittoli PA, Ganapathi M, Duval N, Lavoie P, Roy A, Lavigne M. Randomised controlled trial comparing two methods of acetabular cup positioning during total hip arthroplasty. *Hip Int* 2007;**17**:137–42[473]	Exclude – comparison of different cup positioning methods
Vissers MM, Bussmann JB, Verhaar JA, Arends LR, Furlan AD, Reijman M. Recovery of physical functioning after total hip arthroplasty: systematic review and meta-analysis of the literature. *Phys Ther* 2011;**91**:615–29[474]	Exclude – this systematic review includes only two relevant RCTs, both of which have been included in the present review
Vissers MM, Reijman M, Bussmann HB, Arends LR, Verhaar JA. Recovery of physical functioning after total hip arthroplasty: a systematic review of the literature. *Osteoarthritis Cartilage* 2009;**17**(Suppl. 1):S287–8[475]	Exclude – abstract
von Schewelov T, Sanzen L, Onsten I, Carlsson A, Besjakov J. Total hip replacement with a zirconium oxide ceramic femoral head: a randomised roentgen stereophotogrammetric study. *J Bone Joint Surg Br* 2005;**87**:1631–5[214]	Exclude – publication date before 2008
Wassilew GI, Perka C, Janz V, Konig C, Asbach P, Hasart O. Use of an ultrasound-based navigation system for an accurate acetabular positioning in total hip arthroplasty. A prospective, randomized, controlled study. *J Arthroplasty* 2012;**27**:687–94[182]	Exclude – comparison of different operative approaches
Weissinger M, Grubl A, Poll G. Serum-cobalt levels with metal-on-metal bearings in the cement-free total hip arthroplasty results covering two years; prospective study. *Acta Chir Orthop Traumatol Cech* 2011;**78**:410–15[239]	Exclude – total number of patients < 100
Witzleb WC, Stephan L, Krummenauer F, Neuke A, Gunther KP. Short-term outcome after posterior versus lateral surgical approach for total hip arthroplasty – a randomized clinical trial. *Eur J Med Res* 2009;**14**:256–63[183]	Exclude – comparison of different operative approaches

Paper	Reason for exclusion
Wyness L, Vale L, McCormack K, Grant A, Brazzelli M. The effectiveness of metal on metal hip resurfacing: a systematic review of the available evidence published before 2002. *BMC Health Serv Res* 2004;**4**:39[215]	Exclude – publication date before 2008
Yamauchi Y, Jinno T, Koga D, Asou Y, Morita S, Okawa A. Comparison of different distal designs of femoral components and their effects on bone remodeling in 1-stage bilateral total hip arthroplasty. *J Arthroplasty* 2012;**27**:1538–43[476]	Exclude – comparison of different stem coatings
Yang B, Li H, He X, Wang G, Xu S. Minimally invasive surgical approaches and traditional total hip arthroplasty: a meta-analysis of radiological and complications outcomes. *PLOS ONE* 2012;**7**:e37947[184]	Exclude – comparison of different operative approaches
Zagra L, Bianchi L, Licari V, Champlon C, Giacometti CR. Gait analysis of THA with different head diameters: a prospective randomized study. *J Orthop Traumatol* 2011;**12**(1 Suppl.):S149–50[477]	Exclude – abstract
Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, Part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. *Osteoarthritis Cartilage* 2007;**15**:981–1000[216]	Exclude – publication date before 2008
Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. *Osteoarthritis Cartilage* 2008;**16**:137–62[12]	Exclude – a guideline
Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, *et al* OARSI recommendations for the management of hip and knee osteoarthritis Part III: changes in evidence following systematic cumulative update of research published through January 2009. *Osteoarthritis Cartilage* 2010;**18**:476–99[478]	Exclude – says virtually nothing about THR
Zhang Y, Yang T-T, Zhou Y, Ma B-A. Comparison of postoperative curative effect and the possible survival rate of prosthesis following cemented and cementless total hip replacement. *Chin J Clin Rehabil* 2006;**10**:10[479]	Exclude – observational cohort study
Zhou ZK, Li MG, Borlin N, Wood DJ, Nivbrant B. No increased migration in cups with ceramic-on-ceramic bearing: an RSA study. *Clin Orthop Relat Res* 2006;**448**:39–45[217]	Exclude – publication date before 2008
Zijlstra WP, Bos N, van Raaij JJ. Large head metal-on-metal cementless total hip arthroplasty versus 28 mm metal-on-polyethylene cementless total hip arthroplasty: design of a randomized controlled trial. *BMC Musculoskelet Disord* 2008;**9**:136[241]	Exclude – protocol leading to study published later in which total number of patients is < 100
Zijlstra WP, van den Akker-Scheek I, Zee MJ, van Raay JJ. No clinical difference between large metal-on-metal total hip arthroplasty and 28-mm-head total hip arthroplasty? *Int Orthop* 2011;**35**:1771–6[240]	Exclude – total number of patients < 100
Zwartele RE, Witjes S, Doets HC, Stijnen T, Poll RG. Cementless total hip arthroplasty in rheumatoid arthritis: a systematic review of the literature. *Arch Orthop Trauma Surg* 2012;**132**:535–46[480]	Exclude – no RCTs included in this systematic review

Appendix 7 Clinical trials and health services research identified

Taken from the ClinicalTrials.gov, Current Controlled Trials, UKCRN Portfolio and HSRProj Databases.

Completed or status unknown trials

Title (status), ID number	Condition(s)	Intervention(s)	Start date to estimated completion date	Record last updated
Comparison of hip resurfacing and cementless metal-on-metal total hip arthroplasty (completed), NCT00570167	Hip OA	Procedures: hip RS and cementless THR with metal-on-metal bearings	November 2006 to unavailable	3 August 2009
A randomised single centre study to compare the long-term wear characteristics of Marathon™ and Enduron™ polyethylene cup liners in primary total hip replacement (unknown), NCT00208442	OA, post-traumatic arthritis, collagen disorders, avascular necrosis, traumatic femoral fractures, non-union of femoral fractures, congenital hip dysplasia, slipped capital femoral epiphysis	Devices: Marathon and Enduron polyethylene cup liners	June 2001 to unavailable (primary completion date September 2006)	7 April 2009
A randomized clinical study comparing a fully-coated cementless stem versus a cemented stem for total hip replacements (completed), NCT00689689	OA	Devices: fully coated Prodigy stem and cemented Endurance stem	July 1998–November 2009	15 February 2012
28 mm ceramic-on-ceramic acetabular cup total hip replacement study (completed, has results), NCT00208507	Joint diseases	Devices: 28-mm ceramic head on ceramic acetabular liner and 28-mm ceramic head on polyethylene liner	April 2003–December 2010	30 September 2011
Metal on metal versus ceramic on metal hip replacement (completed, has results), NCT00208494	Joint diseases	Devices: metal-on-metal and ceramic-on-metal THR	August 2005–June 2011	28 February 2012
Multi-centre comparative trial of the ASR™-XL acetabular cup system vs. the Pinnacle metal-on-metal total hip system (completed), NCT00561600	Osteoporosis, arthritis, trauma	Devices: ASR-XL modular acetabular cup system and Pinnacle acetabular shell	August 2006–January 2012	1 June 2012
A randomised controlled trial of total hip arthroplasty versus resurfacing arthroplasty in the treatment of young patients with arthritis of the hip joint (completed), UKCRN ID: 4093	Arthritis of the hip	Procedures: RS and THR	June 2007–May 2010	

Ongoing trials

Title (status), ID number	Condition(s)	Intervention(s)	Start date to estimated completion date	Record last updated
Wear measurements of E-poly in an uncemented THA with either 32 or 36 mm caput of ceramics (active, not recruiting), NCT00804388	Primary arthrosis, OA	Procedure: THR	December 2008–February 2012	6 January 2011
The metaphyseal hip prosthesis – total hip (recruiting), NCT01501955	OA of the hip	Devices: metaphyseal hip prosthesis and Stanmore hip prosthesis	October 2012–December 2023	6 November 2012
Comparison of hip resurfacing to large femoral head total hip arthroplasty (active, not recruiting), NCT00175487	OA, avascular necrosis	Procedure: hip replacement	September 2007–December 2012	28 September 2011
RCT of ceramic bearing primary total hip arthroplasty (active, not recruiting), NCT01522014	THR	Devices: Secure-Fit™ arc-deposited hydroxyapatite shell – surface ceramic and an Alumina Bearing Couple (ABC) ceramic insert and ceramic C-taper head; Secure-Fit arc-deposited hydroxyapatite shell, a Crossfire® insert and ceramic C-taper head	November 1997–June 2014	2 February 2012
Metal-metal articulations versus standard 28 mm cementless total hip arthroplasty (active, not recruiting), NCT01113762	OA, hip	Devices: ASR™ (DePuy); ReCap®/Magnum™ modular head (Biomet); Bimetric® stem, Mallory/Head cup, 28-mm ceramic head (Biomet); 28-mm Colbalt-Chromium head, Trilogy CH cup, VerSys® Fiber stem (Zimmer, Inc.)	February 2007–November 2011	29 April 2010
Efficacy of a new resurfacing hip prosthesis (recruiting), NCT00391937	OA, hip	Device: ASR hip prosthesis	October 2006–March 2013	1 June 2010
RCT comparing ion levels and clinical outcomes of A-class BFH to metal on polyethylene total hip replacement (recruiting), NCT00911599	Hip joint, OA, arthroplasty	Devices: A-class BFH™ and metal-on-polyethylene THR	August 2006–May 2012	20 October 2011
A multicentre trial to compare two articulating bearing surfaces as used in cementless primary hip arthroplasty (recruiting), NCT01247038	OA	Device: ceramic-on-metal articulating bearing surface	January 2011–November 2022	8 November 2012

DOI: 10.3310/hta19100

Title (status), ID number	Condition(s)	Intervention(s)	Start date to estimated completion date	Record last updated
RSA study using two types of uncemented acetabular components and the uncemented hydroxyapatite-coated Symax Stem (recruiting), NCT01618084	OA	Device: Titanium acetabular component	November 2011–December 2014	11 June 2012
A comparison of two total hip replacements: hip resurfacing system versus Mallory-Head/Exeter (active, not recruiting), NCT00116948	OA	Device: ReCap hip resurfacing system (Biomet)	January 2005–January 2012	3 May 2011
A study on M2a magnum total hip arthroplasty (active, not recruiting), NCT01010763	Degenerative joint disease, avascular necrosis	Devices: Metal-on-metal articulation and metal-on-metal acetabular system	November 2009–December 2021	7 February 2012
RSA-study of cemented hip prostheses with five different articulations (active, not recruiting), NCT00698672	Arthritis	Devices: Charnley Ogee, Spectron Reflection Cobalt-Chromium, Spectron XLPE Cobalt-Chromium, Spectron Reflection Oxinium and Spectron XLPE Cobalt-Chromium	November 2004–June 2012	21 May 2012
Cemented Marathon/Corail versus Pinnacle/Corail (recruiting), NCT01693627	Arthritis of the Hip	Devices: Pinnacle/Corail with collar, Marathon/Corail with collar, Pinnacle/Corail without collar and Marathon/Corail without collar	January 2012–January 2022	8 October 2012
Improving orthopedic outcomes through a national total joint replacement registry (Department of Orthopedics and Physical Rehabilitation) (ongoing), HSRProj ID: HSRP20111039	Arthritis	Procedure: Total joint replacement	September 2010–September 2014	NA

NA, not applicable.

Ongoing trials selected for evidence on cost-effectiveness

Title (status), ID	Condition(s)	Intervention(s)	Start date to estimated completion date	Record last updated
Multi-centre study to assess the long-term performance of the DePuy ASR™ system in resurfacing and primary total hip replacement (active, not recruiting), NCT00872547	RA, OA; post-traumatic arthritis, collagen disorders, avascular necrosis, non-union of femoral fractures, congenital hip dysplasia, slipped capital femoral epiphysis	Devices: DePuy ASR hip system and DePuy ASR XL head/ASR acetabular cup system	September 2006 to unknown	1 September 2011
An electronic data capture study to assess the long-term performance of the DePuy Proxima™ hip in primary total hip replacement (recruiting), NCT01134445	RA, OA, post-traumatic arthritis, avascular necrosis, traumatic femoral fractures, congenital hip dysplasia	Device: DePuy Proxima hip	February 2010–January 2026	8 October 2012
AMIStem primary hip system prospective post-marketing multi-centre surveillance study (recruiting), NCT01107340	OA, arthritis, avascular necrosis, fracture of the femoral neck or head, congenital hip dysplasia	Device: AMIStem hip system	February 2010–February 2024	10 January 2012
A multi-centre study to assess the long-term performance of the Silent Hip™ in primary total hip replacement surgery (recruiting), NCT01383824	Primary arthritis, secondary arthritis	Device: Silent Hip	January 2011–March 2029	21 August 2012

DOI: 10.3310/hta19100

Appendix 8 Overview of outcomes of relevance in included randomised controlled trials and systematic reviews

TABLE 134 Overview of outcomes of relevance in included RCTs: studies comparing different types of THR ($n = 14$)

Study	Mortality (all cause)	HHS	OHS	Merle d'Aubigné and Postel score	HOOS	LISOH score	AAOS Hip and Knee Questionnaire score	UCLA activity score	WOMAC score	AIMS score	MACTAR score	SF-36 score	SF-12 score	NHP score	EQ-5D score	Pain score (VAS)	Pain score (other than VAS)	Revision rate	Time to revision	Implant survival rate	Femoral head penetration
Cup fixation																					
Bjørgul 2010,[110] Bjørgul 2010[111]	✓	✓																		✓	
Angadi 2012[112]		✓																✓		✓	
Cup liner bearing surface																					
McCalden 2009[145]	✓	✓							✓				✓								✓
Engh 2012,[113] Engh 2006[114]	✓	✓																✓		✓	✓
Cup shell design																					
Capello 2008,[115] D'Antonio 2005,[116] D'Antonio 2003,[117] Mesko 2011[118]		✓																✓		✓	
Cup/stem fixation																					
Corten 2011,[119] Corten 2011,[122] Laupacis 2002,[120] Bourne 2010[121]	✓	✓		✓					✓		✓							✓		✓	
Femoral head size																					
Howie 2012[123]	✓																	✓			

Study	Mortality (all cause)	HHS	OHS	Merle d'Aubigné and Postel score	HOOS	LISOH score	AAOS Hip and Knee Questionnaire score	UCLA activity score	WOMAC score	AIMS score	MACTAR score	SF-36 score	SF-12 score	NHP score	EQ-5D score	Pain score (VAS)	Pain score (other than VAS)	Revision rate	Time to revision	Implant survival rate	Femoral head penetration
Femoral head bearing																					
Lewis 2008[124]		✓							✓				✓					✓		✓	
Femoral head bearing-on-cup liner bearing																					
Amanatullah 2011[125]		✓											✓					✓			
Capello 2008,[115] D'Antonio 2005,[116] D'Antonio 2003,[117] Mesko 2011[118]		✓																✓		✓	
Kadar 2011[126]		✓																			✓
Stem composition																					
Healy 2009[127]		✓																✓		✓	
Stem design																					
Kim 2011[128]	✓	✓																✓			
Stem fixation																					
Kim 2011[129]		✓						✓	✓									✓		✓	

TABLE 135 Overview of outcomes of relevance in included systematic reviews: studies comparing different types of THR ($n = 5$)

Study	Mortality (all cause)	HHS	OHS	Merle d'Aubigné and Postel score	HOOS	LISOH score	AAOS Hip and Knee Questionnaire score	UCLA activity score	WOMAC score	AIMS score	MACTAR score	SF-36 score	SF-12 score	NHP score	EQ-5D score	Pain score (VAS)	Pain score (other than VAS)	Revision rate	Time to revision	Implant survival rate	Femoral head penetration
Cup fixation																					
Voigt 2012[137]		✓	✓															✓			
Pakvis 2011[138]		✓																✓		✓	
Clement 2012[139]																		✓		✓	
Femoral head bearing-on-cup liner bearing																					
Sedrakyan 2011[140]		✓											✓					✓			
Yoshitmi 2009[141]																		✓			

TABLE 136 Overview of outcomes of relevance in included RCTs: THR vs. RS ($n = 3$)

	Mortality (all cause)	HHS	OHS	Merle d'Aubigné and Postel score	HOOS	LISOH score	AAOS Hip and Knee Questionnaire score	UCLA activity score	WOMAC score	AIMS score	MACTAR score	SF-36 score	SF-12 score	NHP score	EQ-5D score	Pain score (VAS)	Pain score (other than VAS)	Revision rate	Time to revision	Implant survival rate	Femoral head penetration
Costa 2012,[130] Achten 2010[107]		✓	✓												✓						
Garbuz 2010[131]								✓	✓			✓									
Vendittoli 2010,[132] Vendittoli 2006,[133] Girard 2006,[134] Rama 2009,[135] Vendittoli 2006[136]				✓				✓	✓									✓			

TABLE 137 Overview of outcomes of relevance in included systematic reviews: THR vs. RS ($n = 3$)

	Mortality (all cause)	HHS	OHS	Merle d'Aubigné and Postel score	HOOS	LISOH score	AAOS Hip and Knee Questionnaire score	UCLA activity score	WOMAC score	AIMS score	MACTAR score	SF-36 score	SF-12 score	NHP score	EQ-5D score	Pain score (VAS)	Pain score (other than VAS)	Revision rate	Time to revision	Implant survival rate	Femoral head penetration
Jiang 2011[142]	✓	✓		✓				✓	✓									✓			
Smith 2010[143]	✓	✓							✓									✓			
Springer 2009[144]																					

Appendix 9 Table of functional/clinical and quality of life measures

Name of the measure/tool	Construct	Domains addressed	Total score range
Functional clinical scores			
HHS	Hip/disease specific	Pain, function, absence of deformity, range of motion (10 items)	0–100
OHS	Hip/disease specific	Pain and function (12 items)	0–48
Merle d'Aubigné and Postel score	Hip/disease specific	Pain, mobility, ability to walk (three items)	0–18
UCLA activity score	Hip/disease specific	Daily living activity, sports activity (10 items)	1–10
WOMAC	Hip/disease specific	Pain, stiffness, physical function (24 items)	0–100
MACTAR	Hip/disease specific	Domains of function such as domestic care, self-care, professional activities, leisure activities, social interaction and roles; change in limitations of daily activities	0–30
Health-related quality of life scales			
SF-36	Generic	Broad mental and physical domains consisting of physical function, role limitations because of physical problems, role limitations because of emotional problems, energy levels, body pain, psychological distress, general health and social function (36 items)	0–100
SF-12	Generic	Mental and physical components (12 items)	0–100
EQ-5D	Generic	Mobility, self-care, usual activities, pain or discomfort, anxiety, depression, overall health perception (five items)	0–1

Appendix 10 Data extraction table of characteristics of eligible cost-effectiveness total hip replacement and resurfacing arthroplasty studies

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Alberta Heritage Foundation for Medical Research 2006,[208] Canada	*Type*: prospective case–control study and economic (cost–utility) analysis *Aim*: to determine the effectiveness, cost-effectiveness and safety of new alternative hip bearing devices	*Population*: THR patients aged < 56 years suitable for RS ($n = 329$) *Age (years), mean*: metal-on-metal RS ($n = 188$): 47.6; THR ($n = 141$): 47.7 *Outcomes*: quality of life (WOMAC, SF-36 and HUI-3) pre surgery and at 3, 6 and 12 months post surgery and annually thereafter; ng/ml of chromium and cobalt in blood; revision rate; costs (CA$); operating room time; LOS; cost per QALY *Economic model*: Markov model, sensitivity analyses	3 months post-surgery interim analysis of 191 patients (metal-on-metal RS $n = 112$ and THR $n = 79$): ceramic-on-ceramic THR was the most cost-effective option for patients aged < 56 years. An 18% decrease in metal-on-metal RS device costs and a 3-day LOS could result in metal-on-metal RS being the most cost-effective option in year 1	Short-term analysis favours ceramic-on-ceramic THR over metal-on-metal RS	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☐ *Comment*: table of how to break down costs and cost drivers, but costs reported not broken down
Amman 2012,[254] USA	*Type*: retrospective cohort study *Aim*: to compare a cohort of patients who underwent two-incision MIS THR with a matched cohort of patients who received THR using a modified direct-lateral Hardinge approach (STD THR)	*Population*: patients undergoing MIS THR ($n = 49$) matched to STD THR patients ($n = 49$) *Male*: 59% ($n = 29$) *Age (years), mean (SD)*: 59.1 (10.3) *OA*: 92% *Outcomes*: hospital costs and charges, LOS, component position, complication rates	LOS (mean days): MIS THR 2.41 days (95% CI 2.16 to 267 days), STD THR 3.64 days (95% CI 3.25 to 4.04 days) Total direct cost: MIS THR US$9846 (95% CI US$9536 to US$10,156), STD THR US$11,143 (95% CI US$10,651 to US$11,636) ($p < 0.001$) Total indirect cost: MIS THR US$3194 (95% CI US$3009 to US$3378), STD THR US$3661 (95% CI US$3439 to US$3884; $p = 0.002$)	Component position and complication rates were identical for the two groups. Hospital costs and charges and LOS were significantly lower for the two-incision group	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: detailed mean costs for THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Bak 2008,[266] Germany	*Type*: prospective cohort study *Aim*: to evaluate patient-reported functional recovery from primary hip and knee arthroplasty during a multidisciplinary inpatient rehabilitation programme	*Population*: patients with OA undergoing THR ($n = 40$) and TKR ($n = 41$) *Male*: 40% ($n = 16$) *Age (years), mean (SD)*: 65.4 (8.4) *Outcomes*: primary: SF-36, WOMAC and EQ-5D scores on admission, at discharge and 3 months post surgery	Both groups had positive longitudinal changes with moderate to large effect sizes for all outcome measures. Effect sizes were greater for WOMAC than for SF-36 and EQ-5D. Utilisation of resources: the THR group performed significantly better than the TKR group	THR and TKR represent two different states in terms of functional health	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐
Baker 2008,[252] Pollard 2006,[251] UK	*Type*: prospective cohort study *Aim*: to compare the medium-term clinical and radiological results of hybrid THR with those of metal-on-metal Birmingham RS	*Population*: patients who received RS ($n = 54$) matched with patients who received hybrid THR ($n = 54$) *Age (years), mean (range)*: RS 49.8 (18–67), THR 50.4 (21–66) *OA*: RS 78%, THR 74% *Outcomes*: quality of life (UCLA activity score, OHS, EQ-5D score)	THR: nine revisions, four died, five lost to follow-up; RS: five revisions, one died, five lost to follow-up 9- to 10-year follow-up: EQ-5D: THR 0.78 (95% CI 0.06 to 1.00); RS 0.84 (95% CI –0.18 to 1.00) VAS: THR 65.6% (95% CI 9% to 97%); RS 82% (95% CI 30% to 100%) UCLA activity score: RS significantly higher activity score ($p < 0.0001$) There were no significant changes with time in OHS and UCLA activity score	Patients with RS remained more active and had a lower rate of revision	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Batsis 2009,[267] USA	*Type*: retrospective cohort study and economic analysis *Aim*: to determine if hospital costs and LOS following elective THR are impacted by BMI	*Population*: THR patients stratified by preoperative BMI ($n = 5642$) *Male*: normal 32.4% ($n = 441$), overweight 56.9% ($n = 1220$), obese 55.1% ($n = 739$), morbidly obese 45.8% ($n = 363$) *Age (years), mean (SD)*: normal 62.4 (17.2), overweight 64.3 (14.1), obese 62.7 (13.0), morbidly obese 61.2 (12.7) *OA*: 88.1% ($n = 4699$) *Outcomes*: primary: LOS and costs (US$); secondary: composite end point of 30-day mortality, readmission, reoperation or intensive care unit utilisation *Economic analysis*: provider perspective, cost year 2005 (constant US$), direct costs included	Adjusted LOS was similar among BMI groups. Adjusted overall costs were no different ($p = 0.23$). There was no difference in the composite end point	BMI in patients undergoing primary elective THR did not impact on LOS or overall care costs despite higher operative costs in morbidly obese patients	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: Data by BMI category
Bohm 2012,[285] Canada	*Type*: retrospective cohort study *Aim*: to examine three negative outcomes along with resource use changes in the year after hip or knee arthroplasty surgery using a current national data set	*Population*: patients with hip or knee replacement ($n = 58,351$) *Outcomes*: rehospitalisation, revision, infection	In the year before surgery, 12.9% of elective hip patients and 10.2% of knee patients were hospitalised. In the year after surgery, 14.8% of elective hip patients and 15.5% of knee patients were hospitalised (increase of 15% and 52%, respectively). Revision occurred in 2.0% of emergent hip, 1.7% of elective hip and 0.9% of knee patients. Joint infection occurred in 1.3% of patients	The increased hospitalisation rate after the elective hip and knee procedures represents an incremental cost of 10% over the index hospital stay	1. (a) Resource use ☒ (b) Costs ☐ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: average inpatient days per hospital visit pre and post THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Bozic 2006,[259] USA	*Type*: retrospective economic and decision analysis *Aim*: to evaluate the cost-effectiveness of the use of alternative bearings (polyethylene, ceramic-on-ceramic, metal-on-metal) in THR	*Population*: patients with advanced OA of the hip *Male*: 100% *Age (years)*: 50 *Outcomes*: incremental costs (US$) *Decision model*: Markov model; time horizon: remaining life expectancy; direct medical costs considered; multiple sensitivity analysis for age, incremental implant cost and implant failure	In a population of 50-year-old patients alternative bearings (incremental cost of US$2000) would be cost-saving over the individual's lifetime compared with conventional bearings if the reduction in the 20-year implant failure rate is at least 19%. In patients aged > 63 years the same implant would be associated with higher lifetime costs than a conventional bearing, regardless of the reduction in revision rate. An alternative bearing with an incremental cost of US$500 could be cost-saving in patients aged > 65 years. In patients aged > 75 years no alternative bearing would be associated with lifetime cost savings	The cost-effectiveness of alternative bearings is highly dependent on the age of the patient at the time of surgery, the cost of the implant and the associated reduction in the probability of revision	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☒
Bozic 2010,[253] USA	*Type*: retrospective economic (cost–utility) and decision analysis *Aim*: to compare the comparative clinical effectiveness, costs and cost-effectiveness of metal-on-metal RS and THR	*Population*: patients with advanced OA of the hip *Outcomes*: discounted incremental costs, incremental clinical effectiveness, ICERs *Decision model*: Markov model; health-care system perspective; 30-year time horizon; cost year 2008 (US$); Monte Carlo sensitivity analysis for patient characteristics, clinical outcomes, quality of life and costs	RS: ICER < US$50,000 per QALY for men aged < 65 years and women aged < 55 years	RS could be clinically advantageous and cost-effective in younger men and women	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☒ 3. Transition probabilities ☒

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Bozic 2011,[284] USA	*Type*: prospective cohort study *Aim*: to measure health state utility in patients with hip OA and THR	*Population*: patients with hip OA or previous THR ($n = 231$) *Male*: 45.5% ($n = 105$) Age (years), mean: 56.5 *Outcomes*: mean utility score using the time trade-off technique for six cohorts	Mean (SD) utility: chronic OA of the hip ($n = 86$) 0.60 (0.34), successful primary THR ($n = 51$) 0.96 (0.09), failed primary THR ($n = 30$) 0.59 (0.34), successful revision THR ($n = 21$) 0.84 (0.28), failed revision THR ($n = 27$) 0.57 (0.36), chronically infected THR ($n = 16$) 0.46 (0.28)	THR results in substantial improvement in perceived health status in patients with chronic hip OA. Failed THR results in health state utilities that are similar to or worse than that for chronic OA	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐
Briggs 2004,[38] UK	*Type*: retrospective economic (cost–utility) and decision analysis *Aim*: to compare the Charnley and Spectron hip prostheses in terms of lifetime costs and QALYs	*Population*: patients undergoing primary hip replacement ($n = 20,495$) with Charnley ($n = 18,505$) and Spectron ($n = 1990$) prostheses *Male*: Charnley 33.3% ($n = 6168$), Spectron 26.0% ($n = 518$) *Age*: 71–80 years: Charnley 43.7% ($n = 8090$), Spectron: 51% ($n = 1014$) *OA*: Charnley 70.1% ($n = 12,970$), Spectron 76.7% ($n = 1348$) *Outcomes*: costs (£), QALYs, ICERs *Economic model*: Markov model, NHS perspective, 60-year time horizon, cost year 2000–1 (£), sensitivity analysis for failure risks	*Risk of revision HR*: Spectron vs. Charnley: early revisions 0.67 (95% CI 0.32 to 1.02), late revisions 0.26 (95% CI 0.07 to 0.46) Based on mean costs and QALYs, Spectron is cost-saving in younger patients and generates ICERs of between £1000 and £16,000 in older patients	For a threshold of £20,000 per additional QALY, the probability of the Spectron being the more cost-effective prosthesis ranged between 70% and 100%, depending on the age and sex of the patient	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☐ Comments: lifetime costs and QALYs by age group

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Brunenberg 2005,[265] the Netherlands	*Type*: before-and-after trial and economic (cost–utility) analysis *Aim*: to determine the incremental cost-effectiveness of a clinical pathway for patients undergoing joint replacement (the Joint Recovery Programme) compared with usual care	*Population*: patients on waiting list for THR ($n = 98$) TKR ($n = 62$) *Male*: THR intervention 35.4%, THR control 24%, TKR intervention 33.3%, TKR control 31.3% *Age (years), mean (SD)*: THR intervention 63.38 (11.48), THR control 65.40 (13.04), TKR intervention 64.9 (9.43), TKR control 63.94 (12.6) *Outcomes*: quality of life (HHS, EQ-5D) at baseline and 7, 12, 26 and 52 weeks post THR), incremental effects, incremental costs, ICERs *Economic analysis*: societal perspective; direct and indirect medical costs and indirect non-medical costs considered; cost year 2002 (US$); five one-way sensitivity analyses of LOS, costs of personnel and mean QALYs; time horizon 1 year post surgery	THR group: incremental QALYs 0.05 ($p = 0.174$) in favour of intervention; reduced LOS: 4.1 days ($p < 0.001$); incremental costs US$1261 TKR group: incremental QALYs 0.04 ($p = 0.33$) in favour of intervention; reduced LOS: 6.9 days ($p < 0.001$); incremental costs US$3336 Incremental costs per QALYs for THR and TKR groups: dominance	Clinical pathway dominates usual care and is a highly cost-effective approach for joint replacement patients	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Burns 2006,[289] Canada	*Type*: review *Aim*: to explore the economics surrounding revision THR particularly with regard to the financial burden that it will place on health systems in the near future	*Population*: patients with primary THR ($n = 1264$) and revision THR ($n = 133$) *Outcomes*: costs (US$)	Cost per 10-point improvement in WOMAC score: THR US$1383, revision THR US$2458 Revision is 56% as cost-effective as primary THR at a hospital with the funding allocations of US$7331 for THR and US$8850 for revision THR	Revision THR is cost-effective. The incidence of primary THR is increasing and therefore so will the need for revision THR. Durable implants and a reduction in the rate of complications are the solutions to reduce revision rates	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs per 10-point improvement in WOMAC score using the health department funding allocation for one hospital
Chen 2012,[8] UK	*Type*: review *Aim*: to examine all relevant literature on the economic costs of OA in the UK and to compare such costs globally	*Population*: patients with OA *Outcomes*: direct costs, indirect costs, intangible costs	Significant variation in direct and indirect costs across studies. Costs for topical and oral NSAIDs were estimated to be £19.2M and £25.65M, respectively. The cost of hip and knee replacements was estimated to exceed £850M and arthroscopic surgery for OA was estimated to cost £1.34M	Although estimates of economic costs can be made using information from non-published data, there remains a lack of original research looking at the direct or indirect costs of OA in the UK	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs per year per OA patient not divided into knee and hip surgery and conventional management costs to one UK trust per THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Clement 2012,[300] USA	*Type*: prospective cohort study (case matched) *Aim*: to investigate the efficacy and cost-effectiveness of bipolar sealing devices in revision THR because of infection	*Population*: patients with infected THR *Male*: intervention 50% ($n = 19$), control 63% ($n = 24$) *Age (years), median (interquartile range)*: intervention 61.2 (50.0–73.4), control 49.1 (46.7–54.6) *Outcomes*: primary: total blood loss, transfusion requirements; secondary: operative time, LOS, change in blood pressure, complications	Median (interquartile range) operative time (minutes): intervention 134 (110–167), control 158 (125–195) ($p = 0.039$) Gross savings: US$494.91–536.15 per case based on operative time Median (interquartile range) total blood loss (ml): intervention 998 (1106–1875), control 1330 (1106–1875) No significant difference in transfusion requirements	The outcomes may support the use of bipolar sealing in revision THR because of infection	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: Costs expressed as cost-savings only
Coyle 2008,[148] Canada	*Type*: systematic review and economic (cost–utility) analysis *Aim*: to examine the impact of adopting MI THR into the Canadian health system	*Population*: patients eligible for THR because of degenerative, rheumatoid or other arthritic diseases of the hip *Age (years)*: 68 *Outcomes*: QALYs, costs, ICERs *Economic model*: Markov model; Canadian public health-care system perspective; 40-year time horizon; cost year 2006 (CA$); sensitivity analysis for median rather than mean costs, operating room costs, utilities, discount rate, revision rate; Monte Carlo simulation; value of information analysis	MI THR vs. STD THR: costs: CA$20,400 vs. CA$19,100; QALYs: 7.48 vs. 7.47 Incremental cost per QALY: CA$148,300 The probability that MI THR is more cost-effective than STD THR at threshold of CA$50,000 per QALY is 47%. It would be cost-effective to spend up to CA$480M on gathering additional data through field evaluation	There was little evidence of any longer-term differences between MI THR and STD THR, mainly because of a lack of data. The economic analysis found little difference between therapies in terms of costs and QALYs	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☒ 3. Transition probabilities ☒

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Cullen 2012,[264] New Zealand	*Type*: retrospective matched cohort study *Aim*: to compare costs and outcomes for elective hip and knee arthroplasties carried out at the pilot site (Waitakere Hospital) and the main district health board hospital site (NSH)	*Population*: patients ($n = 335$) with THR ($n = 177$) or total knee replacement (TKR) ($n = 158$) *Male*: THR pilot ($n = 100$) 50%, THR NSH ($n = 77$) 43%, TKR pilot ($n = 70$) 51%, TKR NSH ($n = 88$) 43% *Age (years), mean*: THR pilot 63.1, THR NSH 64.9, TKR pilot 65.8, TKR NSH 67.3 *Outcomes*: time in theatre, LOS, costs	Pilot site vs. NSH: total inpatient event costs were 12% and 17% lower for hip and knee replacements, respectively. Significant reduction in operation length (39% THR, 36% TKR) and LOS (38% THR, 39% TKR)	Implementation of a new model in a public hospital setting produced significant increases in productivity and reduced overall costs. Model could potentially be used in other public health-care settings	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐
Cummins 2009,[262] USA	*Type*: economic (cost–utility) analysis *Aim*: to assess the cost-effectiveness of the use of antibiotic-impregnated bone cement for primary THR	*Population*: patients with THR for OA *Age (years)*: 68 *Outcomes*: revision rates (all or because of infection), costs, QALYs *Economic model*: Markov model, cost year 2002 (US$), sensitivity analysis on each of the parameters in the model	ICER US$37,355 per QALY for antibiotic-impregnated cement vs. cement without antibiotics	The use of antibiotic-impregnated bone cement results in an overall decrease in costs	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Davies 2010,[43] UK	*Type*: review *Aim*: to summarise published economic evidence on alternative hip prostheses to examine the potential for the literature to inform resource allocation decisions in the UK	*Population*: patients with THR *Outcomes*: direct medical resource use: prosthesis, operative time, postoperative care, LOS, management of surgical/implant/postoperative complications, medication, revision surgery within follow-up period, long-term revision surgery (prosthesis failure); non-medical resource use: productivity losses, out of-pocket expenses; health effects: postoperative pain, complications, health-related quality of life, mortality, QALYs, ICERs *Cost year*: 2008 (£)	In general, cemented prostheses (£455–1693) were cheaper than cementless prostheses (£691–2845). The average total cost of the THR procedure per patient reported in the studies ranged from £4599 to £8078 Mean LOS was 7.3–23 days Identified methodological problems: lack of observed long-term prosthesis survival data, limited up-to-date and UK-based evidence, exclusion of patient and societal perspectives	More clinical trials including long-term follow-up and economic evaluation are needed. Trials should compare the cost-effectiveness of different prostheses with longer-term follow-up and including a wider perspective	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: Prosthesis costs from the literature for six different brands; LOS and duration of surgery for cemented vs. cementless vs. hybrid prostheses
Dawson 2001,[295] UK	*Type*: prospective cohort study *Aim*: to evaluate the OHS as an outcome measurement tool after revision THR	*Population*: patients with revision THR ($n = 609$) *Male*: 43% ($n = 260$) *OA*: 78.7% ($n = 463$) *Reason for revision*: aseptic loosening 72.3% ($n = 426$), infection 9.8% ($n = 58$), recurrent dislocation 6.8% ($n = 40$), other 11.0% ($n = 65$) *Outcomes*: quality of life (OHS and EQ-5D score) at 2 weeks before and 12 months post surgery	Pre revision THR: EQ-5D score, mean (95% CI): 0.32 (0.29 to 0.36); OHS, mean (95% CI): 43.0 (42.3 to 43.8) Post revision THR: EQ-5D score, mean (95% CI): 0.62 (0.59 to 0.65); OHS, mean (95% CI): 26.4 (25.3 to 27.4)	There is considerable variation in outcomes. High level of agreement between OHS and EQ-5D score. OHS displays a greater effect size; therefore, more sensitive than EQ-5D score. OHS scores deteriorated progressively with the number of previous revisions. EQ-5D score showed less sensitivity to change	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
de Palma 2012,[288] Italy	*Type*: retrospective economic analysis *Aim*: to determine the costs associated with the treatment of early dislocation in a consecutive series of patients with hemiarthroplasty (HA), THR and revision THR (RTHR)	*Population*: patients ($n = 87$) with dislocation following HA ($n = 18$), THR ($n = 44$) [including 73% with OA ($n = 32$)] and RTHR ($n = 25$) *Male*: HA 44% ($n = 8$), THR 43% ($n = 19$), RTHR 36% ($n = 9$) *Age (years), mean*: HA 85.5, THR 72.1, RTHR 75.3 *Outcomes*: cost of implant dislocation (cost of closed reduction, open reduction, revision surgery, average LOS) *Economic analysis*: costs included: implant, length of surgery and LOS postoperatively; cost year 2008 (€)	LOS: closed reduction 27.5 days, open reduction 17 days, revision surgery 42 days Retreatment increased costs by: HA: 223.2% closed reduction, 176.6% open reduction, 352.4% revision surgery; THR: 218.3% closed reduction, 173.2% open reduction, 322.4% revision surgery; RTHR: 281.7% closed reduction, 212.3% open reduction, 472.1% revision surgery	The favourable cost–benefit ratio of primary HA, THR and RTHR may be reduced by costs of treating dislocation	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs (average implant cost, average daily surgeon and nursing costs, average daily cost of hospital stay) for THR and revision

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
de Verteuil 2008,[11] UK	*Type*: systematic review and retrospective economic (cost–utility) analysis *Aim*: to assess the clinical effectiveness and cost-effectiveness of minimal incision approaches to THR for arthritis of the hip	*Population*: patients undergoing primary THR *Age (years)*: 68 *Outcomes*: costs (£), QALYs *Economic model*: Markov model, 1-year and 40-year time horizon, NHS perspective, cost year 2006 (£), probabilistic sensitivity analysis	Costs: 1 year: single mini-incision THR £7060, standard THR £7350; 40-year time horizon: single mini-incision THR £11,618, standard THR £11,899 Mean QALYs: 1 year: mini-incision THR 0.695, THR 0.677; 40 years: mini-incision THR 8.480, THR 8.463 Probabilistic sensitivity analyses: 1 year: mini-incision THR: 95% probability of being cost-effective if society's WTP for a QALY were up to £50,000; 40 years: 55% probability Drivers: 1-month earlier return to usual activities, a decreased hospital LOS, decreased operation duration for mini-incision THR	Minimal incision THR has small perioperative advantages. It may offer a shorter hospital stay. It appears to have a similar procedure cost but evidence on its longer-term performance is limited	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☒ 3. Transition probabilities ☒ Comment: costs of non-operative management of failed THR, revision rates, very detailed costs and utilities for standard THR
di Tanna 2011,[261] Italy	*Type*: retrospective economic (cost-effectiveness) analysis *Aim*: to assess the cost-effectiveness of cementless vs. hybrid prostheses in THR in patients diagnosed with primary OA	*Population*: patients undergoing THR for OA *Age (years)*: 70 *Outcomes*: cost per 'revision-free' life-year, ICERs *Economic model*: Markov model, lifetime time horizon, provider perspective, deterministic sensitivity analysis	ICER: €2401.63 per revision-free life-year	The cementless strategy was more effective but more costly than the hybrid strategy	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☒ Comment: mean costs for hybrid and cementless THR and revision

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Dutka 2008,[271] Poland	*Type*: cohort study *Aim*: to carry out a retrospective comparative analysis of the cost of THR surgery vs. conservative treatment for OA in a variety of sociomedical aspects while patients are awaiting THR	*Population*: patients with hip OA awaiting ($n = 77$) or following ($n = 91$) THR *Outcomes*: quality of life (WOMAC score, SF-8), resource use, costs (Polish zloty or PLN)	Average direct costs of 2-year conservative treatment: PLN 6108; average indirect costs: PLN 9407 Average costs of operative treatment: PLN 7996; cemented THR: PLN 6500, uncemented THR: PLN 9200 WOMAC score: conservative treatment: average 36.1 points (max. 56 points, min. 2 points); THR: average 21.6 points	Pharmacological treatment, rehabilitation, physical therapy and other methods appear to be inefficient in patients with hip OA awaiting THR and their costs are twice as high as those of THR	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs for patients in waiting group expressed as average costs for 2-year period of conservative treatment. Costs for cemented THR vs. uncemented THR
Duwelius 2008,[255] USA	*Type*: prospective cohort study and retrospective economic (cost–utility) analysis *Aim*: to evaluate the 6-week cost-effectiveness of minimally invasive THR techniques relative to that of THR	*Population*: patients with primary THR (two-incision $n = 235$, mini-incision $n = 325$, conventional $n = 31$) *Male*: two-incision 66%, mini-incision 50%, conventional 42% *Age (years), mean*: two-incision 57.7, mini-incision 65.4, conventional 65.3 *OA*: two-incision 86%, mini-incision 89%, conventional 81% *Outcomes*: LOS and operation time, incremental QALYs (SF-36 score and HHS), incremental costs *Economic analysis*: direct and indirect costs considered (US$)	Incremental savings: two-incision US$5620, mini-incision US$5089 Incremental QALYs: two-incision 0.037, mini-incision 0.023	Minimally invasive two-incision hip procedure and the mini-incision technique yield better 6-week outcomes at a lower cost than the conventional technique	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☐ Comment: cost broken down into costs for surgeon, hospital and rehabilitation

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Edlin 2012,[40] Costa 2012,[130] UK	*Type*: RCT and economic (cost–utility) analysis *Aim*: to report on the relative cost-effectiveness of THR and RS in patients with severe arthritis suitable for hip joint RS	*Population*: patients aged > 18 years with severe arthritis of the hip joint suitable for RS ($n = 126$) who received THR ($n = 66$) and RS ($n = 60$) *Outcomes*: primary: hip function (12-month post-surgery OHS and HHS); secondary: quality of life (EQ-5D score), disability rating, physical activity level, complications, cost-effectiveness, incremental costs, ICERs *Economic model*: NHS perspective, 12-month time horizon, cost year 2009/10 (£), univariate sensitivity analyses	Hip function: mean OHS effect size: 2.23 (95% CI –1.52 to 5.98, $p = 0.070$); mean HHS effect size: 6.04 (95% CI –0.51 to 12.58, $p = 0.242$) Complication rates did not differ ($p = 0.291$) Quality of life at 12 months: RS 0.795, THR 0.727 RS vs. THR: incremental QALYs 0.032, incremental cost £564, ICER £17,451 per QALY	No evidence of a difference in hip function was seen in patients with severe arthritis of the hip, 1 year after receiving THR or RS. RS appears to offer very short-term efficiency benefits over THR within a selected patient group	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☒ 3. Transition probabilities ☐
Feeny 2004,[294] Canada	*Type*: prospective cohort study *Aim*: to assess agreement among utility scores and compare estimates of the change in utility following THR using four utility measures (SF-6D, SG, HUI-2 and HUI-3) and to compare the degree of responsiveness among the measures	*Population*: patients referred for THR for OA ($n = 86$) *Outcomes*: quality of life (SF-6D, SG, HUI-2 and HUI-3) at baseline, pre THR and post THR up to 3 months	At baseline: mean SF-6D (0.61), SG (0.62) and HUI-2 (0.62) scores were similar; the mean HUI-3 score (0.52) was lower Agreement between SF-6D and SG scores was 0.13, between SF-6D and HUI-2 scores was 0.47 and between SF-6D and HUI-3 scores was 0.28 Change in scores between post and pre surgery was 0.10 for SF-6D, 0.16 for SG, 0.22 for HUI-2 and 0.23 for HUI-3 Effect sizes were 1.10 for HUI-2, 1.08 for HUI-3, 1.06 for SF-6D and 0.48 for SG	Agreement between SG scores and SF-6D and HUI scores was low. The estimate of change in utility associated with THR was lowest for SF-6D	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: utilities not from EQ-5D

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Feeny 2004,[301] Canada	*Type*: prospective cohort study *Aim*: to assess the stability of standard gamble utility scores for three hypothetical health states describing mild, moderate and severe OA and to provide evidence on the marker state approach to assist in interpreting utility scores	*Population*: patients considered for THR ($n = 114$) *Outcomes*: health-related quality of life (6-minute walking test, SF-36, HUI-2, HUI-3, WOMAC, HHS, State–Trait Anxiety Inventory)	Mean utility scores for mild, moderate and severe OA were 0.69, 0.61 and 0.41, respectively. Time did not affect the scores for the three OA states	Group-level standard gamble scores are stable. THR converted moderate OA to better than mild	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: utilities not from EQ-5D
Fielden 2005,[282] New Zealand	*Type*: prospective cohort study *Aim*: to determine the economic and health costs of waiting for THR	*Population*: patients on waiting list for THR for OA ($n = 122$) *Male*: 35% ($n = 43$) *Age (years), mean*: 66 *Outcomes*: quality of life (WOMAC and EQ-5D scores) at 1, 3 and 6 months post THR, costs (NZ$ and US$)	Mean waiting time 5.1 months; mean total cost of waiting for surgery: NZ$4305 (US$2876) per person Waiting > 6 months: total mean cost NZ$4278 (US$2858) per person; waiting < 6 months: total mean cost NZ$2828 (US$1889) per person ($p < 0.01$) Preoperative to postoperative WOMAC and EQ-5D scores improved ($p \leq 0.01$)	Longer waits for THR incur greater economic costs and deterioration in physical function while waiting	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: costs per patient expressed as medical, personal and societal – not broken down; utilities by dimensions pre THR and 1, 3 and 6 months post THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments	
Fordham 2012,[257] UK	*Type*: retrospective economic (cost–utility) analysis *Aim*: to assess changes in quality of life and costs of patients undergoing primary THR using the Exeter prosthesis compared with those of a hypothetical 'no surgery' group	*Population*: patients undergoing THR ($n = 938$) compared with a hypothetical 'no surgery' group *Male*: 38.7% ($n = 363$) *Age (years), mean (SD)*: 62.2 (11.5) *Outcomes*: incremental costs, incremental quality of life (SF-36), QALYs *Economic analysis*: NHS perspective; included costs of operation, implant and LOS; 5-year time horizon; sensitivity analyses	QALY gain over 5 years for THR group 0.8 Association with increased QALYs: younger age, male, lower BMI and poorer OHS Compared with 'no surgery' the cost per QALY was £7182 (95% CI £6470 to £7678) The most likely cost per QALY was between £7058 and £7220	The study confirmed the long-term benefits and cost-effectiveness of THR in a wide variety of patients using well-established implant models such as the Exeter prosthesis	1. (a) Resource use (b) Costs 2. (a) Utilities (b) QALYs 3. Transition probabilities	☒ ☒ ☒ ☒ ☐
Fujita 2009,[272] Japan	*Type*: prospective cohort study *Aim*: to assess changes in the health outcomes of Japanese patients before and after THR and to assess the impact of THR on commonly performed postures or body positions requiring deep flexion of the hip joint such as the use of Japanese squat toilets	*Population*: patients undergoing primary THR ($n = 451$) *Male*: 15.5% ($n = 70$) *Age (years), mean (SD)*: 60.6 (10.0) *OA*: 87.8% ($n = 396$) *Outcomes*: quality of life before and 6 weeks and 6 months post surgery (WOMAC and EQ-5D scores)	Changes in WOMAC and EQ-5D subscale scores and scores for each item from the three time periods were highly significant ($p = 0.000$) Effect size at 6 months: WOMAC pain 1.56; physical function 1.38 Two items (Japanese toilet and seiza) became significantly worse at 6 weeks postoperatively ($p = 0.000$) and returned to preoperative levels by 6 months	Culturally sensitive physical functions in addition to conventional measurements for the health outcomes of THR patients are important	1. (a) Resource use (b) Costs 2. (a) Utilities (b) QALYs 3. Transition probabilities	☐ ☐ ☒ ☐ ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Higashi 2011,[273] Australia	*Type*: retrospective economic (cost–utility) analysis *Aim*: to evaluate the cost-effectiveness of hip and knee replacements in Australia	*Population*: patients with OA undergoing THR ($n = 68,908$) or total knee replacement (TKR) ($n = 100,657$) *Male*: THR: 44% ($n = 30,347$), TKR: 43% ($n = 42,930$) *Age (years)*: ≥ 40 *Outcomes*: costs, QALYs, ICERs *Economic model*: health-care system perspective, cost year 2003 (AU$), discrete-event simulation model, Monte Carlo probabilistic sensitivity analysis	ICERs: THR: AU$5000 per QALY; TKR: AU$12,000 per QALY < AU$50,000 per QALY threshold level	Hip and knee replacements are cost-effective interventions to improve the quality of life of people with OA	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐
Hulleberg 2008,[258] Norway	*Type*: prospective cohort study *Aim*: to evaluate medium-term clinical radiographic results in patients with a Charnley THR and to compare results with data from the Norwegian Arthroplasty Register	*Population*: Charnley THR operations ($n = 138$) in 123 patients *Male*: 25% ($n = 27$) *Age (years), median (range)*: 66 (50–70) *OA*: 83% *Outcomes*: Charnley category, quality of life (HHS, VAS, EQ-5D score), radiographic evaluation	13-year follow-up: 26 patients died, 20 revision HHS ($n = 93$): 83 (SD 15); HHS pain ($n = 93$): 41 (SD 6.5); VAS satisfaction ($n = 89$): 4.5 (SD 15); EQ-VAS ($n = 88$): 69 (SD 21); EQ-5D ($n = 89$): 0.75 (SD 0.24) Charnley C categories had poorer quality of life outcomes Survival: 89% (95% CI 84% to 95%) at 10 years and 85% (95% CI 79% to 92%) at 13 years	To fully appreciate the clinical effectiveness of an implant, specific hip function, patient satisfaction, quality of life and radiographic analysis must be considered. The functional status of the patient is important for the clinical outcome after THR	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Jansson 2011,[302] Sweden	*Type*: prospective cohort study *Aim*: to assess the effect of orthopaedic surgery as measured by EQ-5D score	*Population*: patients undergoing elective orthopaedic surgery ($n = 2444$) including hip arthroplasty ($n = 370$) *Male*: 43% *Age (years), mean (SD)*: 56 (18) *OA*: 33% *THR*: 15.1% *Outcomes*: EQ-5D score 1 day before and 12 months post surgery	Mean EQ-5D score improved from 0.54 to 0.72 THR, TKR, operations related to previous surgery, trauma-related procedures and RA surgeries: preoperative EQ-5D scores: 0.48–0.52, postoperative: 0.63–0.80 Patients with tumours or diseases of the elbow/hand: pre- and postoperative scores: 0.66–0.77	In most patients the EQ-5D score improved but did not reach the level reported for an age- and sex-matched population sample	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐
Jenkins 2013,[37] UK	*Type*: prospective observational study and retrospective economic (cost–utility) analysis *Aim*: to compare the current cost-effectiveness of THR and TKR using QALY methodology and to investigate the factors associated with variations in the health improvement derived from each procedure and correlate them with preoperative Oxford Hip and Knee Scores	*Population*: patients undergoing primary THR or total knee replacement (TKR) ($n = 671$); THR $n = 348$, TKR $n = 323$ *Male*: THR 42% ($n = 147$), TKR 40.6% ($n = 131$) *Age (years), mean (SD)*: THR 66.1 (12.9), TKR 69.9 (10.7) *Outcomes*: quality of life (Oxford Hip and Knee Scores, EQ-5D score) pre and 12 months post THR, QALYs gained, costs (£), cost per QALY	Number of QALYs gained: after THR 6.5, after TKR 4.0 ($p < 0.001$) Cost per QALY: THR £1372, TKR £2101 Predictors of an increase in QALYs gained: Poorer health before surgery ($p < 0.001$), younger age ($p < 0.001$) General health (EQ-5D VAS) showed greater improvement after THR than after TKR ($p < 0.001$)	THR and TKR are extremely effective both clinically and in terms of cost-effectiveness, with costs that compare favourably with those of other medical interventions	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: costs of primary THR and aseptic and septic revision from the Scottish National Tariff

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Jibodh 2004,[268] USA	*Type*: retrospective cohort study *Aim*: to determine the influence of BMI on perioperative morbidity, functional recovery and hospital use	*Population*: patients with primary THR ($n = 207$) *Male*: BMI (kg/m^2) < 25 ($n = 51$): 19% ($n = 37$); 25–29.9 ($n = 72$): 37% ($n = 51$); 30–39.9 ($n = 66$): 26% ($n = 39$); ≥ 40 ($n = 18$): 4% ($n = 22$) *Age (years), mean (SD)*: BMI < 25 kg/m^2 ($n = 51$): 68 (12); 25–29.9 kg/m^2 ($n = 72$): 64 (13); 30–39.9 kg/m^2 ($n = 66$): 63 (15); ≥ 40 kg/m^2 ($n = 18$): 59 (15) *OA*: 71% *Outcomes*: transfusion requirements, operative complications, functional recovery, assistance needed for transfers, LOS, costs (US$)	Morbidly obese patients (BMI ≥ 40 kg/m^2) had a significantly longer mean operative time and higher mean intraoperative blood loss ($p < 0.05$). There was also a trend towards more complications but no significant difference in functional recovery and hospital use	Patients of varying BMI undergoing primary THR have similar in-hospital outcomes and consume similar resources	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: costs and LOS by BMI group
Jimenez-Garcia 2011,[274] Spain	*Type*: retrospective cohort study *Aim*: to analyse changes in incidence, comorbidity profile, LOS, costs and in-hospital mortality of patients undergoing primary THR	*Population*: patients having undergone elective or emergency THR ($n = 161,791$) *Male*: 45% (73,248) *Age (years)*: ≥ 40 *Outcomes*: incidence, comorbidity profile, LOS, costs (€), in-hospital mortality	2001–8 Incidence: increase from 99 to 105 ($p < 0.001$) cases of THR per 100,000 inhabitants Charlson Index: prevalence of 1–2 and > 2 increased from 18.4% and 0.6% to 20.4% and 1.1%, respectively ($p < 0.001$) Mean LOS: decrease from 13 to 10.45 days ($p < 0.001$) Cost per patient: increase from €6634 to €9474	The health profile of patients undergoing THR seems to be worsening in Spain	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: LOS by age group Costs as mean hospital costs per year

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Jones 2012,[296] Canada	*Type*: review *Aim*: to provide an overview of the different types of disease-specific, generic and utility outcome measures used to assess recovery after THR and TKR and to summarise the reported changes in health-related quality of life after total joint arthroplasty	*Population*: patients with elective primary TKR or THR *Outcomes*: change in health-related quality of life (any validated health-related quality of life measure)	Disease-specific measures reported large and important changes for pain and function over short-term and long-term recovery. Smaller but important changes were reported with generic and utility measures	Changes in health-related quality of life were largest in the pain and physical function domains	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐
Judge 2011,[281] EU (multicentre)	*Type*: prospective cohort study *Aim*: to identify patient characteristics associated with preoperative expectations of THR and to explore whether preoperative expectations predict surgical outcomes 12 months post THR	*Population*: patients receiving primary THR ($n = 1327$) for hip OA *Male*: 44.1% ($n = 559$) *Age (years), mean*: 65.7 *Outcomes*: expectations, quality of life (WOMAC and EQ-5D scores) pre THR, medication use, Osteoarthritis Research Society International (OARSI) classification	Greater preoperative expectations associated with younger age, women, increasing BMI, higher educational level. Improvement after surgery associated with higher expectations – 34% (95% CI 1% to 78%) increase in improvement per expectation. Association is strongest for stiffness and function	There is a large variation in patients' preoperative expectations of THR	1. (a) Resource use ☒ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: resource use is for number of medications

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Kurtz 2012,[291] USA	*Type*: retrospective economic analysis *Aim*: to characterise differences in incidence, LOS and inpatient costs for periprosthetic joint infection between 2001–4 and 2005–9 and to identify patient and clinical factors that influence hospitalisation costs	*Population*: patients with infected hip ($n = 54,292$) and knee ($n = 105,068$) replacements *Outcomes*: LOS, costs (US$)	Decrease in LOS: hip patients: 11.5 (95% CI 10.3 to 12.7) days in 2001 to 9.5 (95% CI 8.8 to 10.2) days in 2009; knee patients: 9.3 (95% CI 8.2 to 10.4) days in 2001 to 7.2 (95% CI 6.9 to 7.5) days in 2009 Treatment costs: average hospital cost hip: US$31,300 (95% CI US$28,300 to US$34,300) in 2001 and US$30,300 (95% CI US$27,600 to US$33,000) in 2009; hip revision: US$72,700 in 2001 and US$93,600 in 2009; average hospital cost knee: US$25,300 (95% CI US$22,500 to US$28,100) in 2001 and US$24,200 (95% CI US$22,800 to US$25,600) in 2009; knee revision: US$58,700 in 2001 and US$74,900 in 2009 Factors having a significant effect on costs: geographic census region in which the patient lived, minority patients of any race	The demand for THR and total knee replacement is expected to increase substantially over the coming decade and so is the economic burden of prosthetic infections	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Larsen 2008,[263] Larsen 2009,[303] Denmark	*Type*: RCT and economic (cost–utility) analysis *Aim*: to investigate the efficacy of an accelerated perioperative care and rehabilitation intervention in patients undergoing primary THR and TKR or UKR and to investigate, from a societal perspective, the cost-effectiveness of such a programme compared with standard protocol following total THR and TKR	*Population*: patients undergoing THR, TKR or UKR ($n = 87$); intervention group ($n = 45$): 28 THR, 15 TKR, 2 UKA; control group ($n = 42$): 28 THR, 12 TKR, 2 UKA *Male*: intervention 44% ($n = 20$), control 55% ($n = 23$) *Age (years), mean (SD)*: intervention 64 (10.8), control 66 (9.2) *Outcomes*: LOS, quality of life (EQ-5D score), adverse effects, costs (DKK), cost per QALY *Economic analysis*: societal perspective, marginal analysis, 1-year time horizon, cost year 2006 (DKK)	Mean LOS: reduction from 8 (95% CI 7.1 to 8.4) days in the control group to 5 (95% CI 4.2 to 5.6) days in the intervention group ($p < 0.001$) Quality of life: QALY gain 0.08 (95% CI 0.004 to 0.16) in the intervention group ($p = 0.03$) Average cost reduction: DKK 18,880 (95% CI DKK 1899 to DKK 38,152) ($p = 0.036$) 98% dominance of the accelerated protocol for THR for a QALY gain of 0.08 ($p = 0.006$)	An accelerated perioperative care and rehabilitation intervention in patients undergoing THR, TKR or UKR is effective and can be cost-saving	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Laupacis 2002,[120] Canada	*Type*: RCT *Aim*: to compare the fixation of a Mallory-Head total hip prosthesis with and without cement	*Population*: patients with OA of the hip undergoing THR ($n = 250$) *Male*: cemented ($n = 124$) 52%, cementless ($n = 126$) 54% *Age (years), mean (SD)*: cemented 64 (8), cementless 64 (7) *Outcomes*: mortality, revision, quality of life (time trade-off, MACTAR score, WOMAC score, HHS, Sickness Impact Profile and Merle d'Aubigné and Postel hip score) pre surgery and 3, 6 and 12 months post surgery and yearly thereafter	Revisions: cemented 13, cementless 6 ($p = 0.11$) All health-related quality-of-life measures improved postoperatively in both groups	Hip replacement has a dramatic and sustained effect on health-related quality of life	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐
Lemon 2008,[304] UK	*Type*: retrospective economic analysis *Aim*: to calculate the in-hospital costs of THR and total knee replacement (TKR) for one NHS treatment centre and compare them with their reimbursement under the national tariff system	*Population*: patients with TKR and THR ($n = 3623$ operations); primary 83% (THR: $n = 1118$, TKR: $n = 1538$), revision 5%, ligament reconstructions and arthroscopies 12% *Outcomes*: costs (£)	Cost (with implant discount coming into effect in April 2005): TKR £6499, THR £6054 The study's THR cost is 2.3% less than and 1% more than the 2004–5 and 2005–6 tariffs, respectively. The TKR cost is 5% and 4.2% less than the 2004–5 and 2005–6 tariffs, respectively The study's cemented and uncemented THR costs are 8% less than and 6.6% more than their respective tariffs in 2005–6	Despite the implant price reduction (equating to a 17% decrease in cost related to the average THR/TKR tariff at the time), the TKR and THR costs are within 5% of the tariff	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
March 2002,[286] Australia	*Type*: prospective cohort study *Aim*: to determine patient out-of-pocket costs during the first year following joint replacement and to explore whether health status pre surgery or 3 months post surgery was a determinant of costs	*Population*: patients with OA scheduled for THR ($n = 76$) or TKR ($n = 98$) *Male*: THR 54%, TKR 47% *Age (years), mean (SD):* THR 63.3 (11.7), TKR 70.4 (7.00) *Outcomes*: out-of-pocket costs (AU$), quality of life (WOMAC and SF-36 scores)	Out-of-pocket costs fell considerably over the first postoperative year, the proportion of patients who experienced no out-of-pocket costs increased and the proportion of patients who made no use of health services increased THR patients: pension status, preoperative SF-36 score and 3-month postoperative WOMAC score were significant independent predictors of postoperative costs TKR patients: pre-surgery WOMAC score and pension status were significant independent predictors of postoperative costs	Out-of-pocket costs pre surgery fell dramatically over the first postoperative year. Poorer pre-surgery health status predicted greater expenditure during the first postoperative year	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs are out-of-pocket costs from the patient perspective
Marinelli 2008,[260] Italy	*Type*: retrospective economic (cost–utility) analysis *Aim*: to establish a framework within which to evaluate the cost-effectiveness of cementless and cemented implants and to analyse how device cost and revision rates affect the model	*Population*: theoretical cohort of patients with unsuccessful management of femoral neck fracture or arthritis of the hip *Age (years)*: 70 *Outcomes*: costs, QALYs, ICERs *Economic model*: cost year 2006 (€), payer perspective, considered average hospital costs	Risk of early revision (at 5 years of follow-up): cementless 1.6%, cemented 1.4% Equal QALYs No cost savings for the two implant types Analysis of mean costs and QALYs indicated that use of either implant is not associated with cost savings	Management with cementless or cemented total hip prostheses was not significantly different according to the model	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☐ Comment: mean implant costs for cementless and cemented THR and revision

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Montin 2009,[287] Finland	*Type*: prospective cohort study *Aim*: to determine the economic outcomes (service use, health-care and non-health-care out-of-pocket costs) related to THR from the perspective of patients	*Population*: patients with OA of the hip undergoing THR ($n = 100$) *Male*: 46% *Age (years), mean (SD)*: 63.9 (11.6) Cemented THR 50%, RS 40%, special implant 10% *Outcomes*: service use, out-of-pocket costs (€), health-related quality of life (Sickness Impact Profile) pre THR and 3 and 6 months post THR	Age, pain, sex, civil status, type of surgery and discharge destination showed association with service use Health-care costs comprised > 90% of total out-of-pocket costs. Non-health-care costs comprised <10% of total out-of-pocket costs No significant correlations between costs and health-related quality of life were observed	When deciding the timing of surgery, patients' characteristics, especially level of pain and health-related quality of life, should be carefully evaluated, as they may predict patients' service use and ability to manage at home after surgery	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs are out-of-pocket costs from the patient perspective
Mota 2013,[275] Italy	*Type*: retrospective economic (cost–utility) analysis *Aim*: to assess the cost-effectiveness of early primary THR for functionally independent older adult patients with OA compared with (1) non-surgical therapy followed by 'delayed THR' and (2) non-surgical therapy alone	*Population*: patients with choice of THR for OA *Age (years)*: 50–59 years, 60–74 years, ≥ 75 years *Outcomes*: QALYs, ICERs *Economic model*: Markov model, IPD, Italian NHS perspective, cost year 2010 (€), utilities from EQ-5D, lifetime time horizon (100 years of age)	Patients aged 65 years: ICER for THR vs. delayed THR €987 in men and €466 in women; ICER for delayed THR vs. medical therapy €463 in men and €82 in women Patients aged 80 years: early THR is (extendedly) dominant Longer female life expectancy implies longer later periods of low health-related quality of life with early THR; therefore, delaying surgery may be more appealing in women than in men in their 50s	THR is cost-effective Patients' health-related quality of life benefits forgone with delayed THR are worth more than the costs it saves to the Italian NHS	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☒ Comment: detailed costs for THR and non-surgical management

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
O'Shea 2002,[278] Ireland	*Type*: prospective economic (cost–utility and cost–benefit) analysis *Aim*: to assess the cost–utility and cost–benefit of THR	*Population*: patients undergoing primary THR ($n = 668$) from the FC-2 study *Outcomes*: LOS, quality of life (SF-36 score, WOMAC score, HHS) pre THR and 1 and 2 years post THR, costs per QALY [in punts (Irish pounds)], costs per 10-point increment in HHS	Average unit cost per THR performed IR£6472.06 LOS 16.4 days Cost per 10-point increment in HHS IR£2575 Cost per QALY for 60- to 69-year-olds: IR£1863.55 men, IR£1467.27 women; cost per QALY for 70- to 79-year-olds: IR£3152.00 men, IR£2454.90 women	THR is cost-effective for patients with OA	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comments: no information on population or sources of costs
Ostendorf 2004,[297] the Netherlands	*Type*: prospective cohort study *Aim*: to define the minimum set of PROMs that are required to assess health status after THR	*Population*: Patients undergoing THR ($n = 114$) *Male*: 37.7% *Age (years), mean (SD)*: 67.6 (10.1) *OA*: 83.3% *Cemented*: 90.4% *Outcomes*: quality of life pre THR and 3 and 12 months post THR (OHS and WOMAC, SF-12, SF-36 and EQ-5D scores)	Large effect sizes in disease-specific measures and the physical domains of the SF-12, SF-36 and EQ-5D. SF-36 and EQ-5D scores at 1 year after operation approached those of the general population	Use of the OHS and SF-12 is recommended in the assessment of THR. The SF-36 is useful in circumstances in which smaller changes in health status are investigated. The EQ-5D is useful in situations in which utility values are needed	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Parvizi 2010,[293] USA	*Type*: retrospective economic analysis *Aim*: to evaluate the in-hospital costs of periprosthetic joint infections caused by methicillin-resistant and methicillin-sensitive organisms	*Population*: patients treated for periprosthetic joint infection ($n = 391$), methicillin resistant ($n = 231$) and methicillin sensitive ($n = 160$) *Male*: resistant 47% ($n = 108$), sensitive 52% ($n = 83$) *Age (years), mean (SD)*: resistant 63.11 (12.65), sensitive: 61.64 (14.84) *Outcomes*: LOS, costs (US$)	LOS is significantly longer for patients with methicillin-resistant infections [mean 41.77 (SD 15.32) days] than for patients with methicillin-sensitive [mean 19.97 (SD 13.67) days] infections ($p = 0.0001$) Significantly higher cost of care for treatment of methicillin-resistant infections (mean US$107,264 per case) than for treatment of methicillin-sensitive infections (mean US$68,053 per case) ($p < 0.0001$)	More effective strategies for preventing the spread of infections caused by resistant organisms need to be implemented to ease the social and economic strains	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐
Pennington 2013,[44] UK	*Type*: retrospective economic (cost–utility) and decision analysis *Aim*: to evaluate the relative cost-effectiveness of cemented, cementless and hybrid prostheses for elective THR surgery	*Population*: patients undergoing primary THR for OA ($n = 30,203$ for quality-of-life analysis) *Male*: cemented 35.1% ($n = 4195$), cementless 44.6% ($n = 6548$), hybrid 38.0% ($n = 1350$) *Age (years), mean (SD)*: cemented 72.4 (6.7), cementless 67.8 (7.2), hybrid 70.4 (7.2) Outcomes: quality of life 6 months post surgery (OHS, EQ-5D score), lifetime cost-effectiveness, costs (£), ICERs *Economic model*: health service perspective; cost year 2010/11 (£); sensitivity analysis of QALYs post 2 years, revision rates using different hazard function, failed hip category without revision, excluding metal-on-metal prostheses	Lifetime costs: lowest with cemented prostheses Postoperative quality of life and lifetime QALYs: highest with hybrid prostheses Mean costs, women aged 70 years: cemented £6900, cementless £7800, hybrid £7500 Mean postoperative EQ-5D scores: cemented 0.78, cementless 0.80, hybrid 0.81 Lifetime QALYs: cemented 9.0, cementless 9.2, hybrid 9.3 years ICER: hybrid vs. cemented £2500	Cemented prostheses were the least costly type for THR. For most patient groups hybrid prostheses were the most cost-effective. Cementless prostheses did not provide sufficient improvement in health outcomes to justify their additional costs	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: initial costs (including prosthesis, operating theatre and hospital stay), utilities and revision rates; costs and utilities by sex, year group and prosthesis type

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Rana 2011,[277] USA	*Type*: retrospective cohort study *Aim*: to determine whether or not the use of alternative bearing surfaces alter a hospital's ability to profit or break even for THR	*Population*: patients with primary THR in 1990 ($n = 104$) and 2008 ($n = 269$) *Age (years), median (range)*: 1990: 68 (53–87), 2008: 69 (18–93) *OA*: 1990: 82% ($n = 85$), 2008: 88% ($n = 237$) *Outcomes*: costs (US$)	From 1990 to 2008, hospital payment for primary THR increased by 29% whereas inflation increased by 58% Lahey Clinic converted a US$3848 loss per primary THR in 1990 to a US$2359 profit per case in 2008 Reduction in LOS and implant costs were the most important drivers of expense reduction	If hospital revenue for THR decreases to managed Medicare levels, it will be difficult to make a profit on THR	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs as total hospital costs and broken down into operating room, hospital room, recovery room, medical and operative supply, anaesthesiology, pharmacy, laboratory, physical therapy and other costs
Rasch 2010,[305] Sweden	*Type*: prospective cohort study *Aim*: to examine the hypothesis that there is persisting muscular weakness in lower limb muscles and an impaired balance and gait 2 years after THR	*Population*: patients undergoing THR for OA ($n = 22$) *Male*: 18% *Age (years), mean (SD)*: 67 (7) *Outcomes*: quality of life (HHS, SF-36 score, EQ-5D score) pre THR and 6 and 24 months post THR, voluntary isometric strength, gait, postural stability	Hip muscles showed a remaining 6% weakness 2 years after THR Single-stance phase recovered at 6-months' follow-up Balance improved after operation Quality of life improved ($p < 0.001$), mean (range): HHS: 52 (34–65) to 86 (46–100); EQ-5D: 0.44 (0.03–0.69) to 0.85 (0.03–1.0); VAS: 5.2 (0–8) to 0.05 (0–1)	To accelerate improvement in muscle strength following THR, postoperative rehabilitation should be more intense and target hip abductors	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Rolfson 2011,[298] Sweden	*Type*: prospective cohort study *Aim*: to describe the development of the Swedish PROMs programme for patients with THR and analyse the response rate and investigate the pre-and 1-year post-THR PROMs data and examine how age, sex, diagnoses and comorbidities are associated with patient-reported outcomes	*Population*: patients with THR ($n = 34,960$) *Male*: 42% ($n = 14,740$) *Age (years), mean (SD)*: 68.1 (10.4) *OA*: 93.2% *Outcomes*: quality of life (EQ-5D score, VAS) pre and 12 months post THR	1-year post THR mean EQ-5D score increased to above the level of an age- and gender-matched population and resulted in reduction of pain ($p < 0.001$) Pre- to post-THR mean (SD) difference in EQ-5D: 0.37 (0.35) Less improvement in quality of life in male, older and Charnley C category patients ($p < 0.001$)	Patients' response rates were good. PROMs data permit improvement work and allow for health economic evaluation	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: EQ-5D data by age and sex
Rolfson 2009,[306] Sweden	*Type*: prospective cohort study *Aim*: to examine the hypothesis that anxiety/depression is a significant variable in predicting satisfaction and pain relief after THR	*Population*: patients with THR for hip OA ($n = 6158$) *Male*: 43% ($n = 2652$) *Outcomes*: quality of life (EQ-5D score, VAS) pre and 12 months post THR, anxiety and pain	Preoperative EQ-5D anxiety/depression dimension was a strong predictor for pain relief and patient satisfaction ($p < 0.001$)	Mental health may influence postoperative quality of life and pain. Appropriate assessment of mental health may support patient management and optimise THR outcome	1. (a) Resource use ☐ (b) Costs ☐ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: EQ-5D scores by sex and preoperative anxiety

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Scheerlinck 2004,[276] Belgium	*Type*: prospective cohort study *Aim*: to assess and identify factors influencing hospital costs, hospital stay and the hospital discharge policy related to implantation of a THR in a Belgian university hospital	*Population*: patients scheduled for an elective primary THR ($n = 102$) *Male*: 40% *Age (years), mean (SD)*: 70.3 (9.6) *Outcomes*: major complications, costs (€), LOS, quality of life (SF-12 score, WOMAC score, HHS, Merle-d'Aubigné and Postel score) pre THR and 6 weeks and 3, 6 and 12 months post THR	Average LOS 14.4 days; average hospital cost €9500 Hospitalisation represented > 50% of the hospital cost and hip implants between 16.1% and 25.6% depending on prosthesis type. Complications and discharge to a rehabilitation unit increased hospital stay and cost 6 months after surgery, functional hip scores as well as WOMAC and mental and physical SF-12 scores improved significantly	Surgical techniques and faster rehabilitation programmes are probably the best ways to control the cost of THR in Belgium	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs not broken down; given as percentage of overall hospital costs
Segal 2004,[269] Australia	*Type*: retrospective economic (cost–utility) analysis *Aim*: to identify all potential interventions for preventing and managing OA, to conduct economic evaluations of the selected interventions and to compare the cost-effectiveness of the interventions	*Population*: patients with OA *Outcomes*: utility benefits (SF-36), cost (AU$) per individual, cost (AU$) per QALY for surgery, use of NSAIDs, primary prevention, management *Evidence-based priority-setting model*: time horizon 15 years for THR and total knee replacement (TKR); management 1–3 years, primary care 20 years; societal perspective; univariate sensitivity analyses investigating outcomes, period of benefit and discounting	THR and TKR highly cost-effective: cost per QALY: AU$7500 for THR and AU$10,000 for TKR Exercise and strength training for knee OA < AU$5000 per QALY; knee bracing and use of capsaicin or glucosamine sulphate < AU$10,000 per QALY	There was strong evidence of effectiveness and favourable cost–utility ratios for THR, TKR, intensive clinic-based exercise and strength training, and knee bracing Identified need for better targeting of cyclo-oxygenase-2 (COX-2) NSAIDs to subpopulations Arthroscopy for knee OA may be a poor use of resources	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: mean QALY gain per person and mean cost per person of total THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Stargardt 2008,[280] EU countries (multicentre)	*Type*: prospective observational study *Aim*: to assess variations in the cost of primary THR between and within nine member states of the EU and to compare the cost of service with public-payer reimbursements	*Population*: patients with hip OA requiring primary THR *Age (years), range*: 65–75 *Outcomes*: cost per treatment *Cost year*: used average exchange rates from 2005 (€)	Total cost of treatment ranged from €1290 (Hungary) to €8739 (the Netherlands). Mean cost was €5043 (SD €2071) Main cost drivers: implants (34% of total cost on average), ward costs (20.9% of total cost on average) 74.0% of variation was between countries and only 26% of variation was within countries Purchasing-power parities explained 79.4% of the explainable between-country variation Percentage of uncemented implants used and the number of beds explained 12.1% and 1.6% of explainable within-country variation, respectively	The large differences in costs and reimbursement between Poland, Hungary and the other EU member states shows that primary THR is a highly relevant case for cross-border care	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs per country and broken down into costs for staff, implant, material, drugs and overheads
Straumann 2006,[256] Switzerland	*Type*: retrospective economic (cost–minimisation) analysis (study says cost–benefit analysis) *Aim*: to show model-based economic consequences of MIS THR	*Population*: patients with THR ($n = 13,101$) *Outcomes*: total cost (€) per standard THR *Model*: societal perspective	Calculated cost-savings for MIS THR were 17.2% Cost savings ranged between €7.8M (30%, conservative) and €12.0M (50%, optimistic) resulting from shorter LOS	MIS THR techniques may allow a reduction in health-care costs	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: average costs broken down into operation (with definition), care lodging, therapy and infrastructure/ administration costs

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Tien 2009,[270] Taiwan (Province of China)	*Type*: retrospective cohort study *Aim*: to explore the increasing prevalence of factors affecting hospital charges for primary THR and total knee replacement (TKR)	*Population*: patients with THR ($n = 39,569$) or TKR ($n = 76,727$) *Male*: THR 60%, TKR 30% *Age*: THR: 65% < 64 years; TKR: 70% > 65 years *OA*: THR: 39.52–42.52% across three time periods; TKR: 91.37–94.54% across three time periods *Outcomes*: average LOS, costs (US$) [cost year 2004 Taiwan currency and converted to US$ (rate 31.5 : 1], differences in LOS and costs across three time periods (1996–8, 1999–2001, 2002–4), predictors for hospital charges	THR: average LOS declined from 11.82 days to 8.96 days (–24.24%) and total hospital charges decreased from US$4523.47 to US$3960.06 (–12.46%) TKR: average LOS decreased from 12.83 days to 9.07 days (–29.31%) and hospital charges decreased from US$5063.17 to US$4330.50 (–14.47%) Factors associated with increased hospital charges: age < 65 years, increased disease severity, absence of primary diagnoses of OA, RA or avascular necrosis, treatment at a hospital or by a surgeon performing a high volume of operations and longer average LOS	Prevalence of THR and TKR has increased annually; however, average LOS has decreased dramatically in the same time period. Factors that increase hospital charges must be carefully managed	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs are total hospital charges (no definition)

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Tso 2012,[307] Canada	*Type*: economic (cost–utility) analysis of observational matched-cohort study (outcomes prospective, costs retrospective) *Aim*: to compare the lifetime ICERs for decompression and decompression with fusion with those for THR and TKR	*Population*: patients with THR and TKR for OA and patients with decompression with/without fusion for spinal stenosis *Male*: 40% *Age (years), mean*: spinal stenosis 64.2, THR 63.0, TKR 64.6 *Outcomes*: costs, QALYs (using SF-36), ICERs *Economic analysis*: direct costs included; provincial health insurance perspective; cost year 2009 (CA$); lifetime time horizon; single and multiway sensitivity analyses including utilities, cost of primary surgery, revision rate and discount rate	The lifetime ICERs discounted at 3% were CA$5321 per QALY for THR, CA$11,275 per QALY for TKR, CA$2307 per QALY for spinal decompression and CA$7153 per QALY for spinal decompression with fusion Sensitivity analyses did not alter the ranking of the lifetime ICERs	In appropriately selected patients who have failed medical management, the lifetime ICER for surgical treatment of lumbar spinal stenosis is similar to those for THR and TKR for the treatment of OA	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☒ (b) QALYs ☒ 3. Transition probabilities ☐ Comment: costs for THR broken down into surgical costs (and further into operating room, nursing, medical imaging, laboratory, pharmacy and allied health costs), post-discharge rehabilitation costs and revision costs
Tuominen 2009,[283] Finland	*Type*: RCT *Aim*: to identify the effects of waiting time on health and health-related quality of life outcomes and the use and costs of disease-specific medication among two patient groups: SWT and NFWT	*Population*: patients undergoing THR for hip OA ($n = 309$); SWT $n = 140$ and NFWT $n = 169$ *Male*: 42% *Age (years), mean (SD)*: 65 (9.9) *Outcomes*: quality of life (HHS, 15D) at time of placement on the waiting list, at admission and at 3 and 12 months post THR; QALYs; costs of medication (€)	Mean waiting time (days): SWT 74, NFWT 194 Intention-to-treat analyses: no statistically significant differences between the groups in the weekly use and costs of medication, health-related quality of life or HHS at baseline, at admission or at 3 or 12 months after surgery Total medication costs during the waiting time period: SWT €83, NFWT €171 ($p < 0.001$) SWT resulted in a gain of 0.028 QALYs	The length of the waiting time did not generate different effects on the studied health and quality of life outcomes. Those in the short waiting time group reached a better health-related quality of life earlier	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: costs of medication use at baseline, admission and 3 and 12 months post THR

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Urquhart 2008,[290] Australia	*Type*: retrospective cohort study *Aim*: to examine the in-hospital outcomes and resource utilisation of primary and revision hip replacement	*Population*: episodes of care for hip replacement procedures ($n = 7724$): 86.8% primary procedures, of which 79.3% were THRs Male: partial: 24.8%, THR: 43%, revision: 40.6% *OA*: 87.7% *Outcomes*: LOS, admission to intensive care unit or coronary care unit, in-hospital mortality, discharge location	Revision vs. primary THR: 22.9% more revisions remained in hospital for more than a week ($p < 0.0001$), 14.6% more required intensive care ($p < 0.0001$) and 10.9% less were discharged to a private residence ($p < 0.001$) Primary partial and revision replacements utilised up to 27.5% and 34.6% of hospital resources, respectively	Partial and revision hip replacements are resource intensive Incidences need to be reduced, as failure to do so will have important implications for the allocation of health-care funding	1. (a) Resource use ☒ (b) Costs ☐ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: LOS in days only for partial hip replacement. THR and revision unit is > 1 week
Vale 2002,[19] McKenzie 2003,[299] UK	*Type*: systematic review and retrospective economic (cost–utility) analysis Aim: to compare the effectiveness and cost-effectiveness of metal-on-metal RS with that of watchful waiting, THR, osteotomy, arthrodesis and arthroscopy of the hip joint	*Population*: patients with hip disease *Age (years)*: 45–50 and 65–70 years *Outcomes*: Costs (£), QALYs, ICERs *Economic model*: Markov model; 20-year time horizon; NHS perspective; cost year 2000 (£); subgroup analysis considering those who would not outlive a THR; sensitivity analyses for revision rates, operation times, watchful waiting costs, time horizon and quality of life	Revision: RS over 3-year follow-up: 0–14%; THR over 10-year follow-up: ≤ 10%; osteotomy over 10- to 17-year follow-up: between 2.9% and 29% Patients pain free: RS: 91% at 4 years; THR: 84% at 11 years; arthrodesis: 22% at 8 years Costs: RS for a patient aged < 65 years £5515; THR £4195; revision £6027; arthroscopy £951; osteotomy £2731; watchful waiting £642 annually Cost-effectiveness: for patients aged < 65 years RS dominated by THR. RS dominated watchful waiting within a 20-year follow-up. Incremental cost per QALY: RS vs. osteotomy £3039, RS vs. arthroscopy £366. For patients aged > 65 years, THR dominated RS	RS had lower revision rates than THR over an extended time period and resulted in better outcomes overall for those who are likely to outlive a primary THR If RS has lower revision rates than THR over an extended period and results in better outcomes from subsequent THR, then RS could possibly be considered cost-effective or even dominant	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☒ 3. Transition probabilities ☒ Comment: revision rates for RS and THR; costs including prosthesis costs; costs broken down for watchful waiting

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Vanhegan 2012,[292] UK	*Type*: retrospective economic analysis *Aim*: to evaluate the costs associated with revision THR for different indications	*Population*: patients with revision THR ($n = 286$; $n = 305$ procedures) *Male*: aseptic loosening ($n = 194$) 34% ($n = 65$), deep infection ($n = 76$) 42% ($n = 32$), periprosthetic fracture ($n = 24$) 25% ($n = 6$), dislocation ($n = 11$) 28% ($n = 3$) *Age (years), mean (range)*: aseptic loosening 67 (20–89), deep infection 62 (29–83), periprosthetic fracture 76 (31–88), dislocation 79 (54–90) *OA*: aseptic loosening 69%, deep infection 48%, periprosthetic fracture 80%, dislocation 54% *Outcomes*: LOS, costs (£)	Mean (SD) total costs for revision surgery: aseptic cases £11,897 (£4629), septic revision £21,937 (£10,965), periprosthetic fracture £18,185 (£9124), dislocation £10,893 (£5476) Surgery for infection and periprosthetic fracture: longer operating times, increased blood loss, increase in complications, longer LOS	Financial costs vary significantly by indication. Variation is not reflected in current NHS tariffs	1. (a) Resource use ☒ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Xie 2007,[308] Singapore	*Type*: retrospective cross-sectional study *Aim*: to estimate and compare the direct and indirect costs of OA in multiethnic Asian patients with OA in Singapore	*Population*: patients with OA ($n = 1179$); no surgery $n = 574$, THR $n = 92$ and TKR $n = 513$ *Male*: 24.7% *Age (years), mean (SD)*: 86.1 (12.0) *Outcomes*: costs *Economic analysis*: societal and patient perspective, direct and indirect costs considered, cost year 2003 (SG$)	Mean direct costs: no surgery: to society SG$3245, to patient SG$1459; TKR: to society SG$11,429, to patient SG$5561; THR: to society SG$15,763, to patient SG$7555 Direct costs to patients (range): Chinese SG$1460–7477, Malays SG$1362–7211, Indians SG$1688–6226, other ethnic groups SG$1437–12,140 Direct costs to society (range): Chinese SG$3351–15,799, Malays SG$2939–15,436, Indians SG$3150–10,990, other ethnic groups SG$2597–17,879 Indirect costs (range): Chinese SG$1215–3834, Malays SG$1138–6116, Indians SG$1371–5292 However, most ethnic variations were not statistically significant	The economic burden of OA to society and patients increased by threefold or more in the patients with TKR/THR compared with those without. Ethnic differences in health resources consumed were more apparent when the disease progressed	1. (a) Resource use ☐ (b) Costs ☒ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: unit cost data for THR (broken down into professional, treatment and procedure, equipment and material, inpatient, diagnosis, medication and miscellaneous costs)

Study and country	Study design	Methods	Results	Main conclusion	Usefulness for model/comments
Zhang 2012,[279] China	*Type*: retrospective cohort study *Aim*: to examine the costs of unilateral and bilateral hip arthroplasties as well as unilateral and bilateral knee arthroplasties performed at Beijing Jishuitan Hospital in 2008 and 2009	*Population*: patients with primary hip or knee replacement ($n = 1009$); unilateral THR $n = 331$, unilateral TKR $n = 407$, bilateral THR $n = 88$, bilateral TKR $n = 183$ *Male*: 29.7% ($n = 300$) *Age (years), mean*: 56.7 *Outcomes*: costs (cost year 2009 yuan)	The prosthesis and pharmacy charges at this institution accounted for approximately 80% of the total charges (total charges unilateral THR 55,813 yuan, unilateral TKR 50,580 yuan, bilateral THR 94,299 yuan, bilateral TKR 89,186 yuan) Labour costs in China accounted for a lower percentage of total charges in comparison to North America and Taiwan	Because the percentage of costs covered by medical insurance was relatively low, a substantial financial burden was imposed on patients, which may limit access to joint arthroplasty in China	1. (a) Resource use ☒ (b) Costs ☐ 2. (a) Utilities ☐ (b) QALYs ☐ 3. Transition probabilities ☐ Comment: mean hospital stay, total costs for THR in yuan, costs broken down as % of total

DKK, Danish Krone; FC-2, The Wrightington Frusto-Conical 2; MI, minimally invasive; MIS, minimally invasive surgery; NFWT, non-fixed waiting time > 3 months; NSH, North Shore Hospital; SF-6D, Short Form questionnaire-6 Dimensions; SG, standard gamble; STD THR, standard THR; SWT, short waiting time ≤ 3 months; TKR, total knee replacement; UKR, unicompartmental knee replacement.

Appendix 11 Key studies reporting costs for resurfacing arthroplasty

Resurfacing arthroplasty

Study and country	Cost data	Definition	Source	Cost year (currency)	Comment
Edlin 2012,[40] UK	Initial operation/care: £6275 (SD £557) LOS: 5.7 days NHS and social care costs: £7217 (SD £1320) Implant plus operative consumables: £1826	Operation including implant and LOS Cost to NHS per RS	Implant costs: finance department University Hospitals Coventry and Warwickshire NHS Trust Non-prosthesis average cost: national-level HRG4 frequencies for primary hip replacements used to calculate an average cost, average LOS and average cost per excess bed-day. Non-prosthesis average cost obtained by deducting the expected RA cost from the average cost LOS: hospital records Other cost data: patient-reported data	2009/10 (£)	Cormet™ metal-on-metal resurfacing (Corin Group, Cirencester, UK) $n = 58$
Vale 2002,[19] McKenzie 2003,[299] UK	Implant: £1730 (Midland Medical Technologies, Birmingham, UK) and £1890 (Wright Cremascoli Ortho, Memphis, TN, USA) Total cost per patient per RS: 10 days: £5396; 8 days: £4947.09 LOS: 10 days: £2244.47; LOS: 8 days: £1795.58 Theatre overheads (134 minutes): £731.40 Theatre staff: £265.37 Radiography: £149.63 Outpatient visits: £281.40		Industry estimate Costs other than prosthesis: published cost estimates	2000 (£)	Acetabular cup and femoral head set (Midland Medical Technologies); Conserve Plus® femoral head and acetabular component (Wright Cremascoli Ortho)
Bozic 2010,[253] USA	RS: US$17,178 (SD US$2191)	Hospital and professional fees	Hospital costs: average Medicare payments Implant: published resources	2008 (US$)	NA

NA, not applicable.

Successful resurfacing arthroplasty follow-up

Study and country	Cost data	Definition	Source	Year/currency
Edlin 2012,[40] UK	NHS plus social care costs minus initial operation/care: £942	At 12 months: includes subsequent inpatient care, outpatient care, primary/community care, aids and adaptations paid for by the NHS and medication	Patient-reported data for resource use, converted into costs using NHS reference costs, Personal Social Services Research Unit (PSSRU) unit costs, NHS electronic drug tariff and previous studies for unit costs of acupuncture and chiropractic	2009/10 (£)
Vale 2002[19] (McKenzie 2003[299]), UK	First-year follow-up: £118 (same as for THR and revision THR)	Includes two outpatient visits with one radiograph	Published cost estimates	2000 (£)

Resurfacing arthroplasty revision

Study and country	Cost data	Definition	Source	Year/currency	Comment
Bozic 2010,[253] USA	RS to THR: US$18,460 (SD US$2335) Major total revision: US$21,195 (SD US$2704) Major partial revision: US$18,155 (SD US$2316) Minor revision: US$16,367 (SD US$2088)	Conversion of RS to THR	Hospital costs: average Medicare payments Implant: published resources	2008 (US$)	No distinction made between revision following THR and revision following RS

Appendix 12 Key studies reporting costs for total hip replacement

Total hip replacement

Study and country	Cost data	Definition	Source	Cost year (currency)
Pennington 2013,[44] UK	*Initial cost*: Men: Age 60 years: cemented £5996, cementless £6811, hybrid £6610 Age 70 years: cemented £6096, cementless £6919, hybrid £6711 Age 80 years: cemented £6459, cementless £7227, hybrid £6989 Women: Age 60 years: cemented £6065, cementless £6882, hybrid £6694 Age 70 years: cemented £6193, cementless £7018, hybrid £6822 Age 80 years: cemented £6581, cementless £7351, hybrid £7126	Prosthesis, operating theatre and hospital stay	Prosthesis: prices paid by NHS provider (lower than list prize) Operation theatre and hospital stay: national tariff modified by study data on LOS by prosthesis type	2010/11 (£)
Edlin 2012,[40] Costa 2012,[130] UK	Initial operation/care: £6091 (SD £532) LOS: 5.5 days NHS and social care costs: £6653 (SD £917) Implant plus consumables: ceramic–ceramic £2042, metal–metal £1625, metal–polyurethane £843	Operation including implant and LOS Cost to NHS per THR	Implant costs: finance department, University Hospitals Coventry and Warwickshire NHS Trust Non-prosthesis average cost: identified that there are national-level HRG4 frequencies for primary hip replacements and that these are used to calculate an average cost, average LOS and average cost per excess bed-day. By deducting the expected THR cost from the average cost, we obtain a non-prosthesis average cost LOS: hospital records Other cost data: patient-reported data	2009/10 (£)

Study and country	Cost data	Definition	Source	Cost year (currency)
Vale 2002,[19] McKenzie 2003,[299] UK	Prosthesis cost: Charnley £341.70, cement £68.12 Total cost per patient per THR: 10 days: £4075.78; 12 days: £4456.56 LOS: 10 days: £2244.47; 8 days: £2693.37 Theatre overheads (134 minutes): £731.40 Theatre staff: £265.37 Radiography: £149.63 Outpatient visits: £281.40	NA	Published cost estimates (previous HTA report[373])	2000 (£)
	Prosthesis cost: Charnley £306.00, cement £61.00 Total cost per patient per THR: £4052.00 LOS: 12 days: £2412.00 Theatre overheads (134 minutes): £655.00 Theatre staff: £232.00 Radiography: £134.00 Outpatient visits: £252.00			1996 (£)

NA, not applicable.

Successful total hip replacement follow-up

Study and country	Cost data	Definition	Source	Year/currency
Edlin 2012,[40] UK	NHS plus social care costs minus initial operation/care: £562	At 12 months: includes subsequent inpatient care, outpatient care, primary/community care, aids and adaptations paid for by the NHS and medication	Patient-reported data for resource use, converted into costs using NHS reference costs, Personal Social Services Research Unit (PSSRU) unit costs, NHS electronic drug tariff and previous studies for unit costs of acupuncture and chiropractic	2009/10 (£)
Vale 2002,[19] McKenzie 2003,[299] UK	First-year follow-up: £118.74	Includes two outpatient visits and one radiograph	Published cost estimates	2000 (£)

Total hip replacement revision

Study and country	Cost data		Definition	Source	Year/currency
Pennington 2013,[44] UK	Men: Age 60 years: cemented £8167, cementless £8748, hybrid £8726 Age 70 years: cemented £6912, cementless £7712, hybrid £7516 Age 80 years: cemented £6819, cementless £7690, hybrid £7481 Women: Age 60 years: cemented £7864, cementless £8551, hybrid £8487 Age 70 years: cemented £6937, cementless £7704, hybrid £7486 Age 80 years: cemented £6853, cementless £7762, hybrid £7521		Replacement-related costs	NA	2010/11 (£)
Vanhegan 2012,[292] UK	*Aseptic loosening (n = 194)*:		NA	National tariff	2007/8 (£)
	Mean (SD) inpatient stay (days)	9.3 (8.6)			
	Mean (SD) operative time (minutes)	173 (51)			
	Mean (SD) investigation costs (£)	342 (78)			
	Mean (SD) drug costs (£)	200 (33)			
	Mean (SD) implant costs (£)	2298 (1320)			
	Mean (SD) theatre costs (£)	1216 (256)			
	Mean (SD) total cost (£)	11,897 (4629)			
	Deep infection (n = 76):				
	Mean (SD) inpatient stay (days)	16.8 (22.3)			
	Mean (SD) operative time (minutes)	183.3 (71)			
	Mean (SD) investigation costs (£)	988 (212)			
	Mean (SD) drug costs (£)	854 (175)			
	Mean (SD) implant costs (£)	3345 (2183)			
	Mean (SD) theatre costs (£)	1744 (1002)			
	Mean (SD) total cost (£)	21,937 (10,965)			
	Periprosthetic fracture (n = 24):				
	Mean (SD) inpatient stay (days)	17.1 (17.8)			

Study and country	Cost data		Definition	Source	Year/currency
	Mean (SD) operative time (minutes)	193 (79)			
	Mean (SD) investigation costs (£)	394 (143)			
	Mean (SD) drug costs (£)	320 (69)			
	Mean (SD) implant costs (£)	3123 (1613)			
	Mean (SD) theatre costs (£)	1315 (394)			
	Mean (SD) total cost (£)	18 185 (9124)			
	Dislocation (n = 11):				
	Mean (SD) inpatient stay (days)	9.1 (4.2)			
	Mean (SD) operative time (minutes)	89 (39)			
	Mean (SD) investigation costs (£)	369 (161)			
	Mean (SD) drug costs (£)	220 (48)			
	Mean (SD) implant costs (£)	2890 (1540)			
	Mean (SD) theatre costs (£)	1280 (240)			
	Mean (SD) total cost (£)	10 893 (5476)			
	Implant costs: Range £750–3000				
Vale 2002,[19] McKenzie 2003,[299] UK	*Revision*:		NA	Published cost estimates (previous HTA report[368])	2000 (£)
	Prosthesis: Charnley £754.86, cement £136.23				
	Theatre overheads (195 minutes): £1065.29				
	Total cost per patient per revision: £5908.21				
	LOS: £3142.26				
	Theatre staff: £378.55				
	Radiography: £149.63				
	Outpatient visits: £281.40				
	Revision:				1996 (£)
	Prosthesis: Charnley £676.00, cement £122.00				
	Theatre overheads (195 minutes): £954.00				
	Total cost per patient per revision: £5291.00				
	LOS (14 days): £2814.00				
	Theatre staff: £339.00				
	Radiography: £134.00				
	Outpatient visits: £252.00				

NA, not applicable.

Appendix 13 Excluded cost-effectiveness papers and reasons for exclusion

Note: excluded papers are those excluded at full-text stage or those that were unavailable ($n = 35$).

Excluded papers	Reason for exclusion
Bellamy N, Hendrikz J, Wilson C. Comparison of transformed visual analogue and native numerical rating scaled patient responses to the WOMAC® index. *Intern Med J* 2011;**41**(Suppl. s1):23	Exclude – WOMAC scores only
Bernath V. *Hip Resurfacing in Patients with Osteoarthritis*. Clayton, Victoria: Centre for Clinical Effectiveness; 2002	Exclude – unavailable
Bertin P, Grange L, Rannou F, Taieb C. The cost of hospitalization of hip osteoarthritis patients in France in 2010. *Osteoarthritis Cartilage* 2012;**20**(Suppl. 1):S189	Exclude – not THR, only rehabilitation costs
BlueCross BlueShield Association. Metal-on-metal total hip resurfacing. Chicago, IL: BlueCross BlueShield Association; 2007.	Exclude – no cost or utility data
Bourne RB, Corten K. Cemented versus cementless stems: a verdict is in. *Orthopedics* 2010;**33**:638	Exclude – no cost data
Bozic KJ, Chiu V, Slover JD, Immerman I, Kahn JG. Patient preferences and willingness to pay for arthroplasty surgery in patients with osteoarthritis of the hip. *J Arthroplasty* 2012;**27**:503–6	Exclude – hip/knee total joint replacement not reported separately
Bozic KJ, Stacey B, Berger A, Sadosky A, Oster G. Resource utilization and costs before and after total joint arthroplasty. *BMC Health Serv Res* 2012;**12**:73	Exclude – WTP/patient preferences
Collins NJ, Roos EM. Patient-reported outcomes for total hip and knee arthroplasty: commonly used instruments and attributes of a 'good' measure. *Clin Geriatr Med* 2012;**28**:367–94	Exclude – usefulness of instruments reviewed
Dobson RL, Osenenko K, Szabo SM, Roy S, Cifaldi M, Maksymowych WP, *et al.* Frequency and cost of joint replacement surgery for patients with rheumatoid arthritis: a population-based study in Canada. *Arthritis Rheum* 2009;**60**(Suppl. 10):718	Exclude – abstract only – only total cost for hip replacement in RA
Ethgen O, Tancredi A, Lejeune E, Kvasz A, Zegels B, Reginster JY. Do utility values and willingness to pay suitably reflect health outcome in hip and knee osteoarthritis? A comparative analysis with the WOMAC index. *J Rheumatol* 2003;**30**:2452–9	Exclude – OA not THR or RS
Geissler A, Scheller-Kreinsen D, Quentin W, EuroDRG group. Do diagnosis-related groups appropriately explain variations in costs and length of stay of hip replacement? A comparative assessment of DRG systems across 10 European countries. *Health Econ* 2012;**21**(Suppl. 2):103–15	Exclude – regression analysis to estimate costs and LOS
Ghoz A, Macdonald D. (iii) New trends in total hip replacement: follow-up is it required and who pays? *Curr Orthop* 2008;**22**:173–6	Exclude – costs too generalised/opinion
Gonzalez Saenz de Tejada M, Escobar A, Herrera C, Garcia L, Aizpuru F, Sarasqueta C. Patient expectations and health-related quality of life outcomes following total joint replacement. *Value Health* 2010;**13**:447–54	Exclude – hip/knee data not reported separately
Grammatico-Guillon L, Baron S, Gettner S, Lecuyer AI, Gaborit C, Rosset P, *et al.* Bone and joint infections in hospitalized patients in France, 2008: clinical and economic outcomes. *J Hosp Infect* 2012;**82**:40–8	Exclude – hip data not reported separately
Hartl A, Schillinger M, Wanivenhaus A. Cemented versus cementless total hip arthroplasty for osteoarthrosis and other non-traumatic diseases. *Cochrane Database Syst Rev* 2004;**3**:CD004850	Exclude – protocol, no costs or utility outcomes reported
Hawker GA, Badley EM, Croxford R, Coyte PC, Glazier RH, Guan J, *et al.* A population-based nested case–control study of the costs of hip and knee replacement surgery. *Med Care* 2009;**47**:732–41	Exclude – hip/knee data not reported separately

Excluded papers	Reason for exclusion
Healy WL, Iorio R. Implant selection and cost for total joint arthroplasty: conflict between surgeons and hospitals. *Clin Orthop Relat Res* 2007;**457**:57–63	Exclude – hip/knee data not reported separately, no costs or utilities data in review
Kim S. Changes in surgical loads and economic burden of hip and knee replacements in the US: 1997–2004. *Arthritis Rheum* 2008;**59**:481–8	Exclude – only total costs and no utilities, hip/knee data not reported separately
Kim S, Koebel S, Duffy R. Cost burden of hip and knee replacements in Ohio: estimates from the National Hospital Discharge Survey, 2000. *Managed Care Interface* 2004;**17**:22–5	Exclude – total costs and not broken down by resource use and unit costs
Krummenauer F, Gunther KP, Witzlebf WC. The incremental cost effectiveness of in-patient versus out-patient rehabilitation after total hip arthroplasty – results of a pilot investigation. *Eur J Med Res* 2008;**13**:267–74	Exclude – matched-pair analysis and differences between groups (in- and outpatients)
Lavernia CJ, Alcerro JC. Quality of life and cost-effectiveness 1 year after total hip arthroplasty. *J Arthroplasty* 2011;**26**:705–9	Exclude – quality of well-being and quality of well-year not QALYs; costs not broken down
Lavernia CJ, D'Apuzzo MR, Hernandez VH, Lee DJ, Rossi MD. Postdischarge costs in arthroplasty surgery. *J Arthroplasty* 2006;**21**:144–50	Exclude – no costs per QALYs
Loza E, Lopez-Gomez JM, Abasolo L, Maese J, Carmona L, Batlle-Gualda E. Economic burden of knee and hip osteoarthritis in Spain. *Arthritis Rheum Arthritis Care Res* 2009;**61**:158–65	Exclude – hip/knee data not reported separately
March LM, Barcenilla AL, Cross MJ, Lapsley HM, Parker D, Brooks PM. Costs and outcomes of total hip and knee joint replacement for rheumatoid arthritis. *Clin Rheumatol* 2008;**27**:1235–42	Exclude – RA not OA
Martineau P, Filion KB, Huk OL, Zukor DJ, Eisenberg MJ, Antoniou J. Primary hip arthroplasty costs are greater in low-volume than in high-volume Canadian hospitals. *Clin Orthop Relat Res* 2005;(437):152–6	Exclude – costs not broken down, no average costs given, costs are based on low-volume and high-volume hospitals
Piscitelli P, Iolascon G, Di Tanna G, Bizzi E, Chitano G, Argentiero A, *et al.* Socioeconomic burden of total joint arthroplasty for symptomatic hip and knee osteoarthritis in the Italian population: a 5-year analysis based on hospitalization records. *Arthritis Care Res* 2012;**64**:1320–7	Exclude – total costs not broken down
Rampersaud YR, Tso P, Walker K, Eagen B, Lewis S, Gandhi R, *et al.* Assessment of the incremental cost-utility of surgery compared to medical management for the treatment of hip, knee, and spine osteoarthritis. *Spine J* 2010;**10**(9 Suppl. 1):S35	Exclude – not relevant, off topic (spinal decompression)
Rasanen P, Paavolainen P, Sintonen H, Koivisto AM, Blom M, Ryynanen OP, *et al.* Effectiveness of hip or knee replacement surgery in terms of quality-adjusted life years and costs. *Acta Orthop* 2007;**78**:108–15	Exclude – costs not broken down
Rat A, Baumann C, Osnowycz G, Mainard D, Delagoutte J, Cuny C, *et al.* Clinically relevant change and patient acceptable quality of life after total hip or knee replacement surgery for osteoarthritis. *Ann Rheum Dis* 2006;**65**:599–600	Exclude – hip/knee data not reported separately, not EQ-5D data for quality of life
Rat AC, Baumann C, Osnowycz G, Mainard D, Delagoutte JP, Cuny C, *et al.* Total hip or knee replacement surgery for osteoarthritis: medium and long term quality of life. *Arthritis Rheum* 2008;**58**:S630–1	Exclude – no utility data
Rolfson O, Strom O, Karrholm J, Malchau H, Garellick G. Costs related to hip disease in patients eligible for total hip arthroplasty. *J Arthroplasty* 2012;**27**:1261–6	Exclude – costs relate to hip disease patients 12 months before surgery
Taieb C. The cost of hospitalization of hip osteoarthritis patients in France in 2010. *Value Health* 2012;**15**:A35	Exclude – not THR, only rehabilitation costs
Uddin S, Hossain L, Kelaher M. Effect of physician collaboration network on hospitalization cost and readmission rate. *Eur J Public Health* 2012;**22**:629–33	Exclude – off-topic, not cost-effectiveness, looks at network analysis
Woo J, Lau E, Lau CS, Lee P, Zhang J, Kwok T, *et al.* Socioeconomic impact of osteoarthritis in Hong Kong: utilization of health and social services, and direct and indirect costs. *Arthritis Rheum Arthritis Care Res* 2003;**49**:526–34	Exclude – not THR, hip and knee data not reported separately

Appendix 14 Key studies reporting utilities for total hip replacement

Total hip replacement

Study and country	Utility data, mean (SD)	Definition	Source	Patient characteristics [THR (*n*) and OA]
Pennington 2013,[44] UK	*Pre THR*: Cemented 0.34 (0.32), cementless 0.36 (0.32), hybrid 0.34 (0.32) *Post THR, 6 months*: Male: Age 60 years: cemented 0.797, cementless 0.807, hybrid 0.810 Age 70 years: cemented 0.819, cementless 0.836, hybrid 0.848 Age 80 years: cemented 0.797, cementless 0.804, hybrid 0.824 Female: Age 60 years: cemented 0.785, cementless 0.787, hybrid 0.800 Age 70 years: cemented 0.781, cementless 0.799, hybrid 0.805 Age 80 years: cemented 0.754, cementless 0.749, hybrid 0.751	EQ-5D Gene matched and regression to allow for observed preoperative differences	Patient reported (PROMs in England)	*n* = 30,203 OA patients
Rolfson 2011,[298] Sweden	*Pre THR*: Male: 0.45 (0.31), female: 0.37 (0.32) *Post THR, 12 months*: Male: 0.81 (0.23), female: 0.76 (0.25) *Pre THR by age*: < 30 years: female (*n* = 32): 0.24 (0.30), male (*n* = 25): 0.25 (0.30) 30–39 years: female (*n* = 159): 0.28 (0.32), male (*n* = 158): 0.39 (0.32) 40–49 years: female (*n* = 611): 0.29 (0.31), male (*n* = 688): 0.40 (0.31) 50–59 years: female (*n* = 2533): 0.33 (0.31), male (*n* = 2428): 0.43 (0.31)	EQ-5D	Patient reported (PROMs in Sweden)	*n* = 34,960 OA: 93.2%

Study and country	Utility data, mean (SD)	Definition	Source	Patient characteristics [THR (*n*) and OA]
	60–69 years: female (*n* = 6542): 0.38 (0.32), male (*n* = 5142): 0.46 (0.30) 70–79 years: female (*n* = 7205): 0.39 (0.31), male (*n* = 4760): 0.48 (0.30) 80–89 years: female (*n* = 3053): 0.35 (0.32), male (*n* = 1491): 0.45 (0.31) ≥ 90 years: female (*n* = 85): 0.23 (0.31), male (*n* = 48): 0.37 (0.31) *Post THR by age*: < 30 years: female (*n* = 32): 0.72 (0.26), male (*n* = 25): 0.69 (0.23) 30–39 years: female (*n* = 159): 0.74 (0.32), male (*n* = 158): 0.77 (0.26) 40–49 years: female (*n* = 611): 0.76 (0.27), male (*n* = 688): 0.83 (0.25) 50–59 years: female (*n* = 2533): 0.75 (0.27), male (*n* = 2428): 0.80 (0.25) 60–69 years: female (*n* = 6542): 0.78 (0.24), male (*n* = 5142): 0.82 (0.22) 70–79 years: female (*n* = 7205): 0.76 (0.24), male (*n* = 4760): 0.81 (0.22) 80–89 years: female (*n* = 3053): 0.72 (0.25), male (*n* = 1491): 0.76 (0.23) ≥ 90 years: female (*n* = 85): 0.65 (0.24), male (*n* = 48): 0.68 (0.30)			
Hulleberg 2008,[258] Norway	13-year follow-up: 0.75 (0.24) 60–69 years: 0.79 (0.12) 70–79 years: 81 (0.20) ≥ 80 years: 0.64 (0.24)	EQ-5D Mean follow-up 13 years (range 12–15 years)	Patient reported	*n* = 89 OA: 83%

Total hip replacement revision

Study and country	Utility data	Definition	Source	Patient characteristics [THR (*n*) and OA]
Briggs 2004,[38] Dawson 2001,[295] UK	Pre revision THR: 0.32 (95% CI 0.29 to 0.36, range –0.43 to 1.00) Post revision THR, 12 months: 0.62 (95% CI 0.59 to 0.65, range –0.59 to 1.00) 12 months post revision with 0 previous revisions (*n* = 137): 0.64 (95% CI 0.61 to 0.67, range –0.59 to 1.00); one previous revision (*n* = 18): 0.58 (95% CI 0.50 to 0.65, range –0.24 to 1.00); two or more previous revisions (*n* = 17): 0.51 (95% CI 0.40 to 0.63, range –0.24 to 1.00)	EQ-5D	Patient reported	*n* = 609 OA: 78.7%
Bozic 2011,[284] USA	Successful THR (*n* = 86): 0.96 (0.09) Failed primary THR (*n* = 30): 0.59 (0.34) Successful revision THR (*n* = 21): 0.84 (0.28) Failed revision (*n* = 27): 0.57 (0.36)	EQ-5D No time post THR reported	Patient reported	*n* = 231 OA patients

Appendix 15 Key studies reporting utilities for resurfacing arthroplasty

Study and country	Utility data	Definition	Source	Comments
Edlin 2012,[40] Costa 2012,[130] UK	Baseline: 0.308 (SD 0.338); 3 months: 0.722 (SD 0.229); 6 months: 0.796 (SD 0.244); 12 months: 0.795 (SD 0.216)	EQ-5D	Patient reported	Cormet™ metal-on-metal resurfacing (Corin Group, Cirencester, UK)
Baker 2011,[252] Pollard 2006,[251] UK	9-year follow-up: 0.84 (–0.18 to 1.00); 5-year follow-up: 0.90 (0.08 to 1.00)	EQ-5D Mean follow-up 61 months (range 52 –71 months)	Patient reported	NA
Bozic 2010,[253] USA	Before RS: 0.5 (SD 0.10); post RS: 0.92 (SD 0.04); post RS conversion to THR: 0.92 (SD 0.04); post major revision: 0.84 (SD 0.04); post minor revision: 0.88 (SD 0.04); post second major revision: 0.76 (SD 0.04); post second minor revision: 0.8 (SD 0.04)	Not clear what tool was used	Published resources	No distinction made between revision following THR and revision following RS

NA, not applicable.

Appendix 16 Excluded studies for registry searches

Final list of excluded studies for the registry searches undertaken on 3 April 2013

Reference	Reason
Aulakh T, Rao C, Kuiper J, Richardson J. Hip resurfacing and osteonecrosis: results from an independent hip resurfacing register. *Arch Orthop Trauma Surg* 2010;**130**:841–5	< 1000 patients
Berger A, Bozic K, Brett S, Edelsberg J, Sadosky A, Oster G. Patterns of pharmacotherapy and health care utilization and costs prior to total hip or total knee replacement in patients with osteoarthritis. *Arthritis Rheum* 2011;**62**:2268–75	Not a formal registry – large US health insurance database, claims database includes 16,527 patients
Bozic K, Lau E, Ong K, Vail T, Rubash H, Berry D. Comparative effectiveness of metal-on-metal and metal-on-polyethylene bearings in Medicare total hip arthroplasty patients. *J Arthroplasty* 2012;**27**(8 Suppl.):37–40	Not a formal registry – large US health insurance database, claims database includes 148,827 patients
Bryan R, Nair P, Taylor M. Influence of femur size and morphology on load transfer in the resurfaced femoral head: a large scale, multi-subject finite element study. *J Biomech* 2012;**45**:1952–58	Computerised modelling
Bucci J, Oglesbyb R, Agodoac L Abbotta K. Hospitalizations for total hip arthroplasty after renal transplantation in the United States. *Am J Transplant* 2002;**2**:999–1004	Renal transplant study
Changulani M, Peel T, Field R. The relationship between obesity and the age at which hip and knee replacement is undertaken. *J Bone Joint Surg* 2008;**90-B**:360–3	Obesity focus
Chen A, Gupte C, Akhtar K, Smith P, Cobb J. The global economic cost of osteoarthritis: how the UK compares. *Arthritis* 2012;**2012**:698709	Cost focus
Culliford D, Judge A, Maskell J, Arden N. Estimating prosthesis survival following total hip replacement in the United Kingdom. *Osteoarthritis Cartilage* 2010;**20**(Suppl. 1):S180–1	GP database study on lifetime risk of OA
Culliford D, Judge A, Maskell J, Arden N. Estimating prosthesis survival following total hip replacement in the United Kingdom. *Abstracts Osteoarthritis Cartilage* 2012;**20**:S180–1	GP database study – lifetime risk of OA
Della Valle C, Nunley R, Raterman S, Barrack R. Initial American experience with hip resurfacing following FDA approval. *Clin Orthop Relat Res* 2009;**468**:72–8	Not a registry study
Derbyshire B, Porter M. A study of the Elite Plus femoral component using radiostereometric analysis. *J Bone Joint Surg* 2007;**89-B**:730–5	Radio analysis study
de Steiger R, Hang J, Miller L, Graves S, Davidson D. Five-year results of the ASR XL Acetabular System and the ASR Hip Resurfacing System: an analysis from the Australian Orthopaedic Association National Joint Replacement Registry. *J Bone Joint Surg* 2011;**93**:2287–93	DePuy ASR study
Devane P, Wraighte P, Ong D, Horne G. Do joint registries report true rates of hip dislocation? *Clin Orthop Relat Res* 2012;**470**:3003–6	New Zealand registry study, < 1000 patients
Dunbar M. Antibiotic bone cements: their use in routine primary total joint arthroplasty is justified. *Orthopaedics* 2009;**32**:660–3	Bone cement study
Engesæter I, Lehmann T, Laborie L, Lie S, Rosendahl K, Engesæter L. Total hip replacement in young adults with hip dysplasia: age at diagnosis, previous treatment, quality of life, and validation of diagnoses reported to the Norwegian Arthroplasty Register between 1987 and 2007. *Acta Orthop* 2011;**82**:149–54	Hip dysplasia study

Reference	Reason
Englund A, Franklin M, Petersson J, Ingemar F. Differences in the observed rate of hip fracture in male and female patients diagnosed with osteoarthritis or ankylosing spondylitis compared with the expected based on the general population seeking health care. *Arthritis Rheum* 2010;**62**(Suppl. 10):963	Not a registry study
Flugsrud G, Nordsletten L, Espehaug B, Havelin L, Engeland A, Meyer H. The impact of body mass index on later total hip arthroplasty for primary osteoarthritis: a cohort study in 1.2 million persons. *Arthritis Rheum* 2006;**54**:802–807	BMI study
Froberg L, Christensen F, Pedersen N, Overgaard S. The need for total hip arthroplasty in Perthes' disease: a long-term study. *Clin Orthop Relat Res* 2011;**469**:1134–40	Perthes' disease study
Gandhi R, Razak F, Tso P, Davey J, Mahomed N. Greater perceived helplessness in osteoarthritis predicts outcome of joint replacement. *J Rheumatol* 2009;**36**:1507–11	Psychological study
Gillam M, Ryan P, Salter A, Graves S. Multi-state models and arthroplasty histories after unilateral total hip arthroplasties: introducing the summary notation for arthroplasty histories. *Acta Orthop* 2012;**83**:220–6	Study on revision rates
Gorenoi V, Schönermark M, Hagen A. Prevention of infection after knee arthroplasty. *GMS Health Technol Assess* 2010;**6**:1–12	Infection prevention study
Gottliebsen M, Rahbek O, Ottosen P, Soballe K, Stilling M. Superior 11 year survival but higher polyethylene wear of hydroxyapatite-coated Mallory-head cups. *Hip Int* 2012;**22**:35–40	< 1000 patients
Hartmann A, Lutzner J, Kirschner S, Witzleb W, Gunther K. Do survival rate and serum ion concentrations 10 years after metal-on-metal hip resurfacing provide evidence for continued use? *Clin Orthop Relat Res* 2012;**470**:3118–26	< 1000 patients
Hirvonen J, Blom M, Tuominen U, Seitsalo S, Lehto M, Paavolainen P, *et al.* Health-related quality of life in patients waiting for major joint replacement. A comparison between patients and population controls. *Health Qual Life Outcomes* 2006;**4**(3)	Prospective study
Jarvholm B, Lundstro R, Malchau H, Rehn B, Vingard E. Osteoarthritis in the hip and whole-body vibration in heavy vehicles. *Int Arch Occup Environ Health* 2004;**77**:424–6	Vibration study
Jarvholm B, Lewold S, Malchau H, Vingard E. Age, bodyweight, smoking habits and the risk of severe osteoarthritis in the hip and knee in men. *Eur J Epidemiol* 2005;**20**:537–42	Smoking study
Jarvholm B, From C, Lewold S, Malchau H, Vingard E. Incidence of surgically treated osteoarthritis in the hip and knee in male construction workers. *Occup Environ Med* 2008;**65**:275–8	Profession study
Kadam U, Holmberg A, Blagojevic M, Nilsson P, Åkesson K. Risk factors for cardiovascular disease and future osteoarthritis-related arthroplasty: a population-based cohort study in men and women from Malmö, Sweden. *Scand J Rheumatol* 2011;**40**:478–85	OA linked to cardiovascular disease
Katz J, Wright E, Wright J, Corbett K, Malchau H, Baron J, *et al.* Choice of hospital for revision total hip replacement. *J Bone Joint Surg* 2010;**92**:2829–34	Revisions on Medicare
Khan A, McLoughlin E, Giannakas K, Hutchinson C, Andrew J. Hip osteoarthritis: where is the pain? *Ann R Coll Surg Engl* 2004;**86**:119–21	Pain study
Kiefer H, Othman A. Ultrasound vs pointer palpation based method in THA navigation: a comparative study. *Orthopedics* 2007;**30**:S152–6	Ultrasound study
Kim H, Koh S, Lee B, Kim I, Seo Y, Song Y, *et al.* Low rate of total hip replacement as reflected by a low prevalence of hip osteoarthritis in South Korea. *Osteoarthritis Cartilage* 2008;**16**:1572–5	The Republic of Korea, < 1000 patients (580 patients)
Knight D, Alves C, Wedge J. Femoral varus derotation osteotomy for the treatment of habitual subluxation and dislocation of the pediatric hip in trisomy 21: a 10-year experience. *J Pediatr Orthop* 2011;**31**:638–43	Study of children
Lai Y, Wei H, Cheng C. Incidence of hip replacement among national health insurance enrollees in Taiwan. *J Orthop Surg Res* 2008;**3**:42	Setting up a registry in Taiwan (Province of China)

Reference	Reason
Leonardsson O, Kärrholm J, Åkesson K, Garellick G, Rogmark C. Higher risk of reoperation for bipolar and uncemented hemiarthroplasty. *Acta Orthop* 2012;**83**:459–66	Hemiarthroplasty after hip fracture study
Lubega N, Mkandawire N, Sibande G, Norrish A, Harrison W. Joint replacement in Malawi establishment of a national joint registry. *J Bone Joint Surg* 2009;**91-B**:341–3	Malawi, < 1000 patients (73 patients)
Lutro O, Langvatn H, Schrama J, Hallan G, Dale H, Espehaug B, *et al.* Resistance of staphylococci isolated from infected hip arthroplasties in Norway. 20th European Congress of Clinical Microbiology and Infectious Diseases, Vienna, Austria, 1–13 April 2010. Abstract no. O484	Abstract comment
Makela K, Häkkinen U, Peltola M, Linna M, Kröger H, Remes V. The effect of hospital volume on length of stay, re-admissions, and complications of total hip arthroplasty. A population-based register analysis of 72 hospitals and 30,266 replacements. *Acta Orthop* 2010;**82**:20–6	Cost of readmission
Medical Advisory Secretariat. Metal-on-metal total hip resurfacing arthroplasty: an evidence-based analysis. *Ontario Health Technol Assess Ser* 2006;**6**(4)	Not a registry study
Meenagh G, McGibbon D, Nixon J, Wright G, Doherty M, Hughes A. Lack of support for the presence of an osteoarthritis susceptibility locus on chromosome 6p. *Arthritis Rheum* 2005;**52**:2040–3	Genetics study
Minns Lowes C, Barker K, Dewey M, Sackley C. Effectiveness of physiotherapy exercise following hip arthroplasty for osteoarthritis: a systematic review of clinical trials. *BMC Musculoskelet Disord* 2009;**10**:98	Physiotherapy study
Mundy G, Harper E. Primary hip replacement in young osteoarthritic patients – current practices in one UK region. *Hip Int* 2005;**15**:159–65	< 1000 patients (911 patients)
Munger P, Röder C, Ackermann-Liebrich U, Busato A. Patient-related risk factors leading to aseptic stem loosening in total hip arthroplasty. A case-control study of 5,035 patients. *Acta Orthop* 2006;**77**:567–74	< 1000 patients
Nikiphorou E, Dixey J, Williams P, Morris S, Young A. Osteoporotic fracture in rheumatoid arthritis (RA): a study of incidence, predictive factors and economic burden from a 25 year RA inception cohort. *Osteoporos Int* 2012;**23**(Suppl. 2):S159	Fracture
Pina M. Epidemiology of total hip and knee replacement and revision surgeries. *Osteoporos Int* 2010;**21**(Suppl. 1):P628	Abstract of epidemiology of OA
Pivec R, Johnson A, Mears S, Mont M. Hip arthroplasty. *Lancet* 2012;**380**:1768–77	Contains background information only
Plate J, Seyler T, Stroh D, Issa K, Akbar M, Mont M. Risk of dislocation using large- vs. small-diameter femoral heads in total hip arthroplasty. *BMC Res Notes* 2012;**5**:553	< 1000 patients (48 patients)

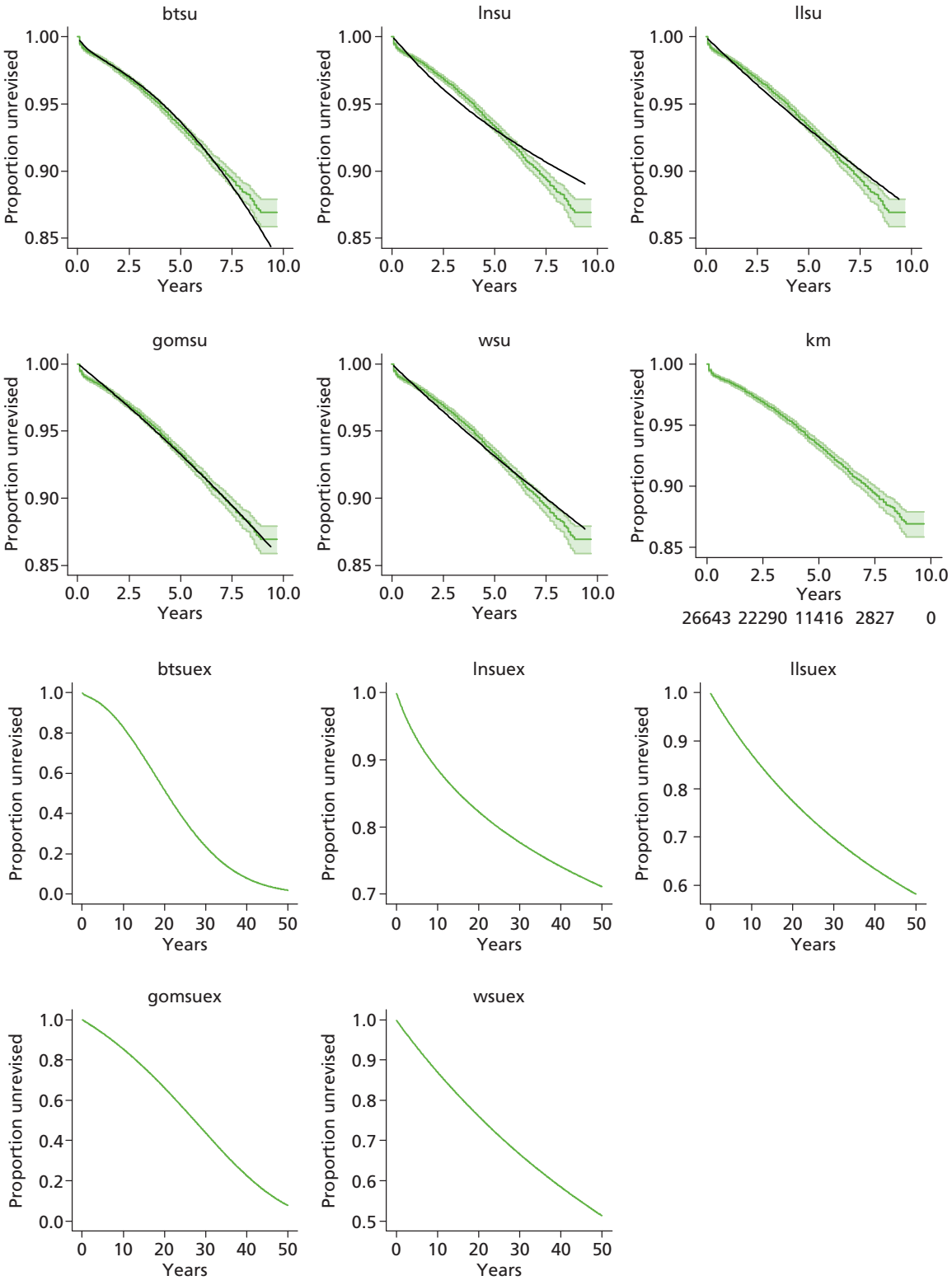
Appendix 17 Catalogue of modelled parametric fits to observed time to revision

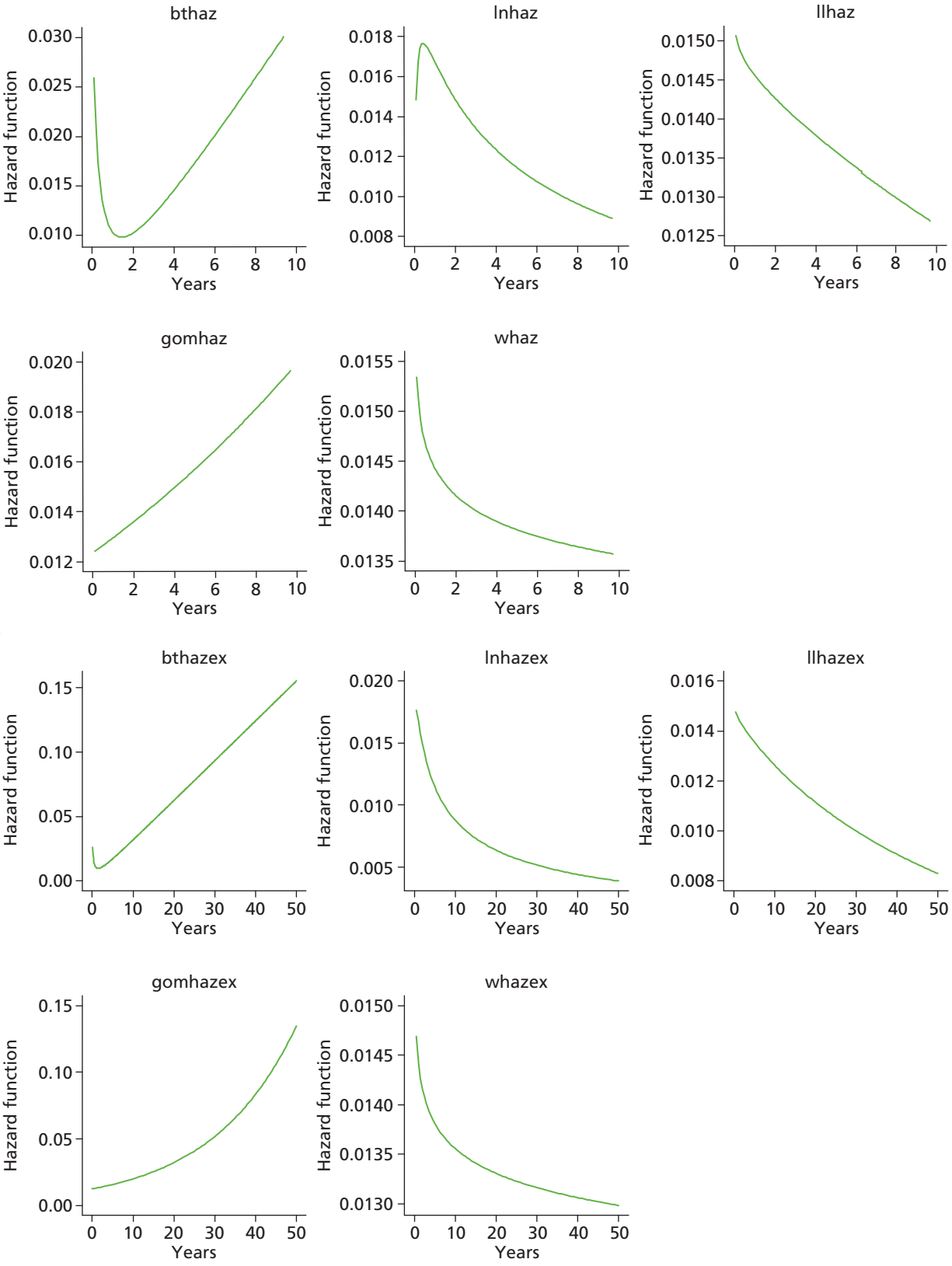
Key to the abbreviations used in the following figures of modelled parametric fits to observed time to revision of hip replacements.

bthaz	Bathtub modelled hazard
bthazex	Bathtub modelled hazard extrapolated
btsu	Bathtub fit
btsuex	Bathtub fit extrapolated
gomhaz	Gompertz modelled hazard
gomhazex	Gompertz modelled hazard extrapolated
gomsu	Gompertz fit
gomsuex	Gompertz fit extrapolated
km	Kaplan–Meier estimate
llhaz	Log-logistic modelled hazard
llhazex	Log-logistic modelled hazard extrapolated
llsu	Log-logistic fit
llsuex	Log-logistic fit extrapolated
lnhaz	Log-normal modelled hazard
lnhazex	Log-normal modelled hazard extrapolated
lnsu	Log-normal fit
lnsuex	Log-normal fit extrapolated
whaz	Weibull modelled hazard
whazex	Weibull modelled hazard extrapolated
wsu	Weibull fit
wsuex	Weibull fit extrapolated

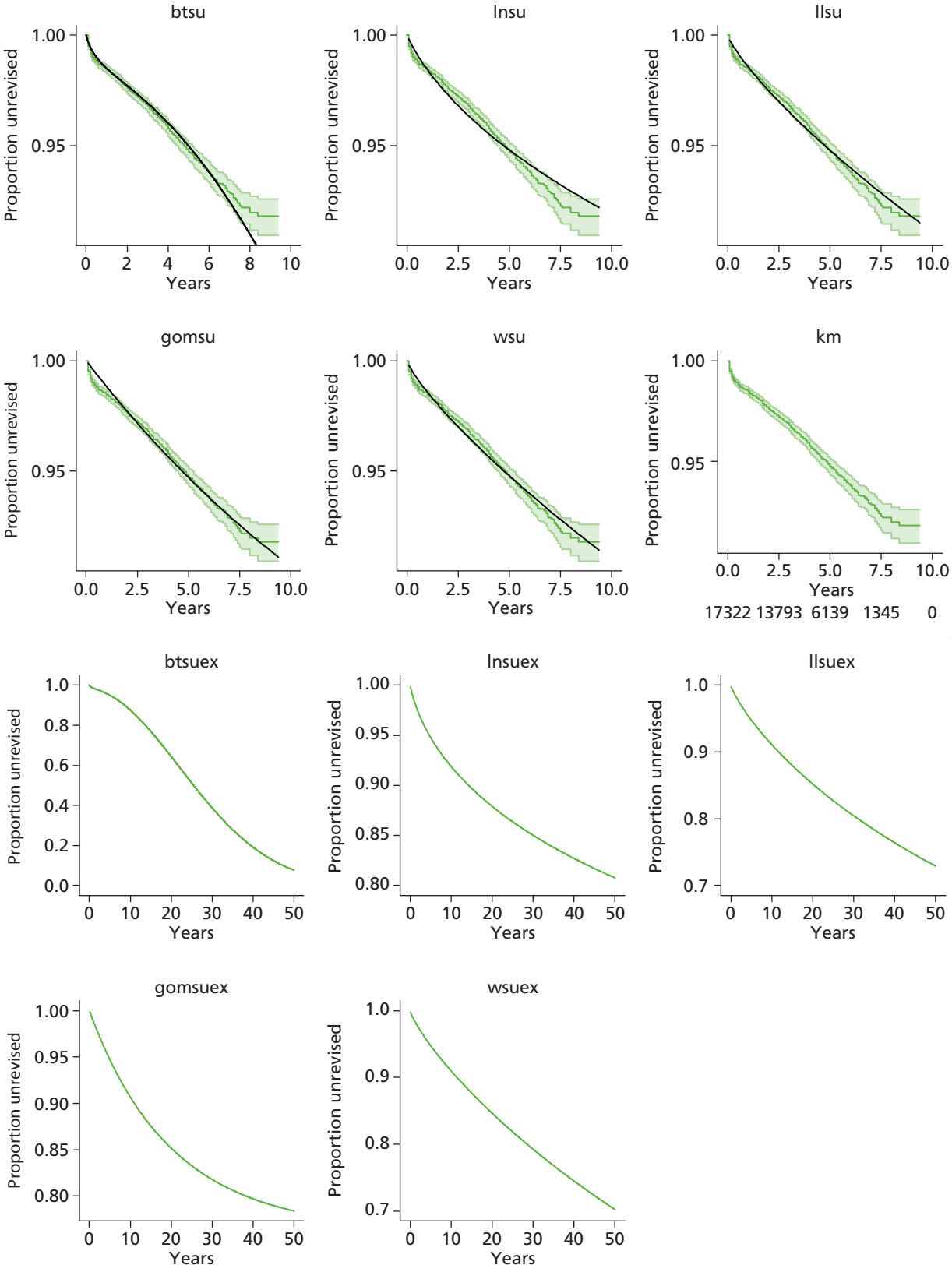
Results from Kaplan–Meier analyses of revision rates for patients receiving total hip replacement and resurfacing arthroplasty interventions

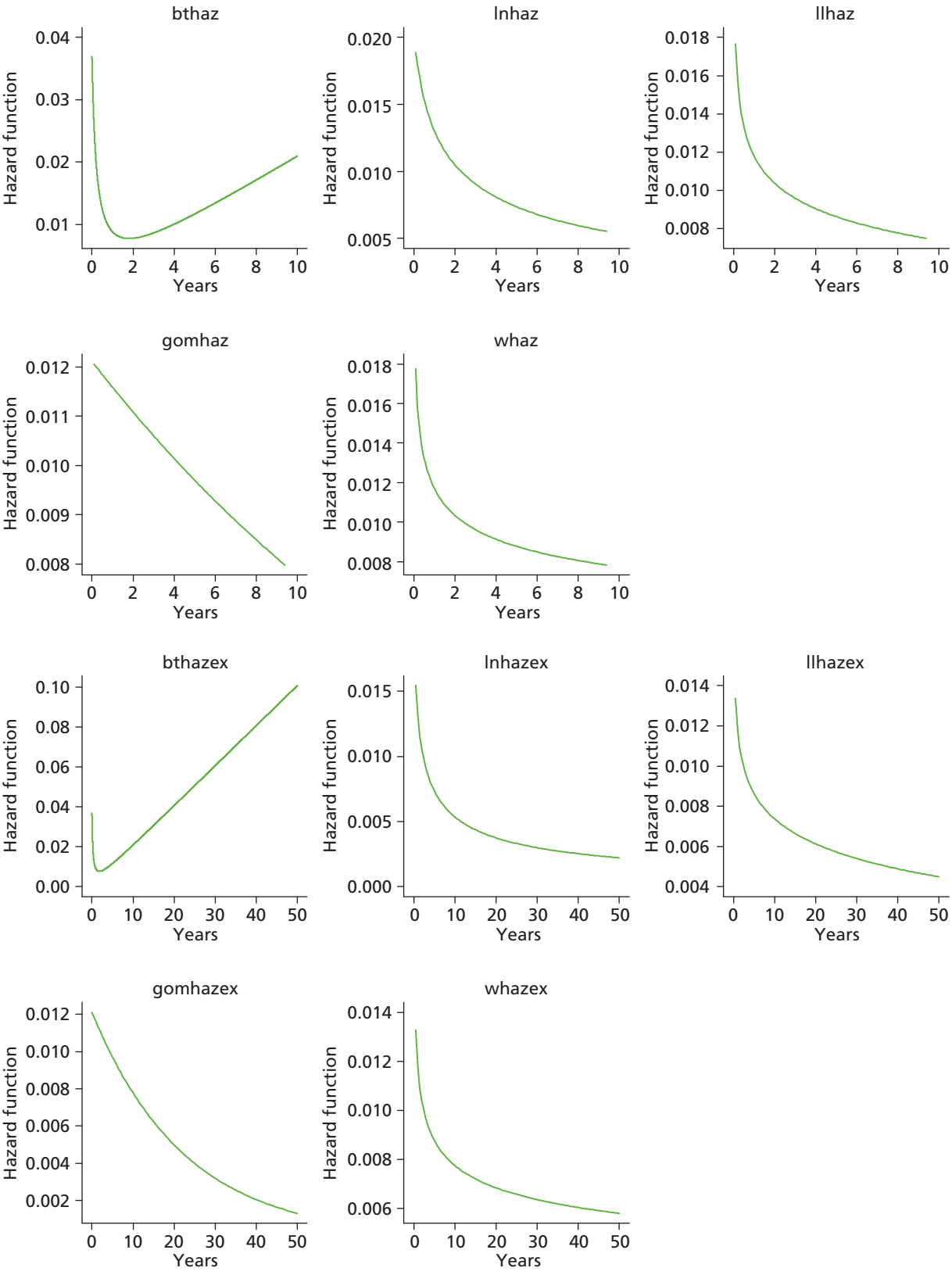
Resurfacing arthroplasty



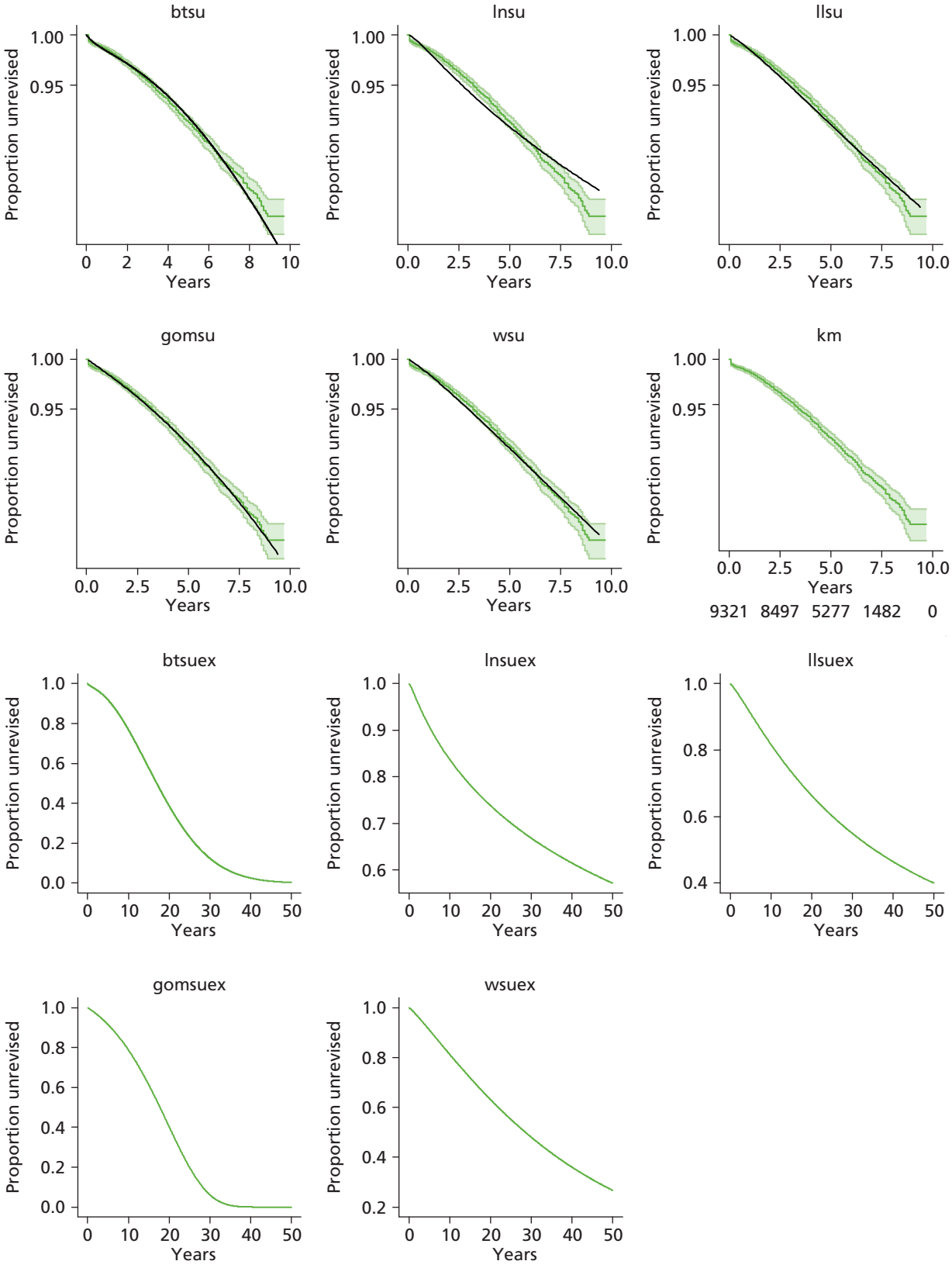


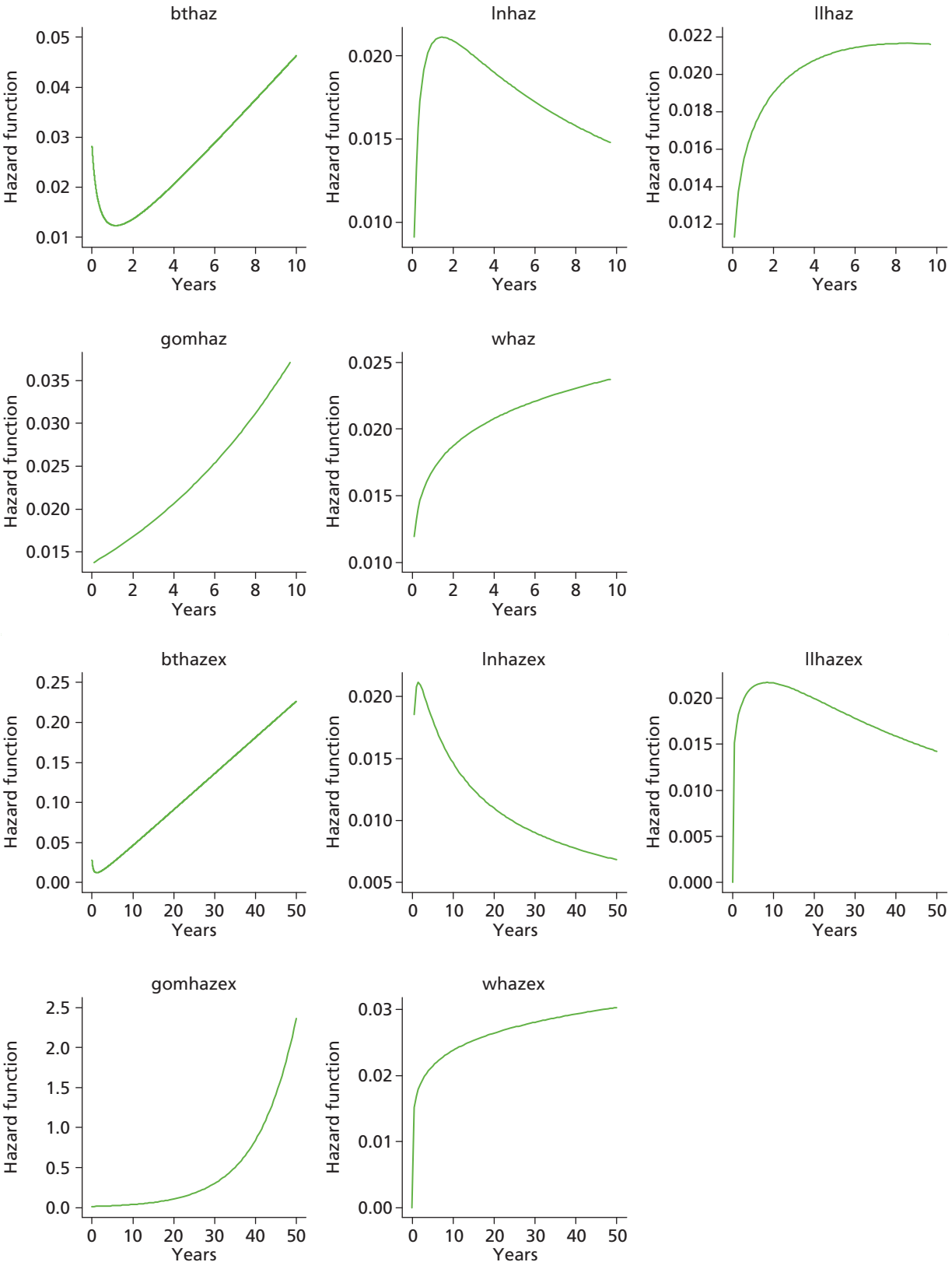
Resurfacing arthroplasty: male



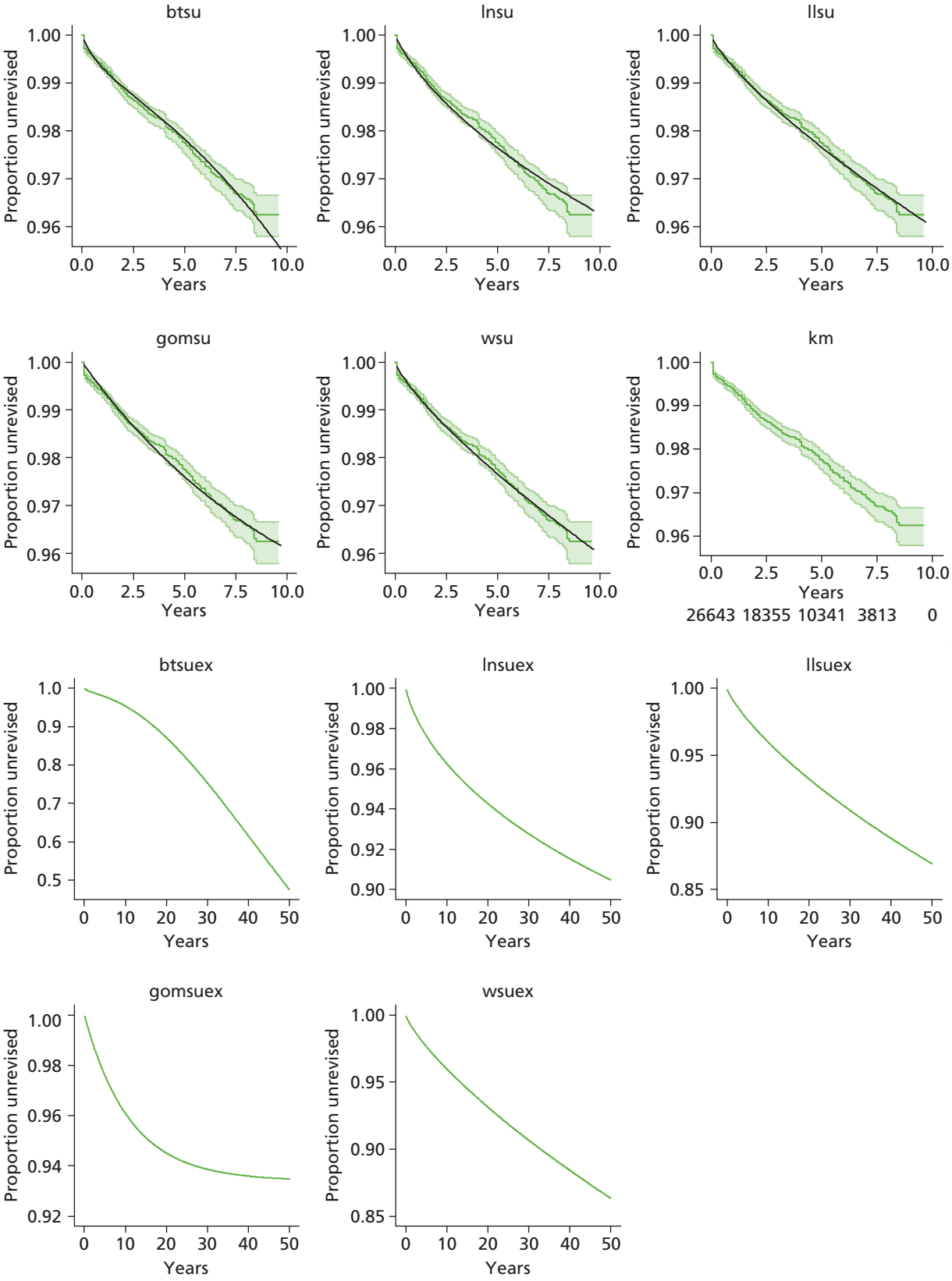


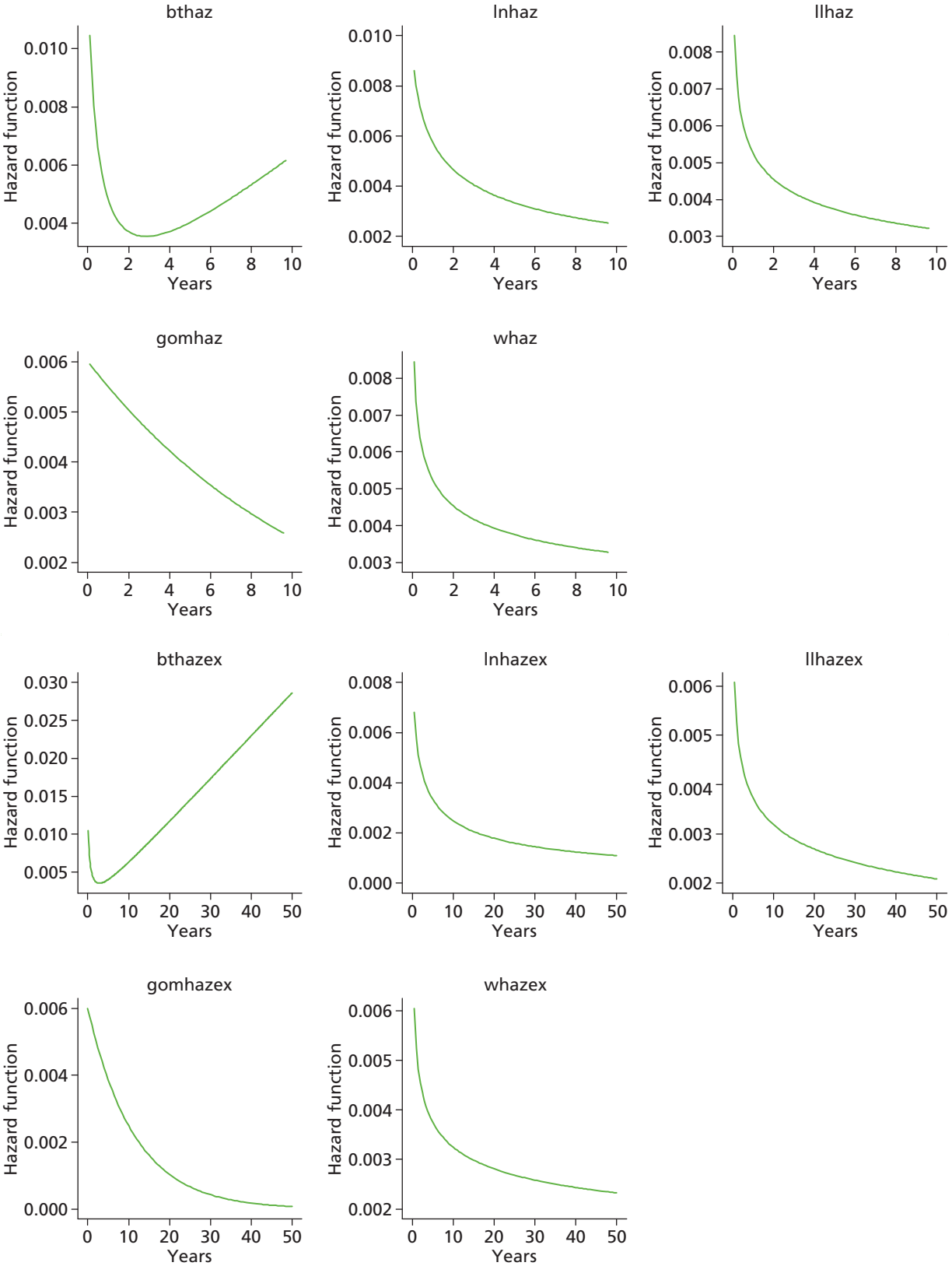
Resurfacing arthroplasty: female



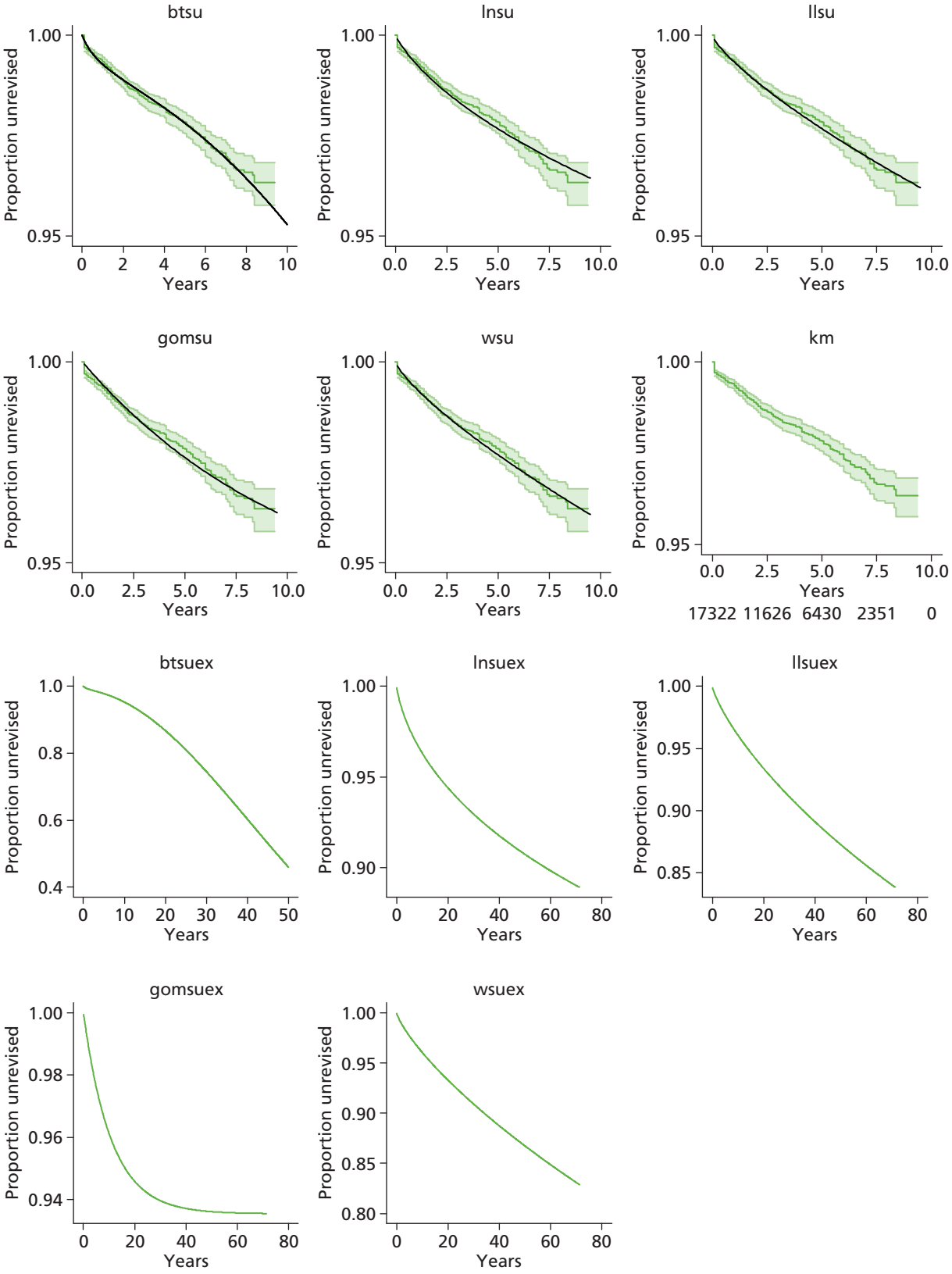


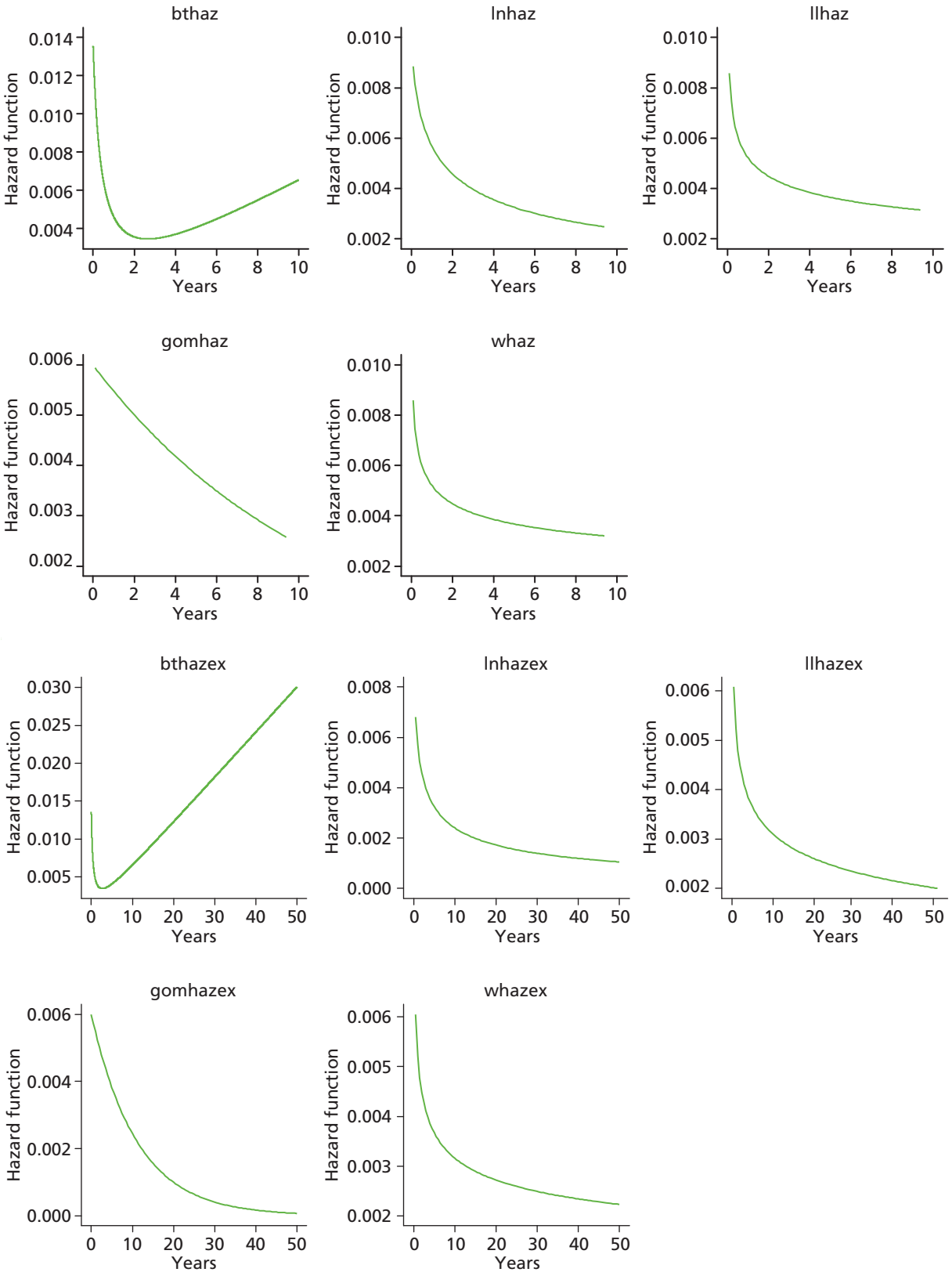
Total hip replacement (matched)



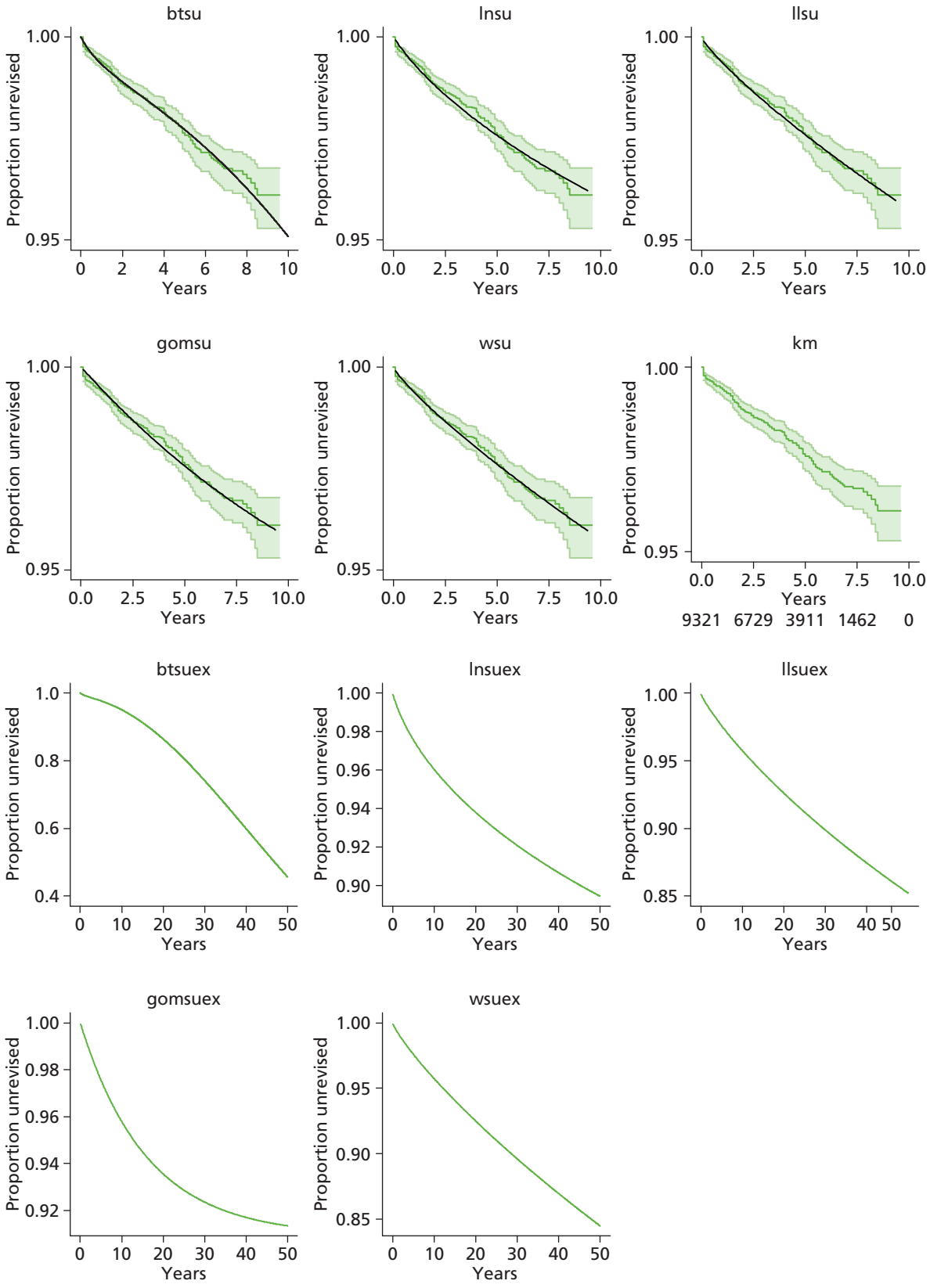


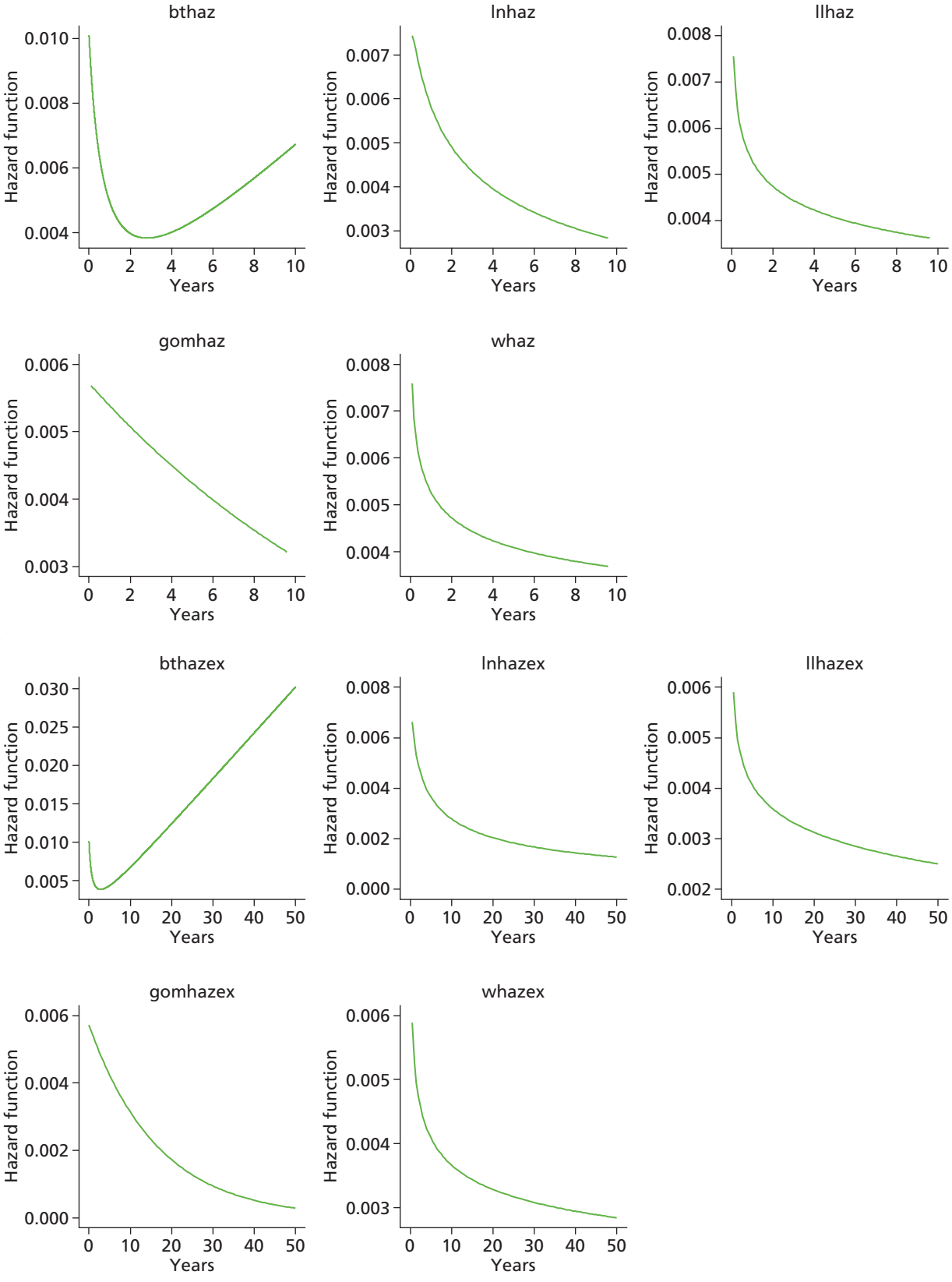
Total hip replacement (matched): male





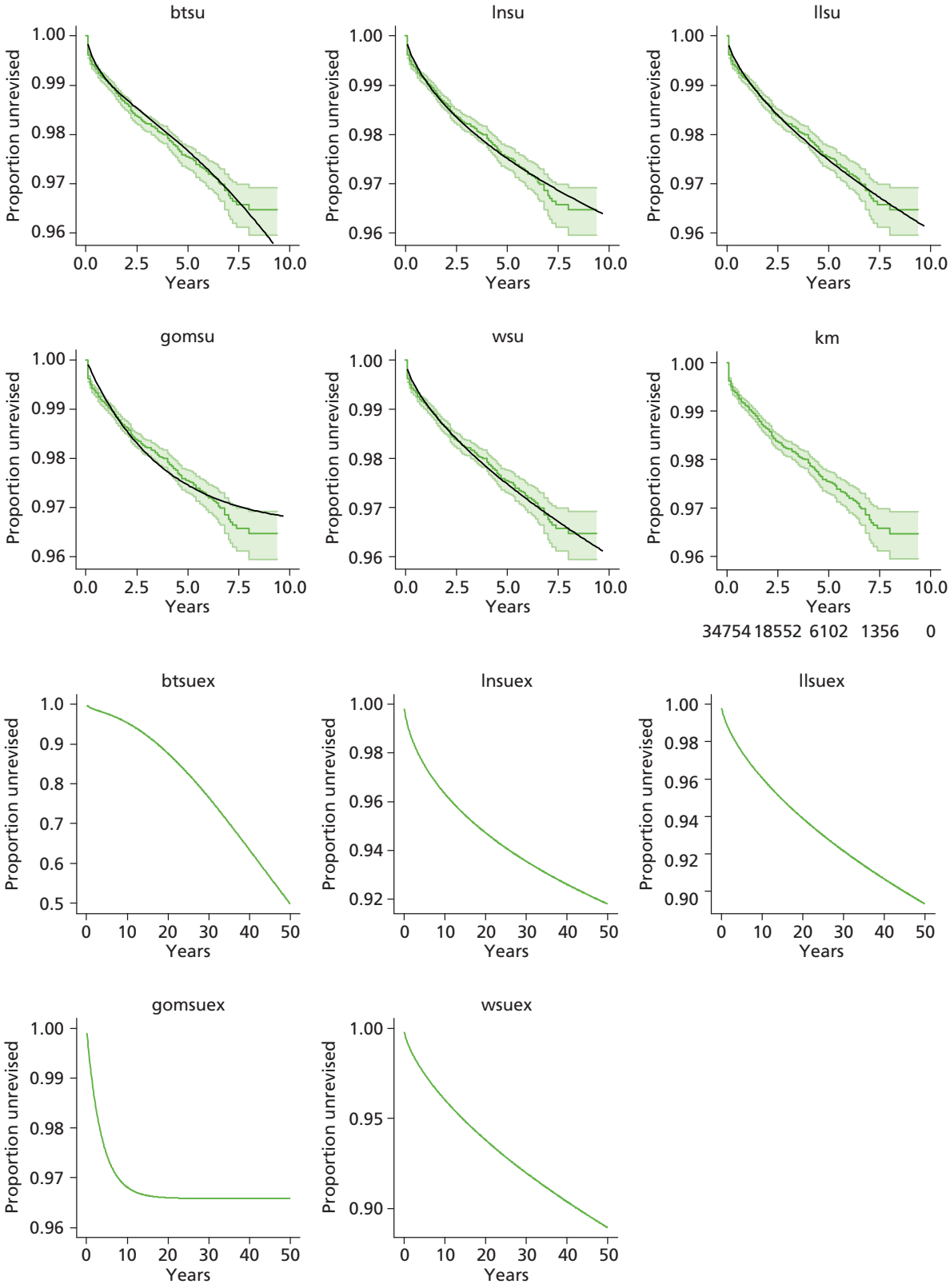
Total hip replacement (matched): female

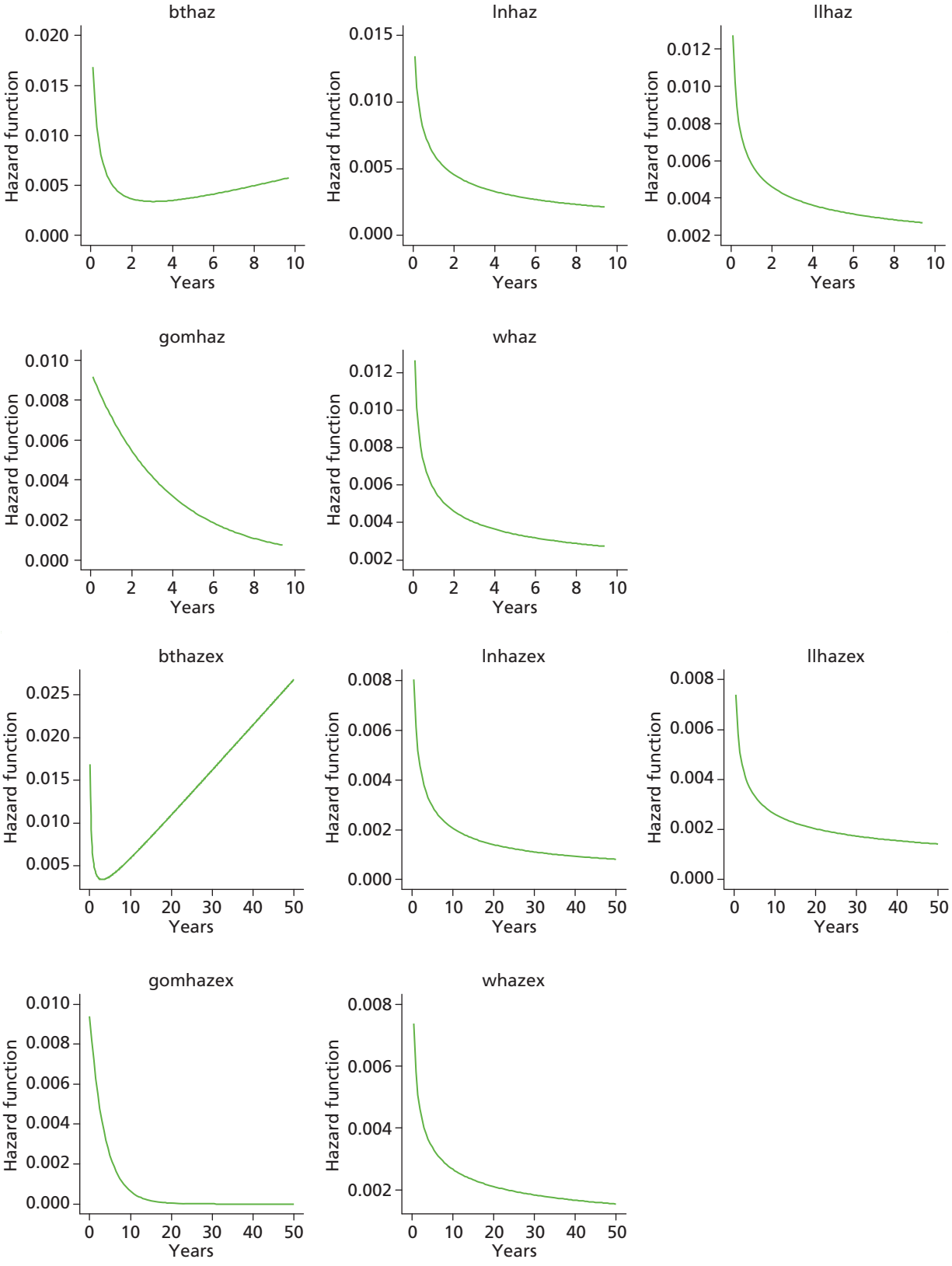




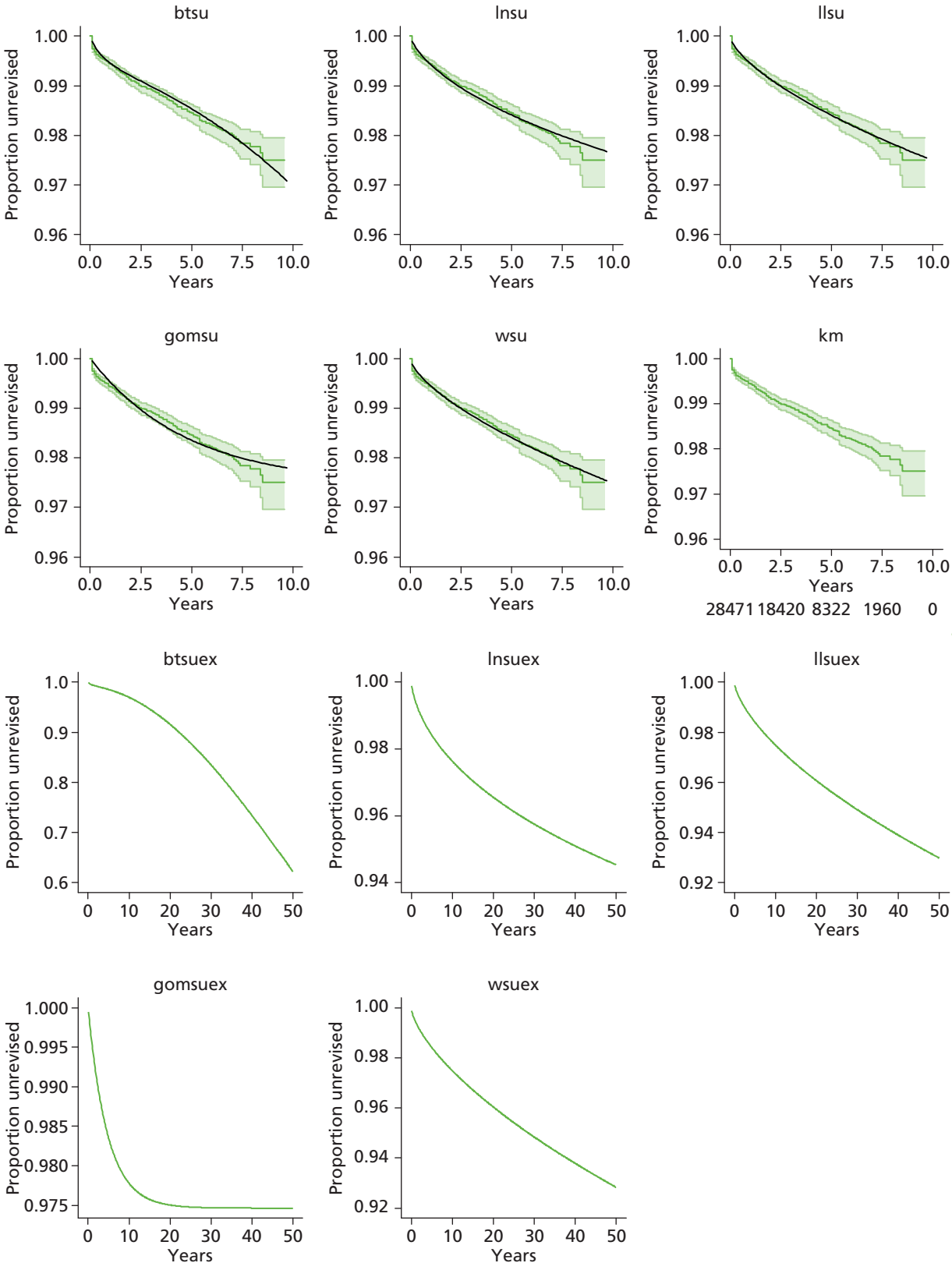
Total hip replacement categories

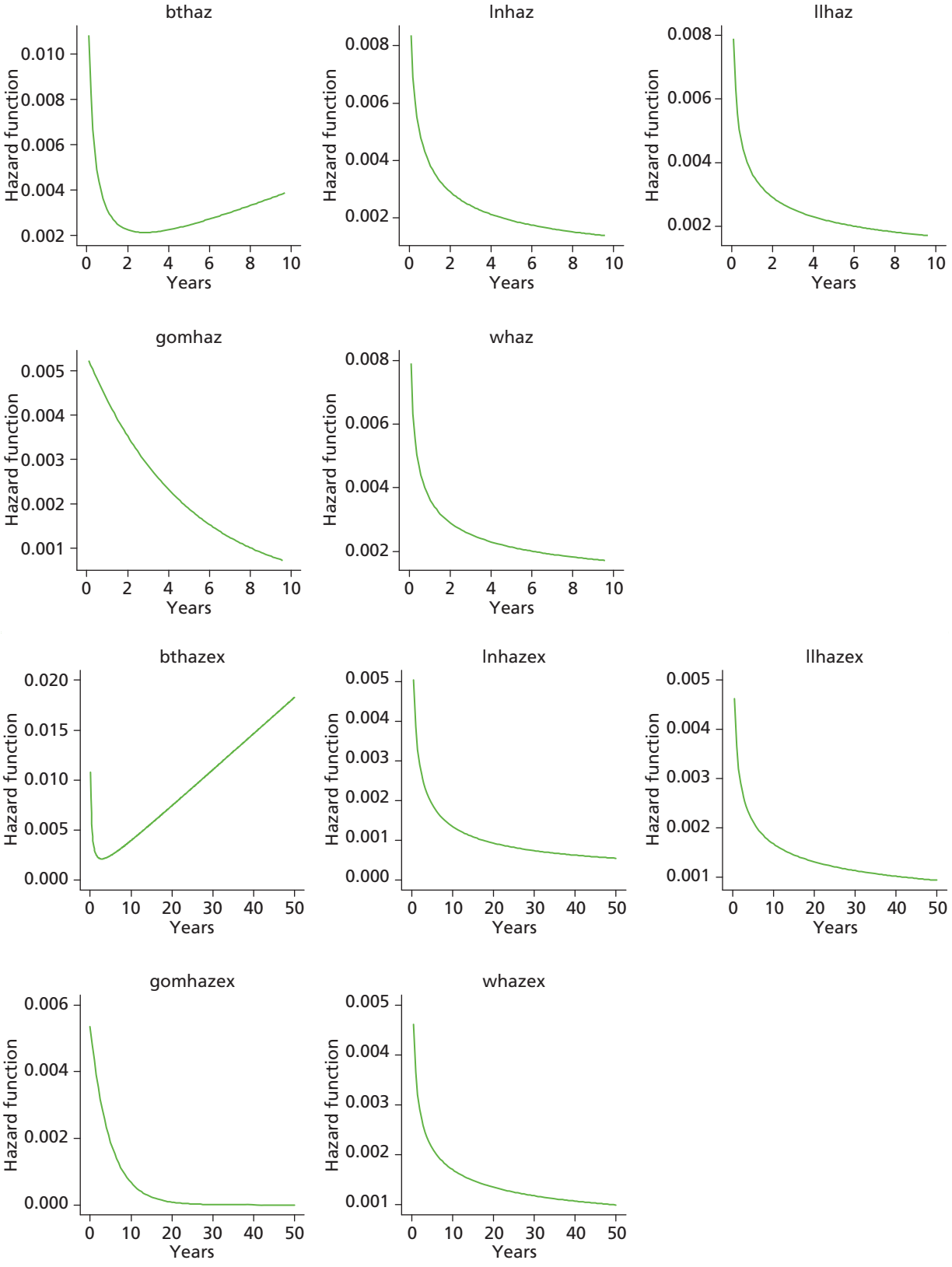
Category C [ceramic head (cementless stem) hydroxyapatite-coated metal cup (polyethylene liner)]



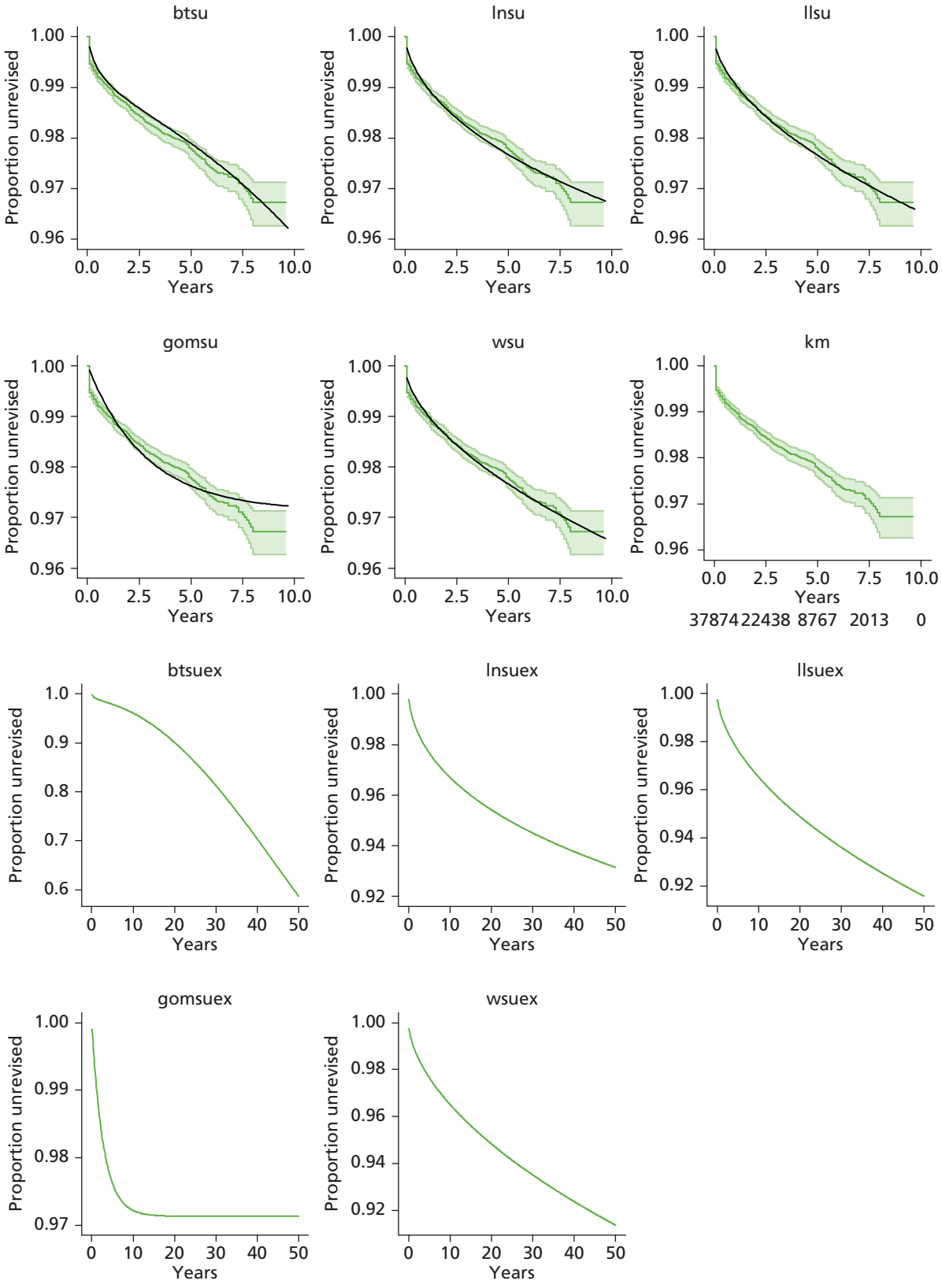


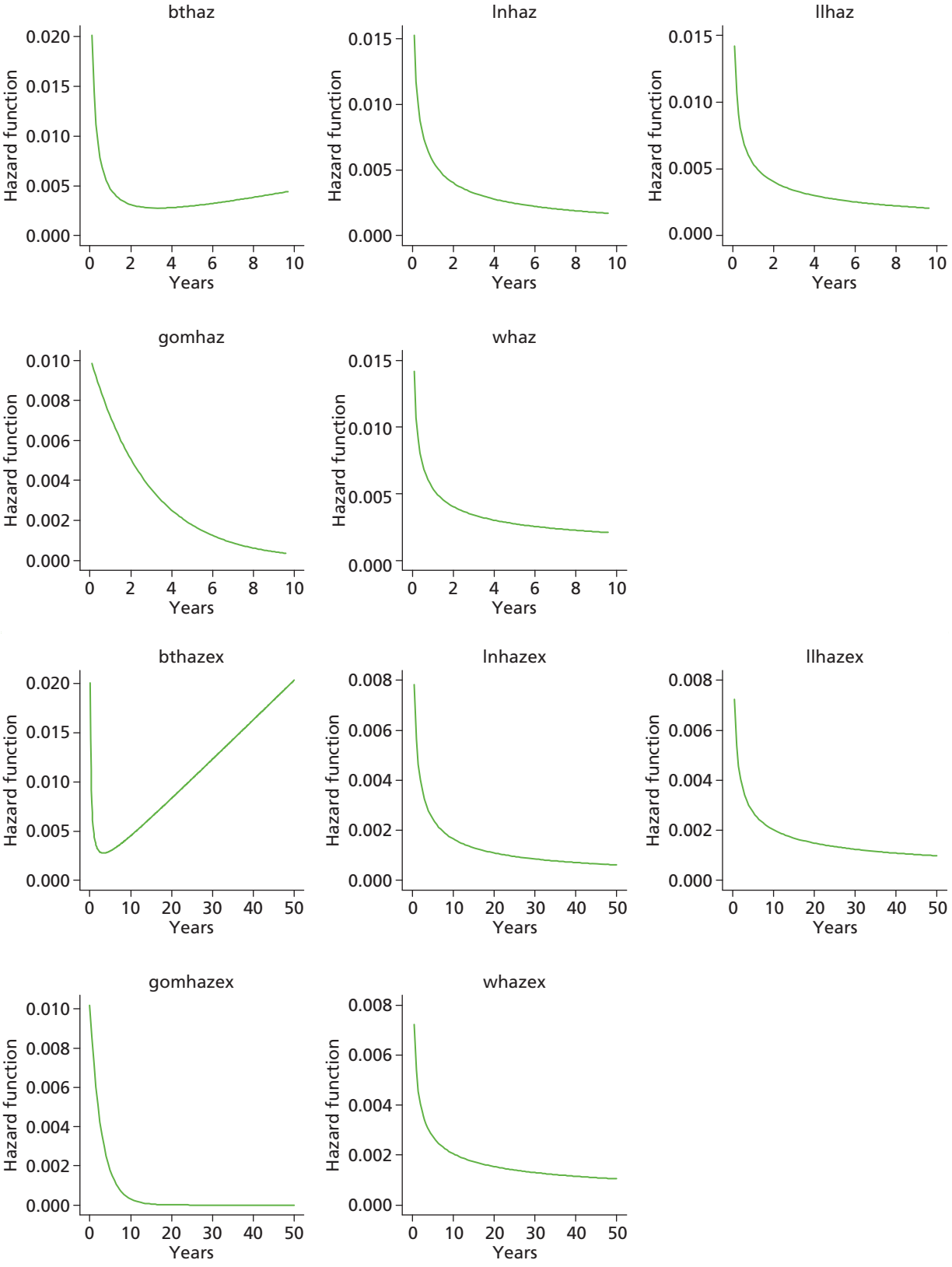
Category D [hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



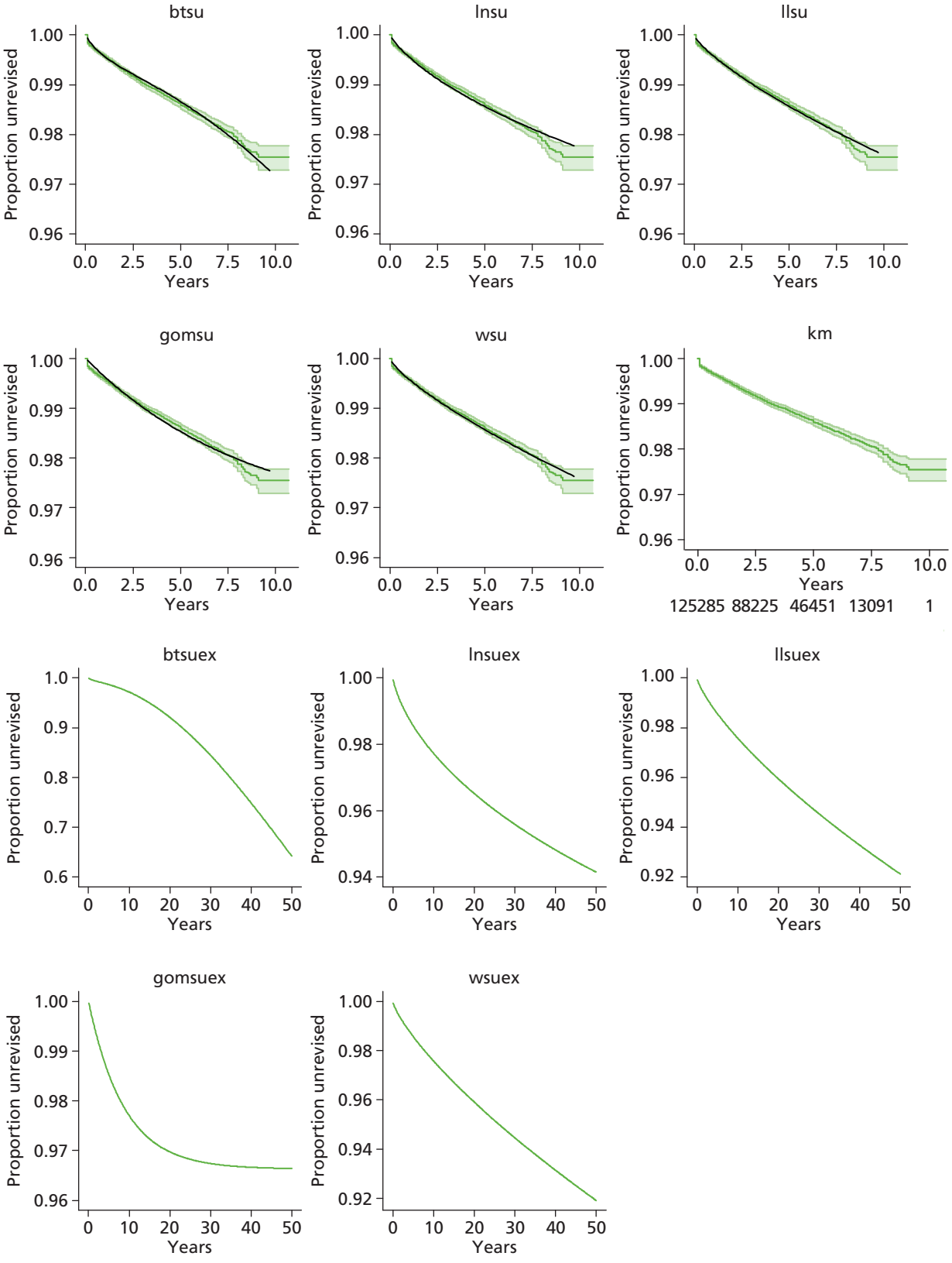


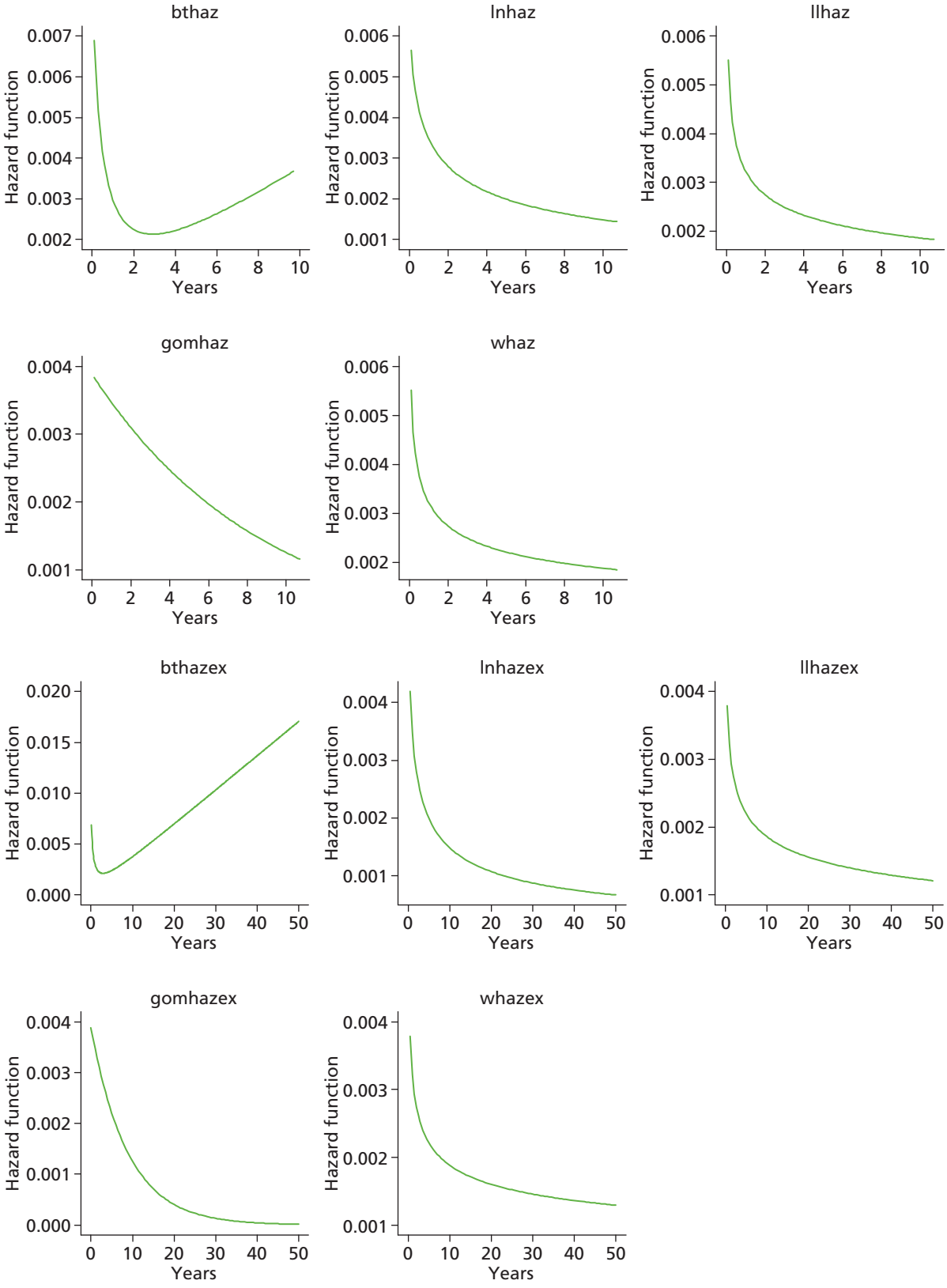
Category B [metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



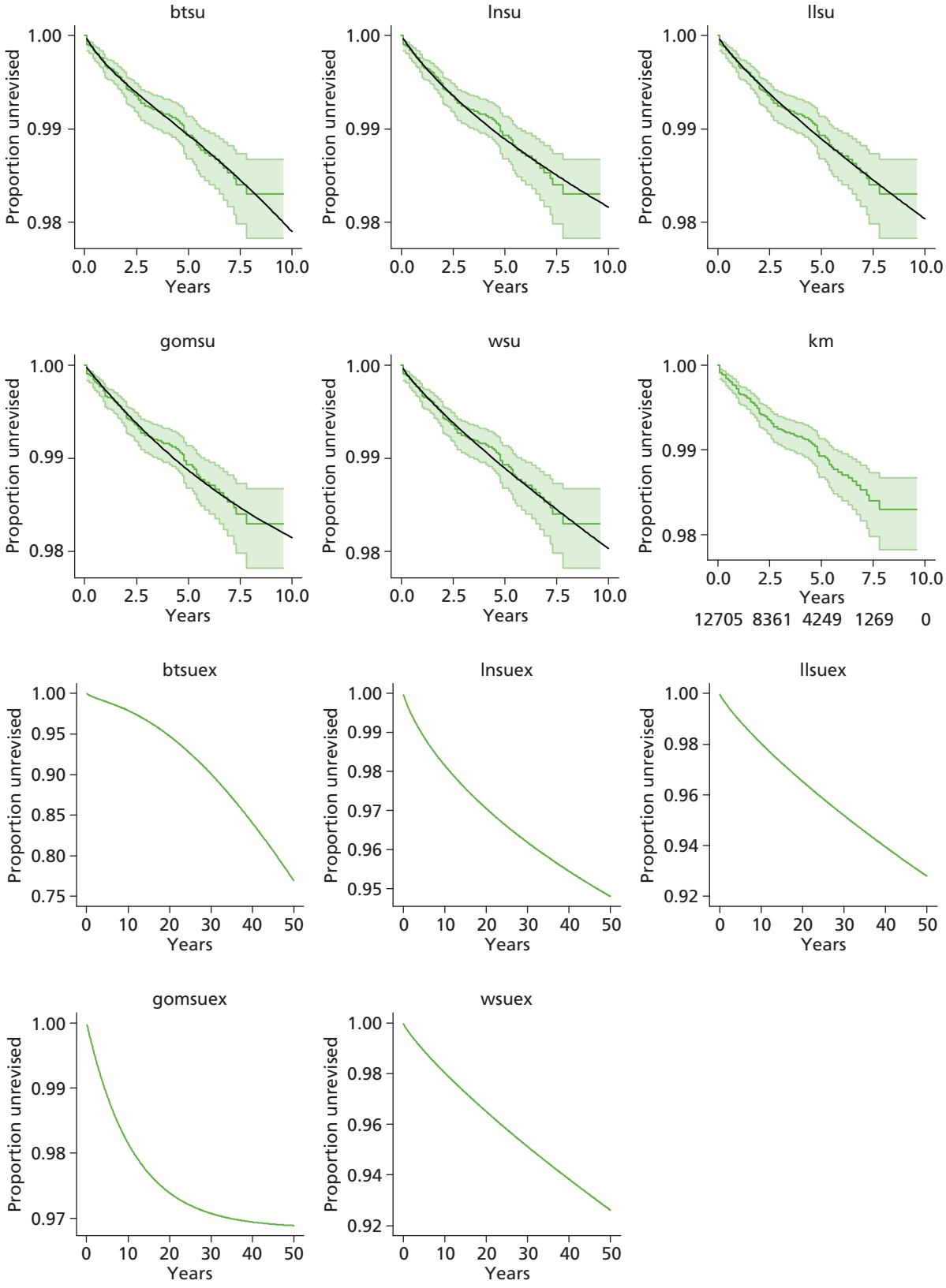


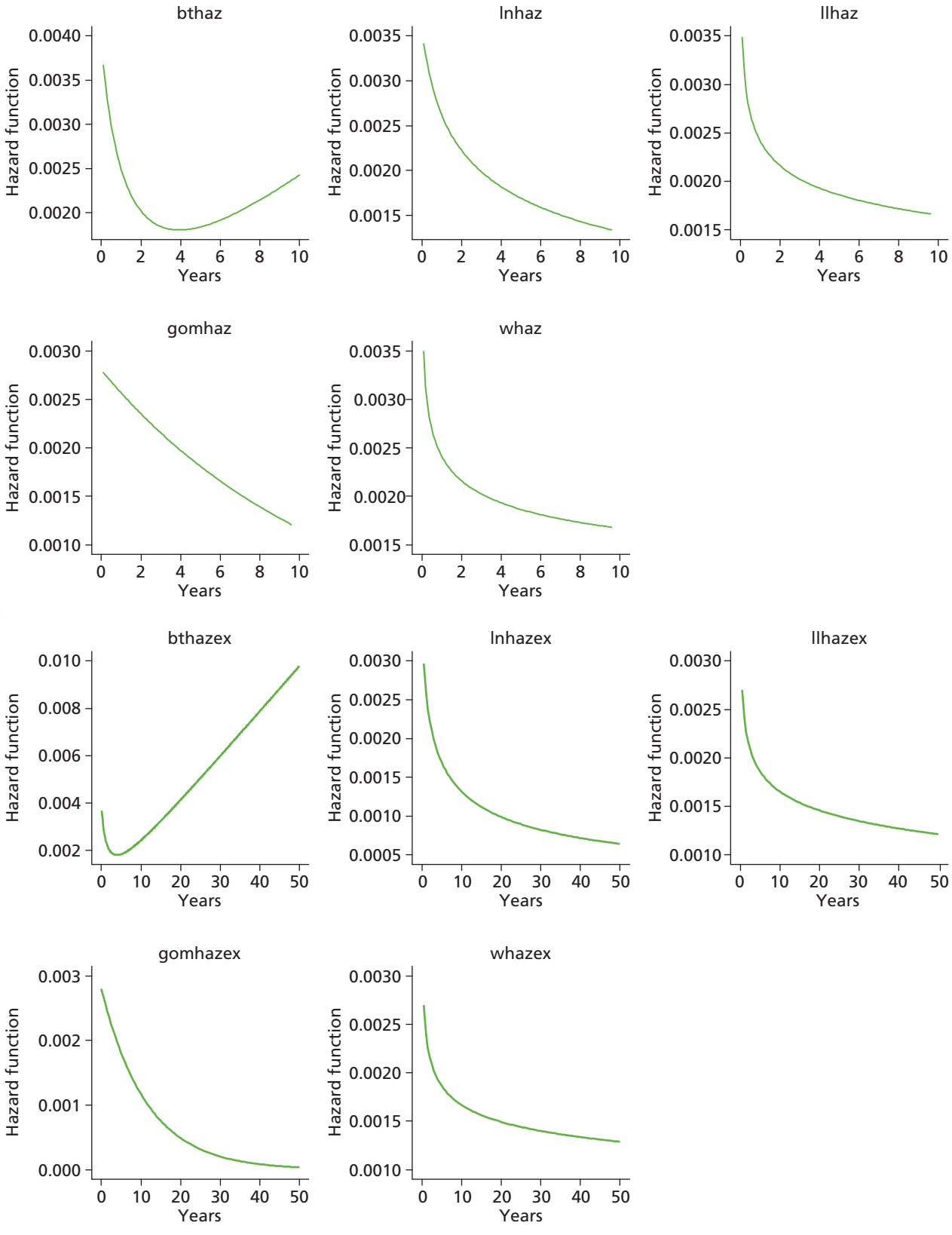
Category A [metal head (cemented stem) on cemented polyethylene cup]





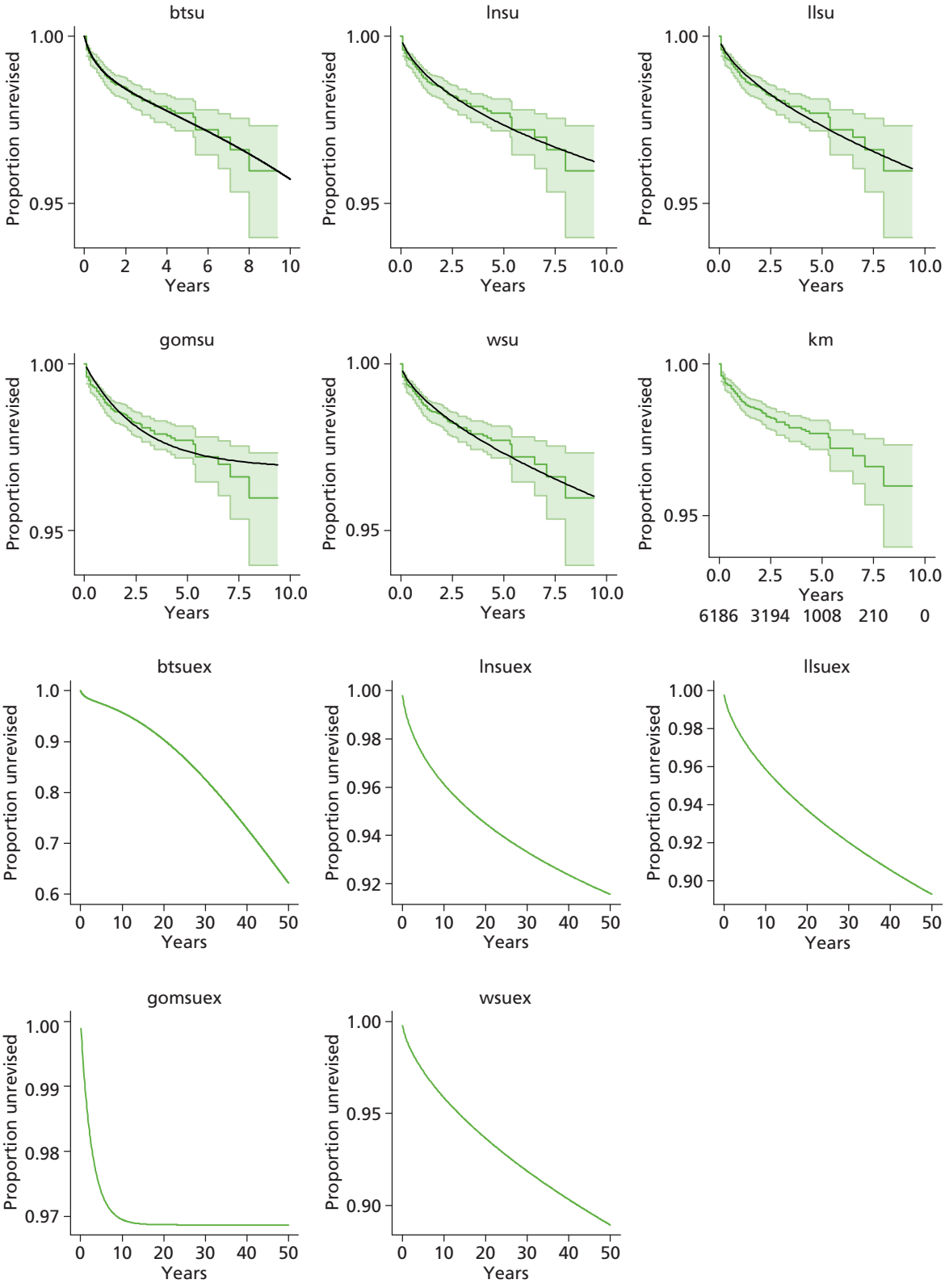
Category E [ceramic head (cemented stem) on cemented polyethylene cup]

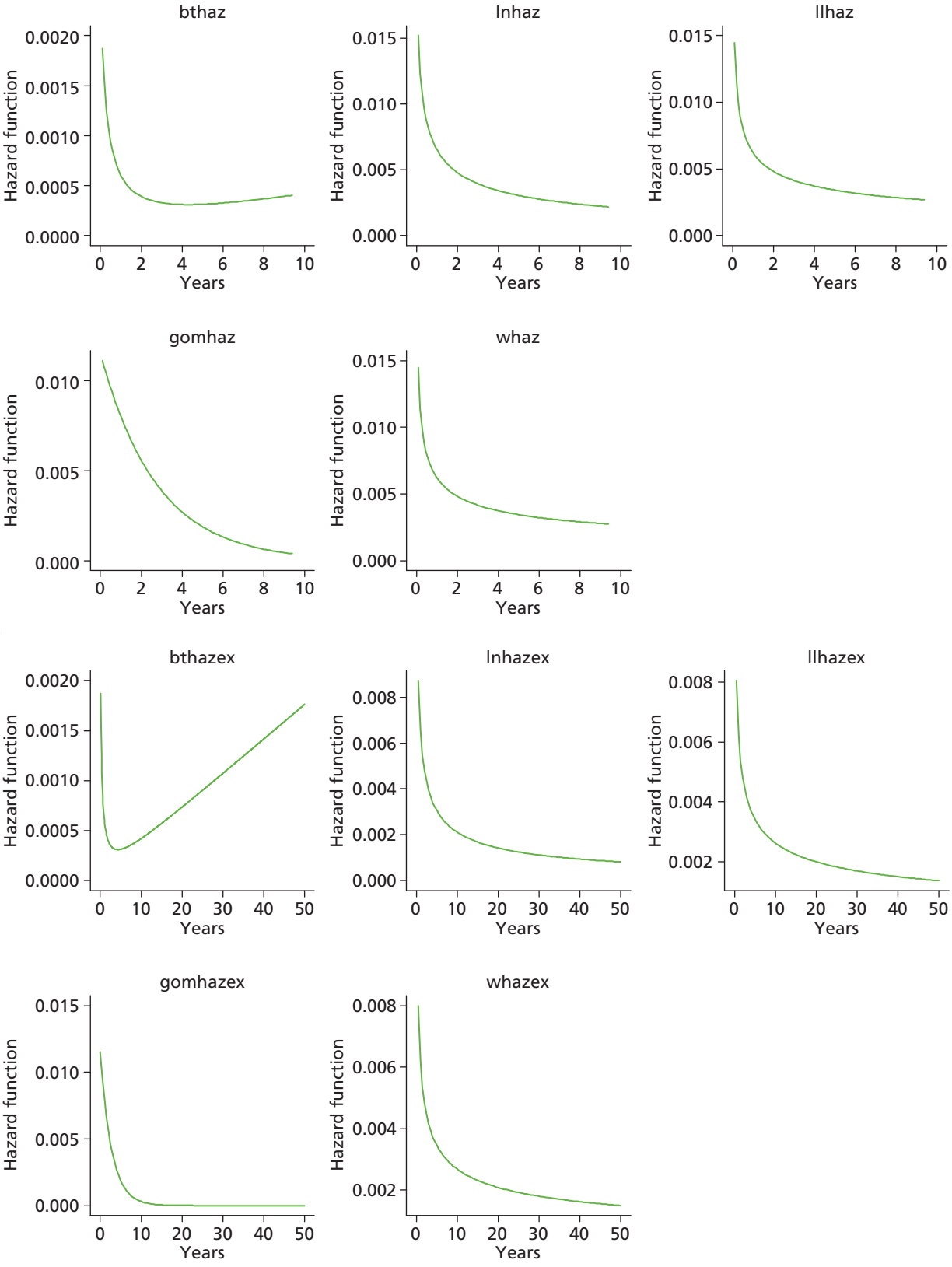




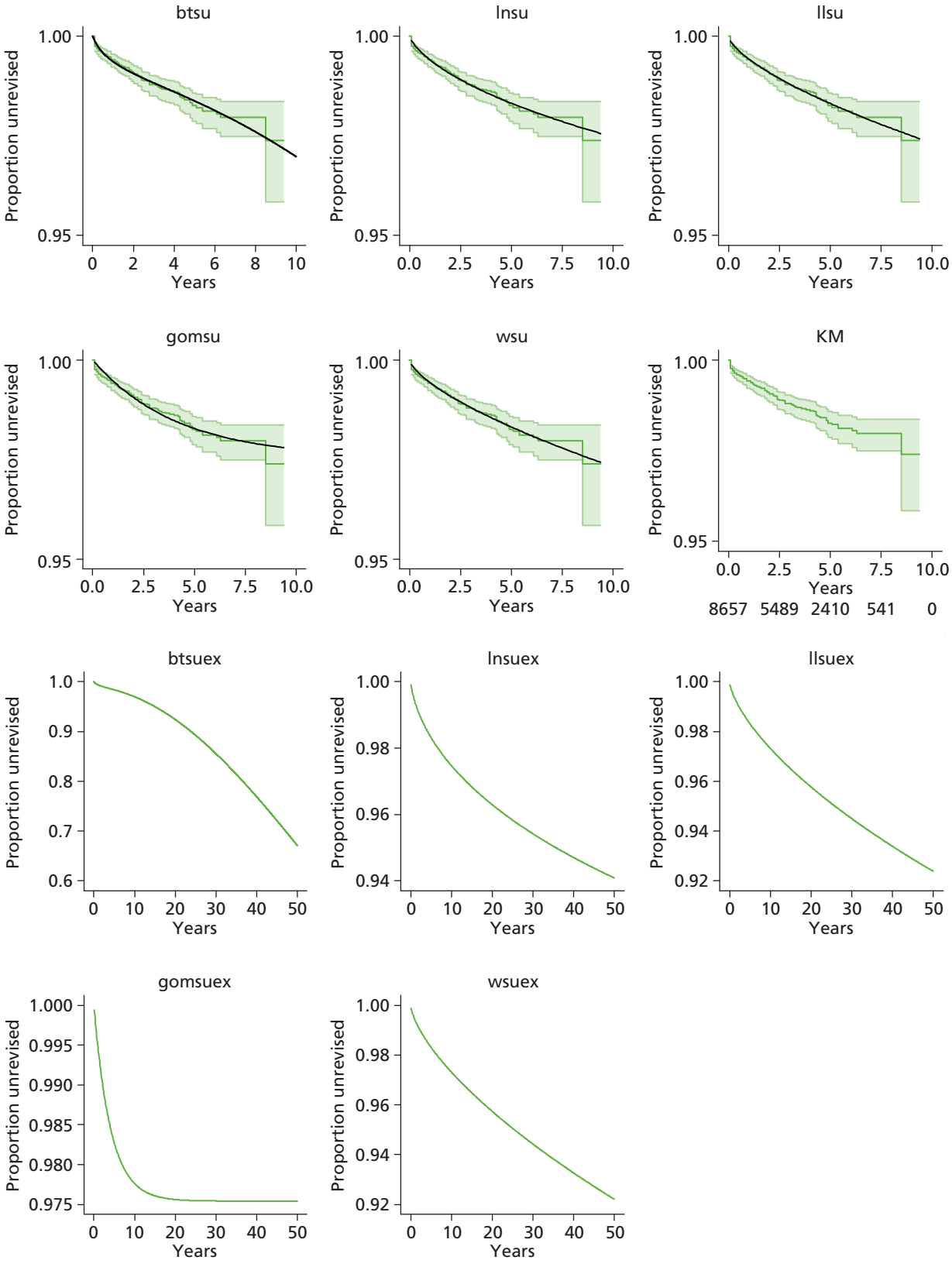
Total hip replacement categories: male aged > 65 years

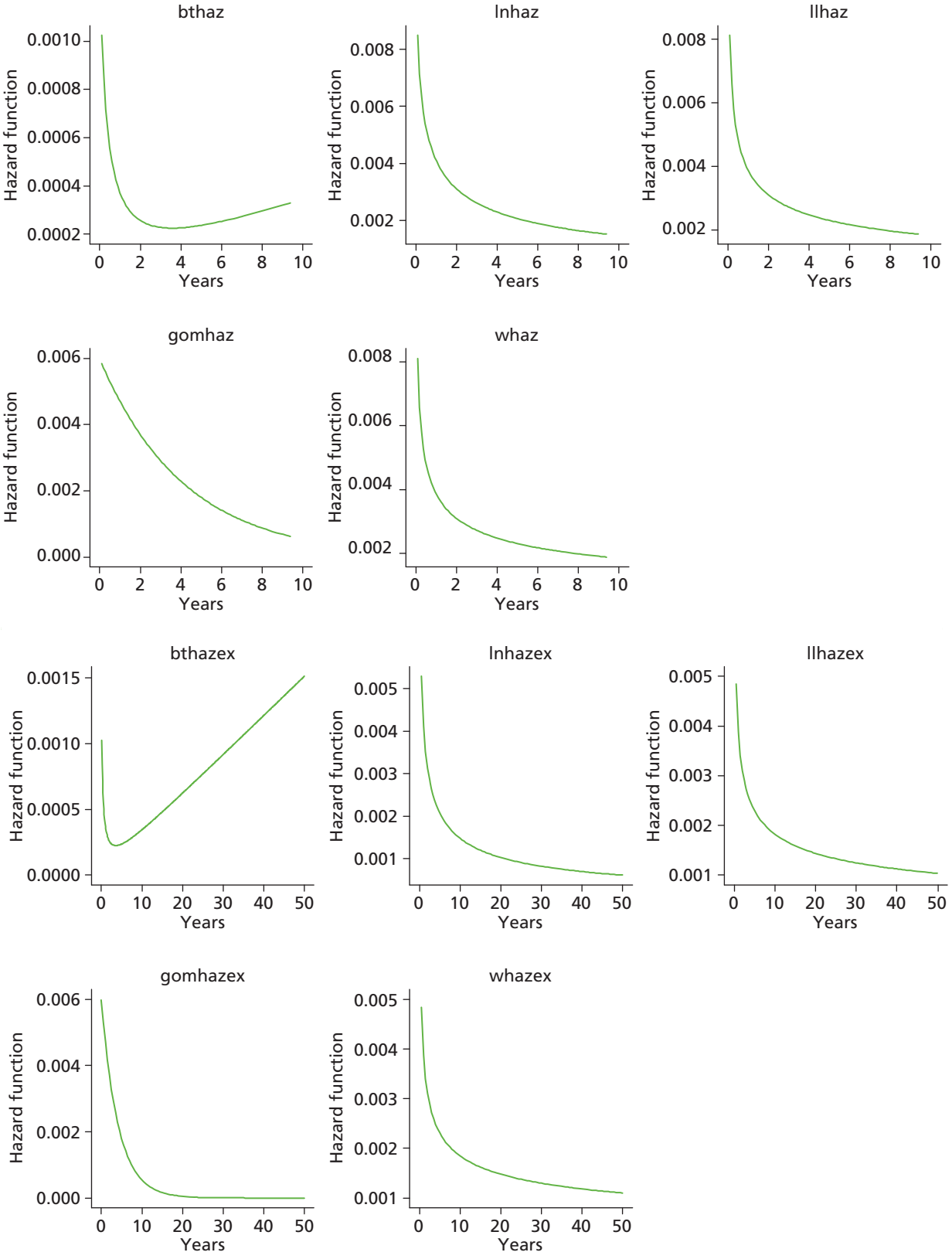
Category C [ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)]



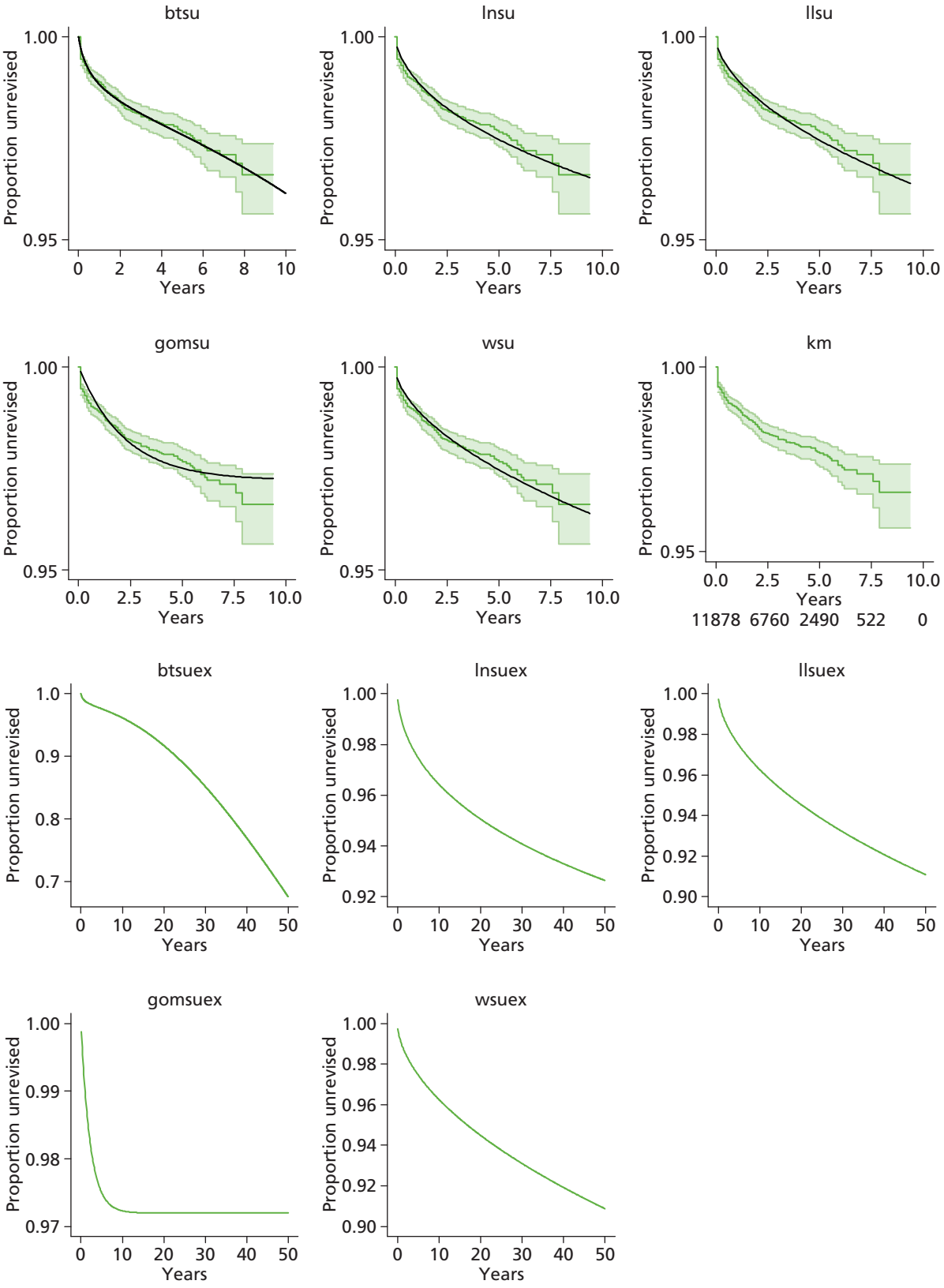


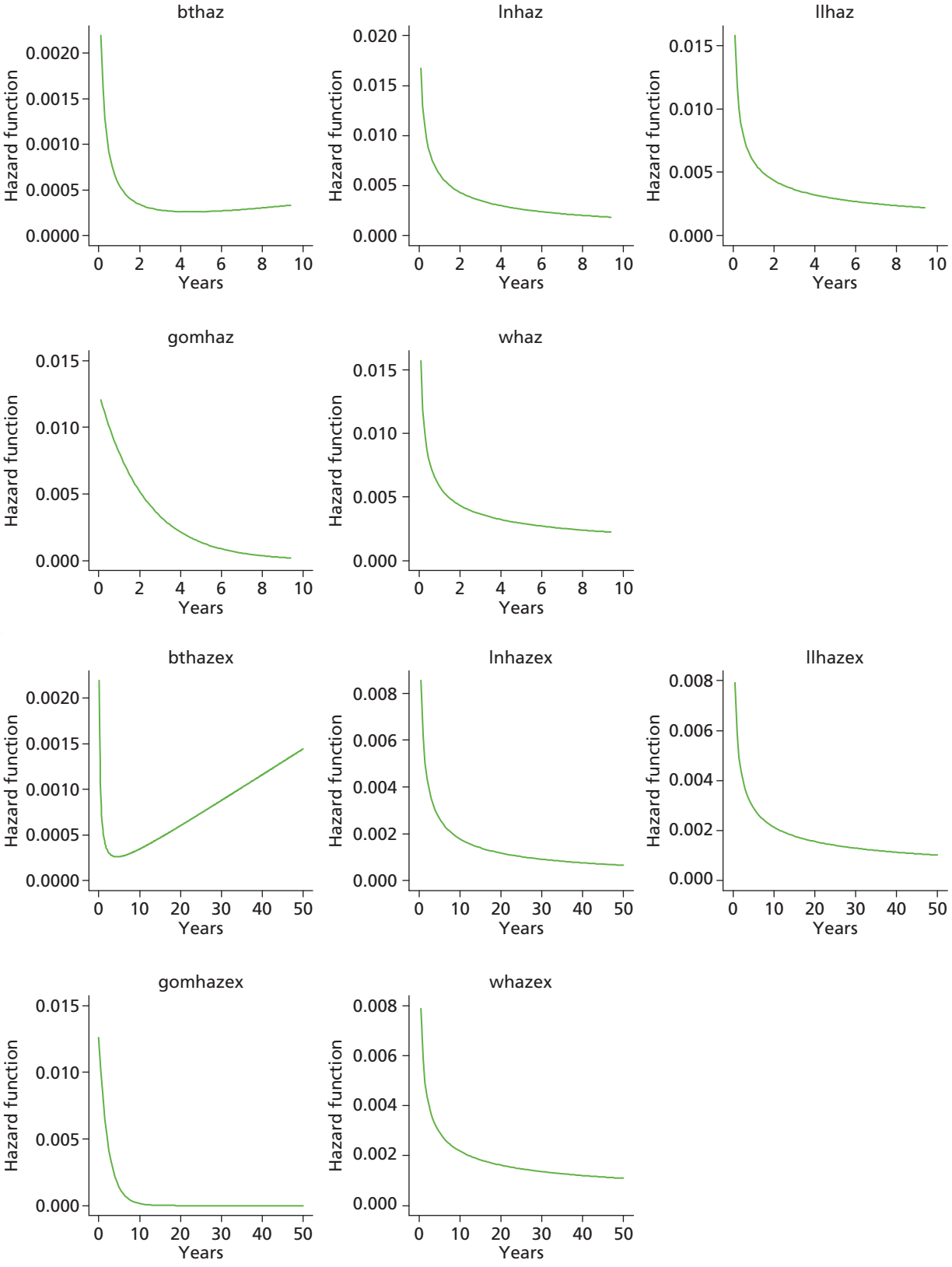
Category D [hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



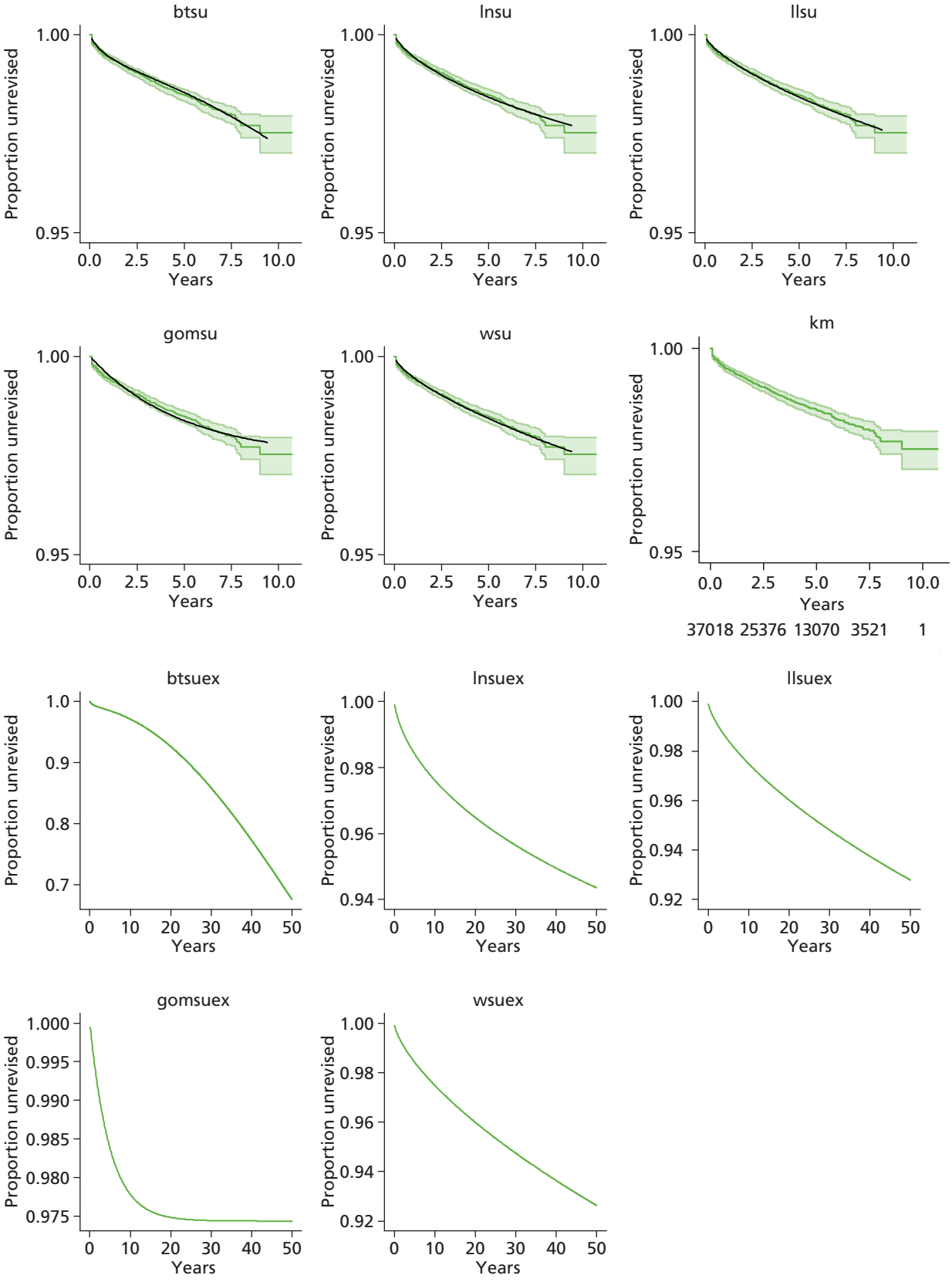


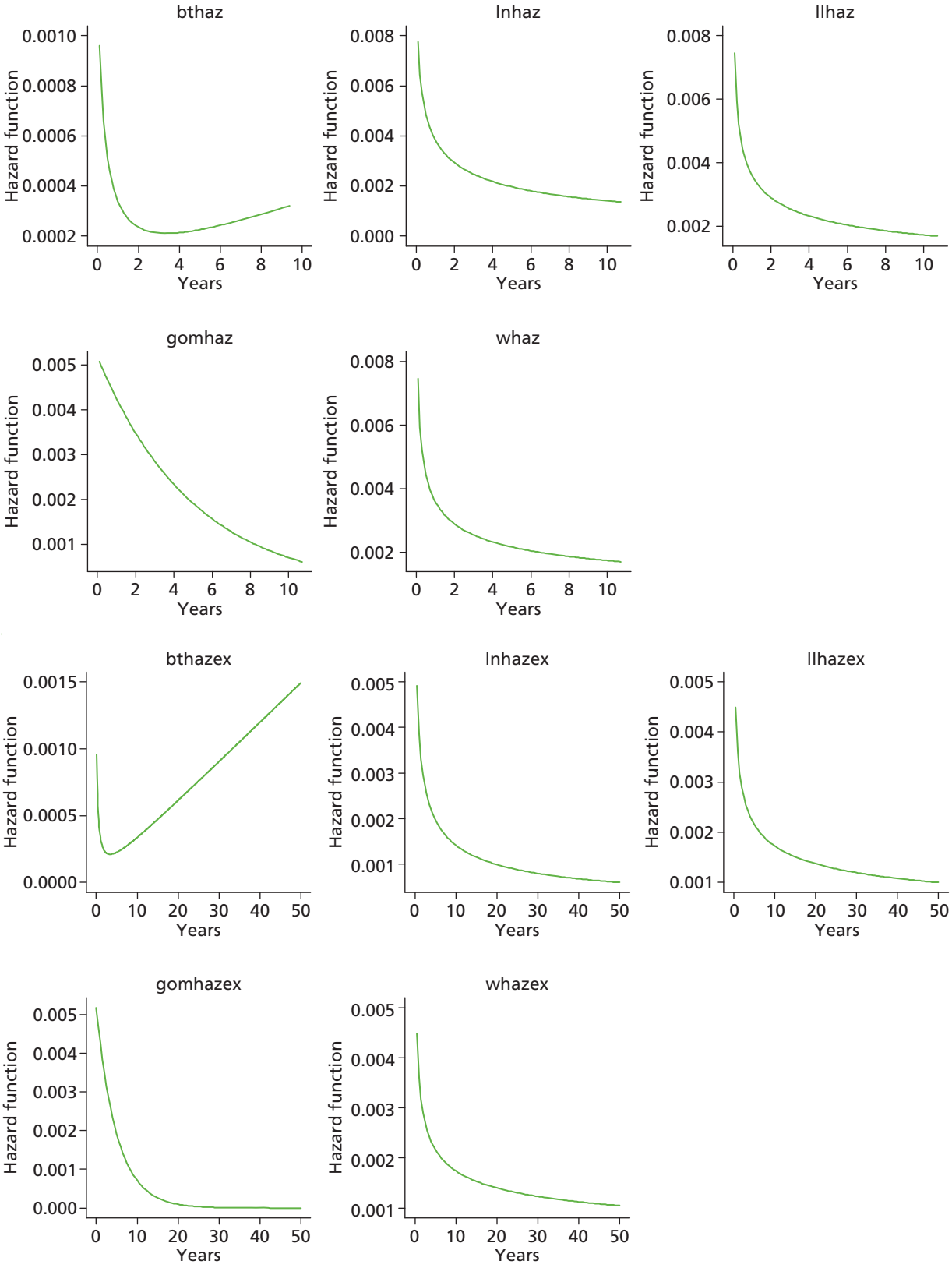
Category B [metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



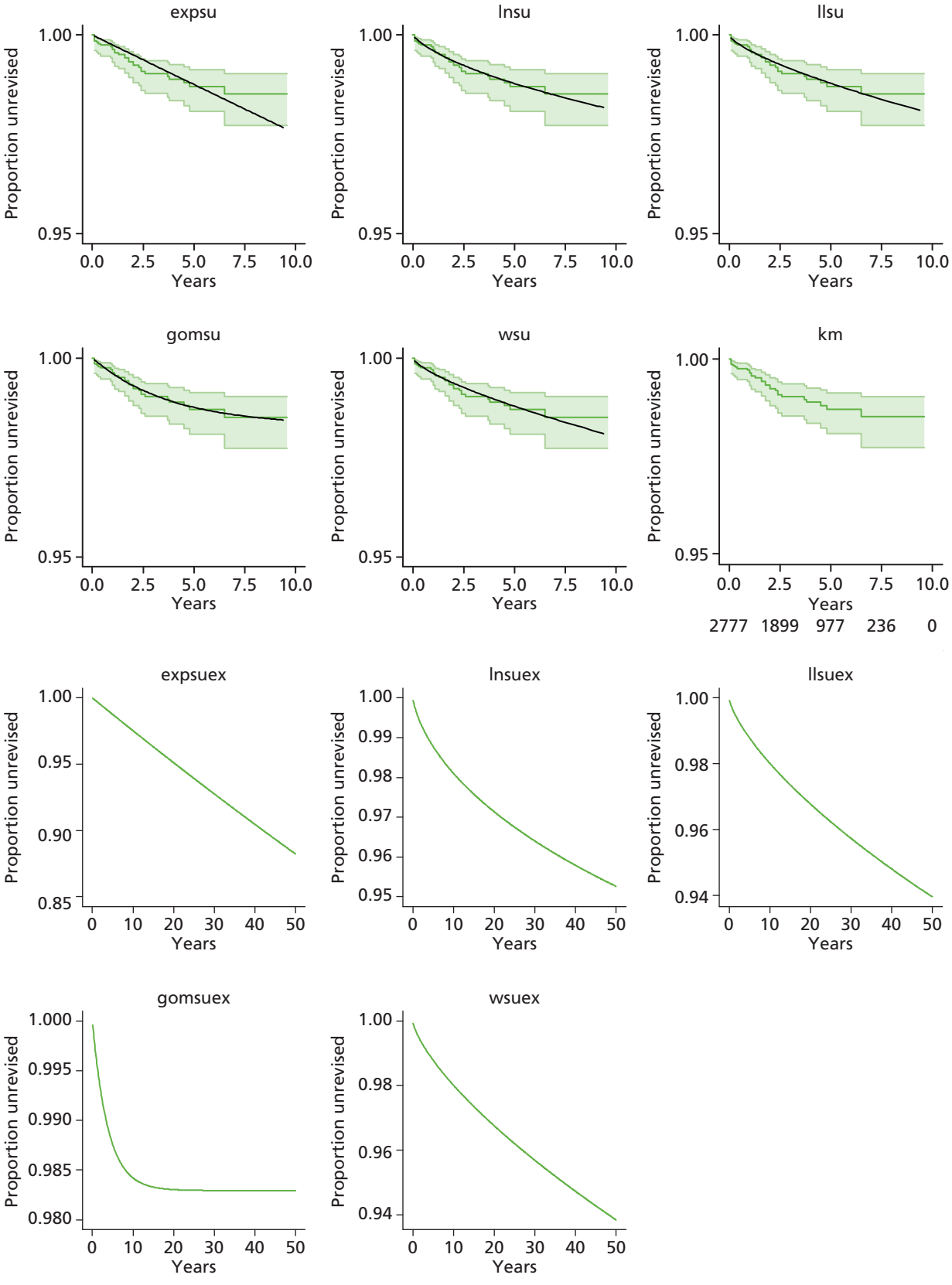


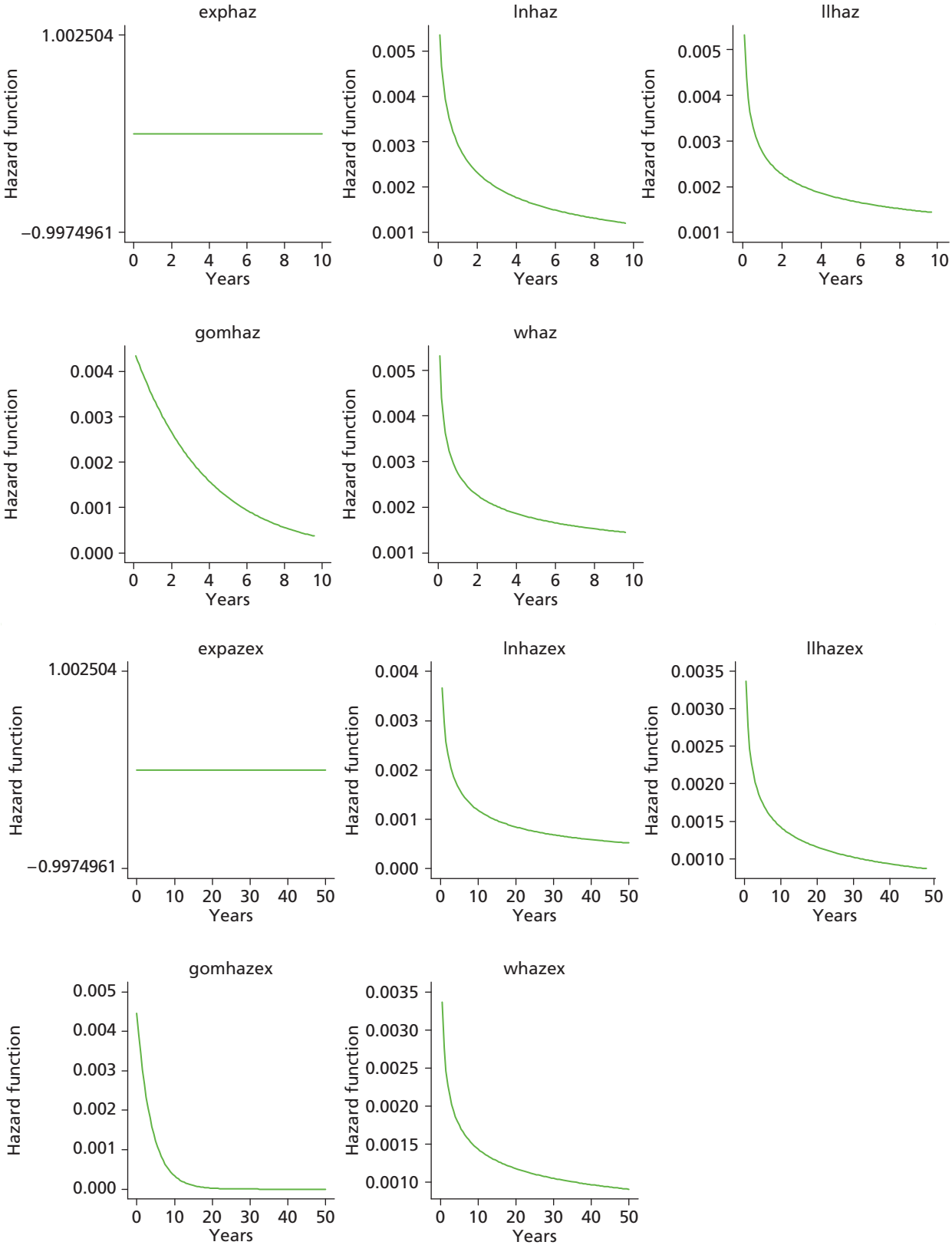
Category A [metal head (cemented stem) on cemented polyethylene cup]





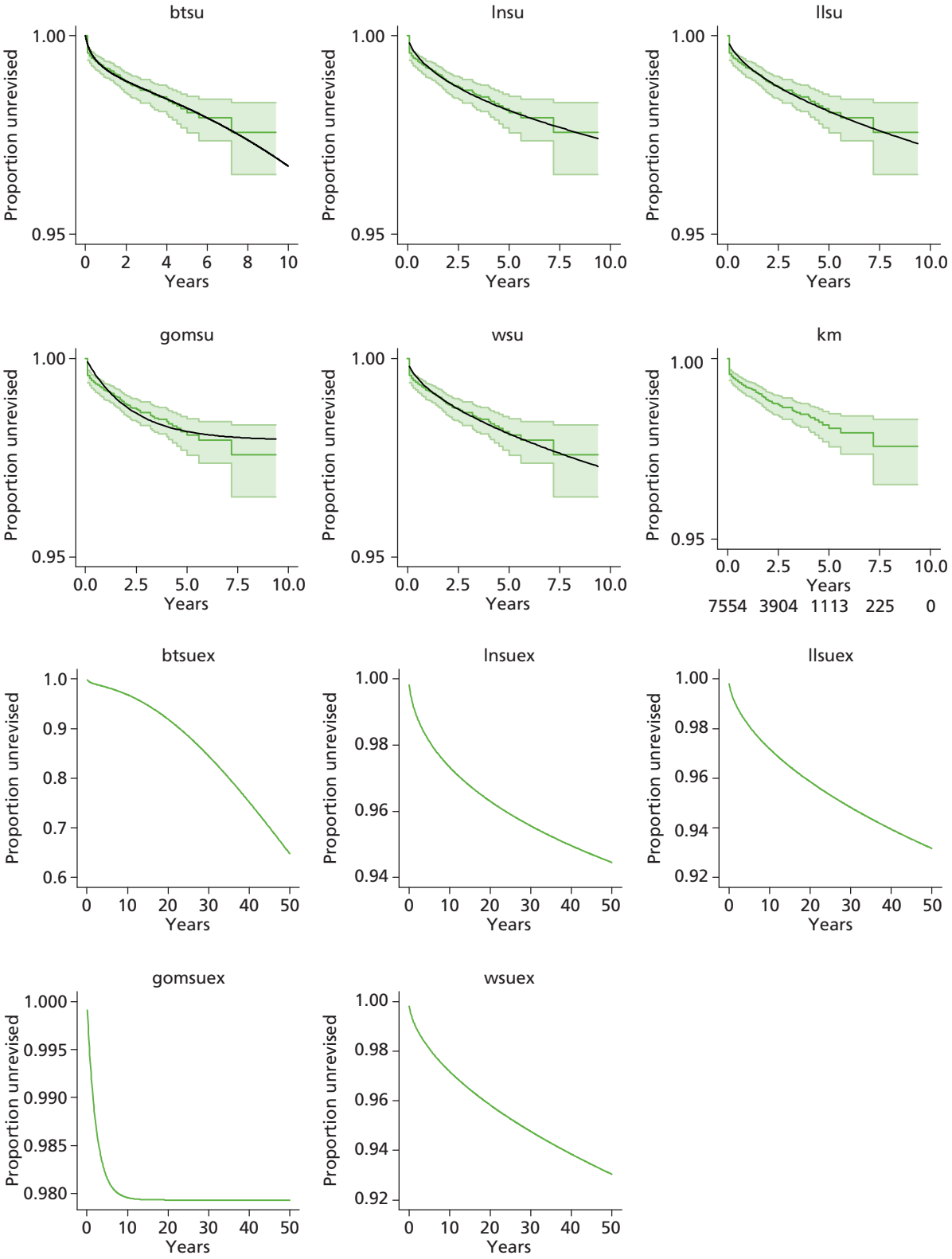
Category E [cemented head (cemented stem) on cemented polyethylene cup]

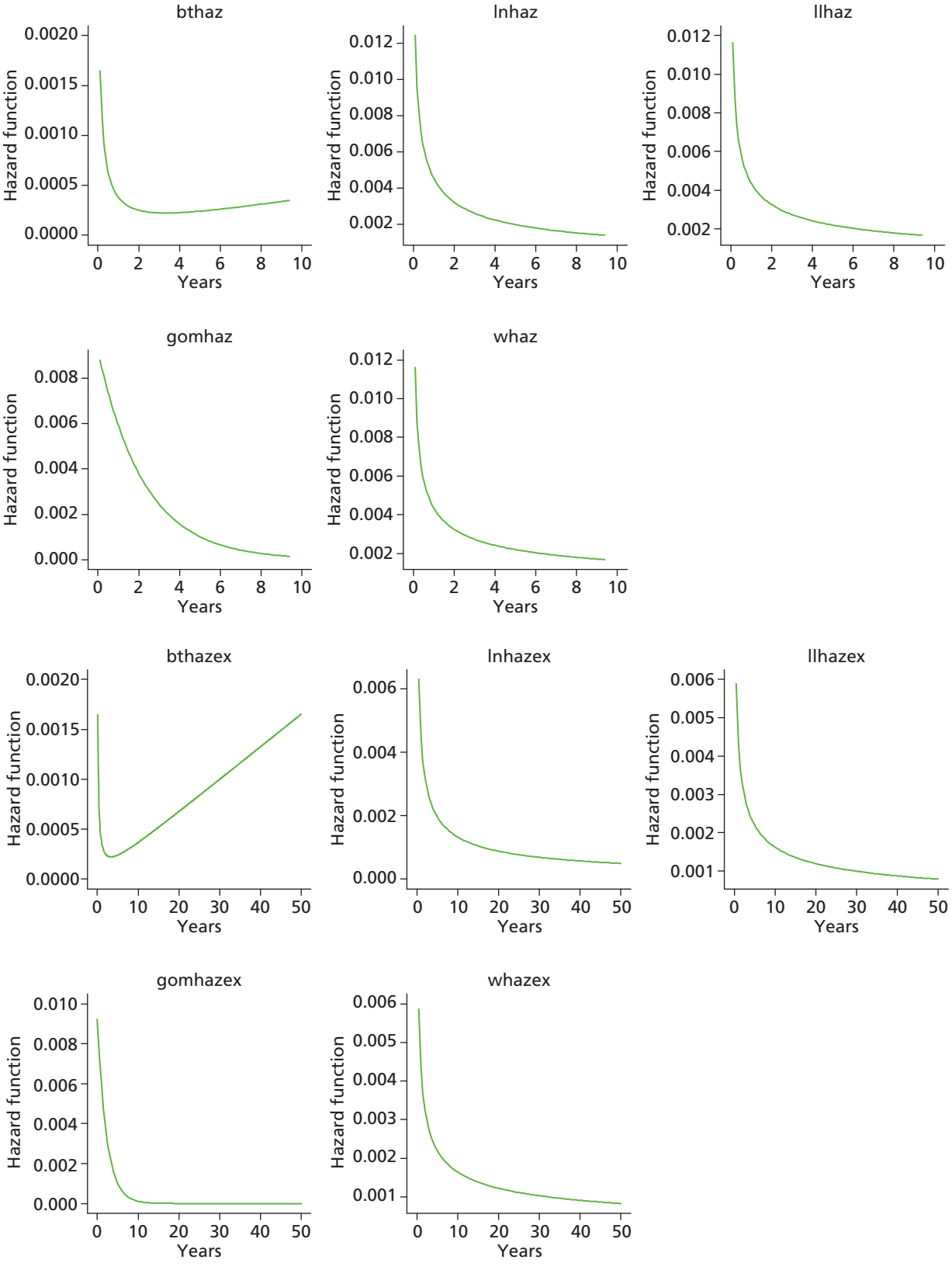




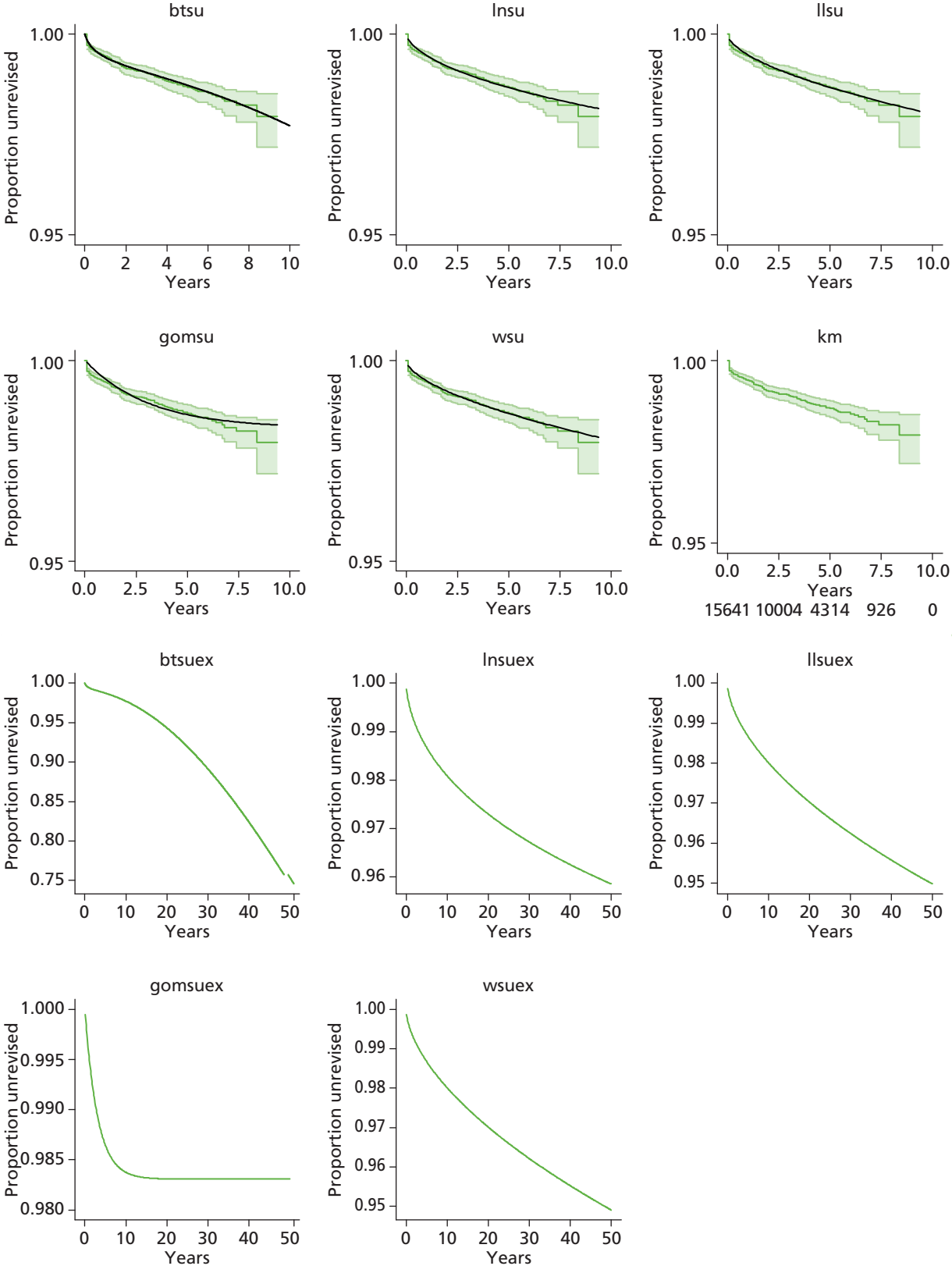
Total hip replacement categories: female aged > 65 years

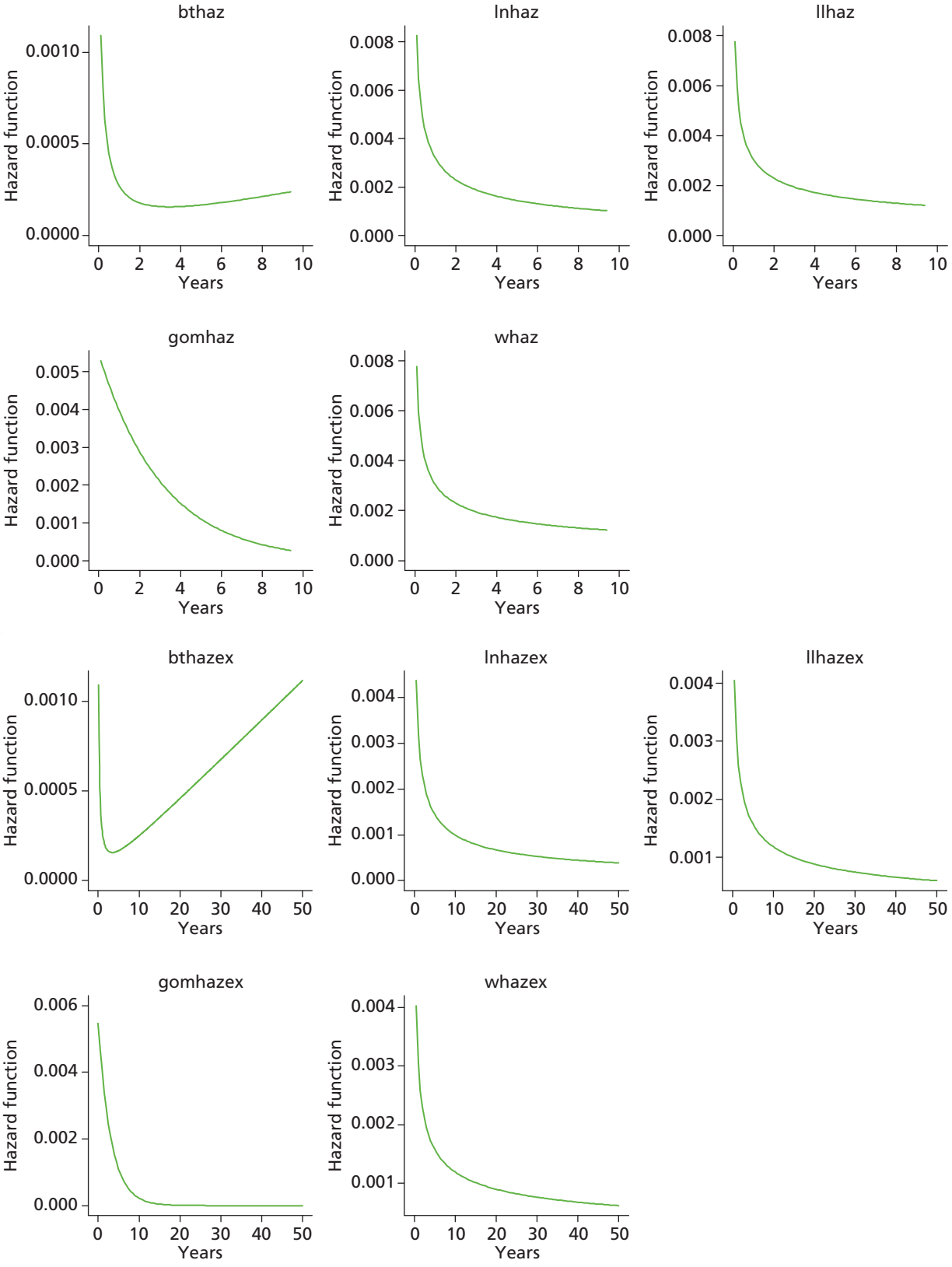
Category C [ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)]



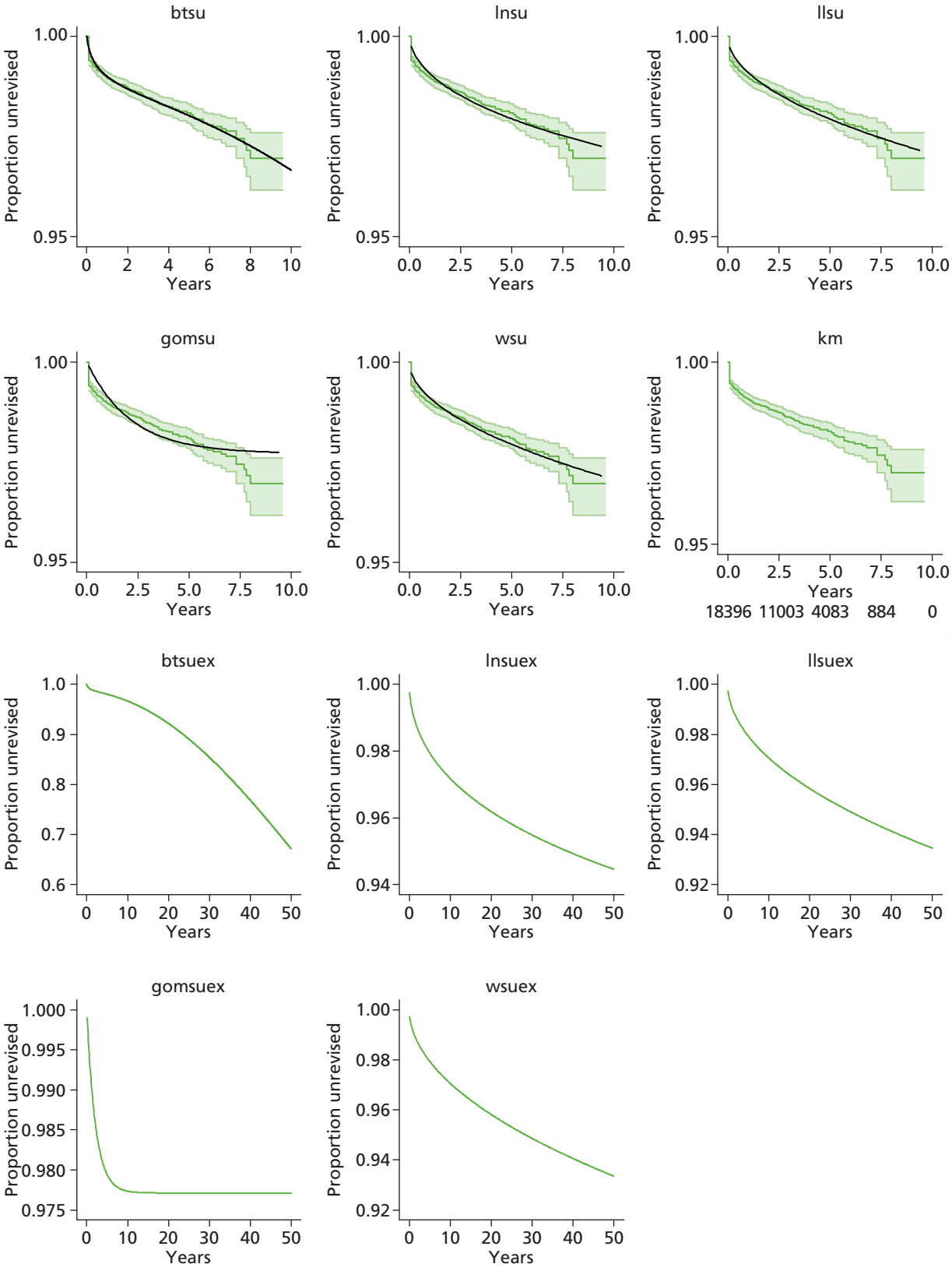


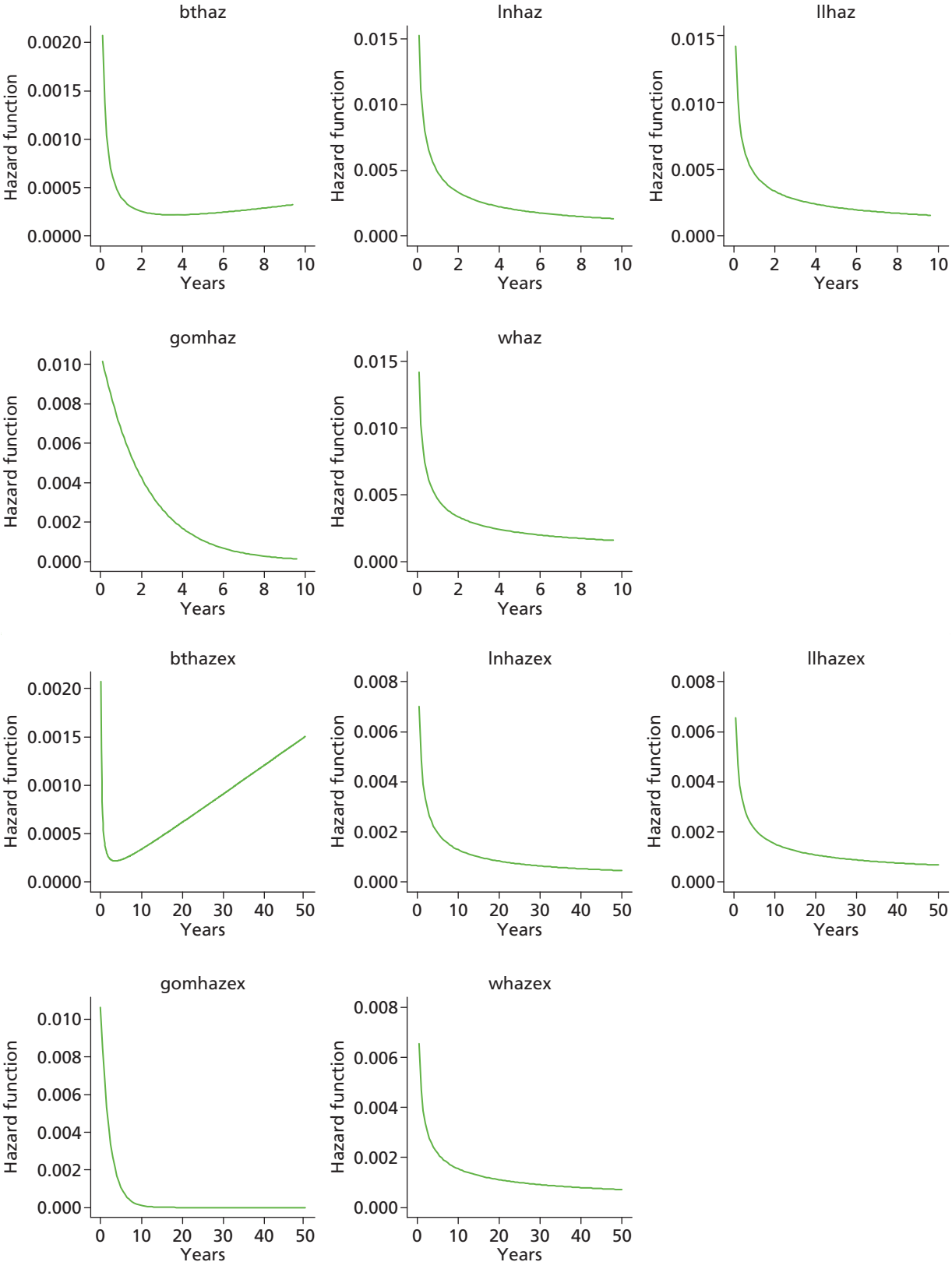
Category D [hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



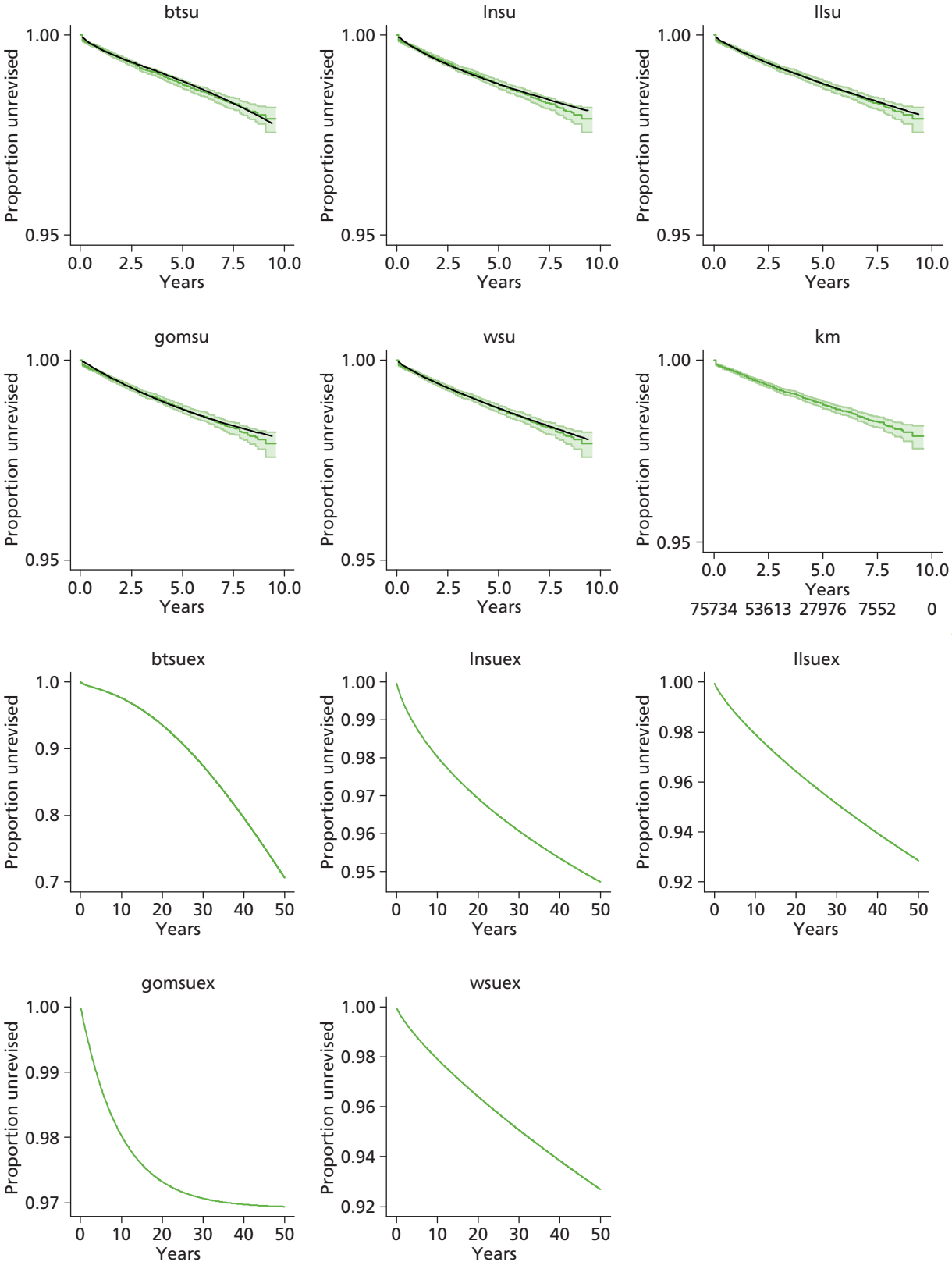


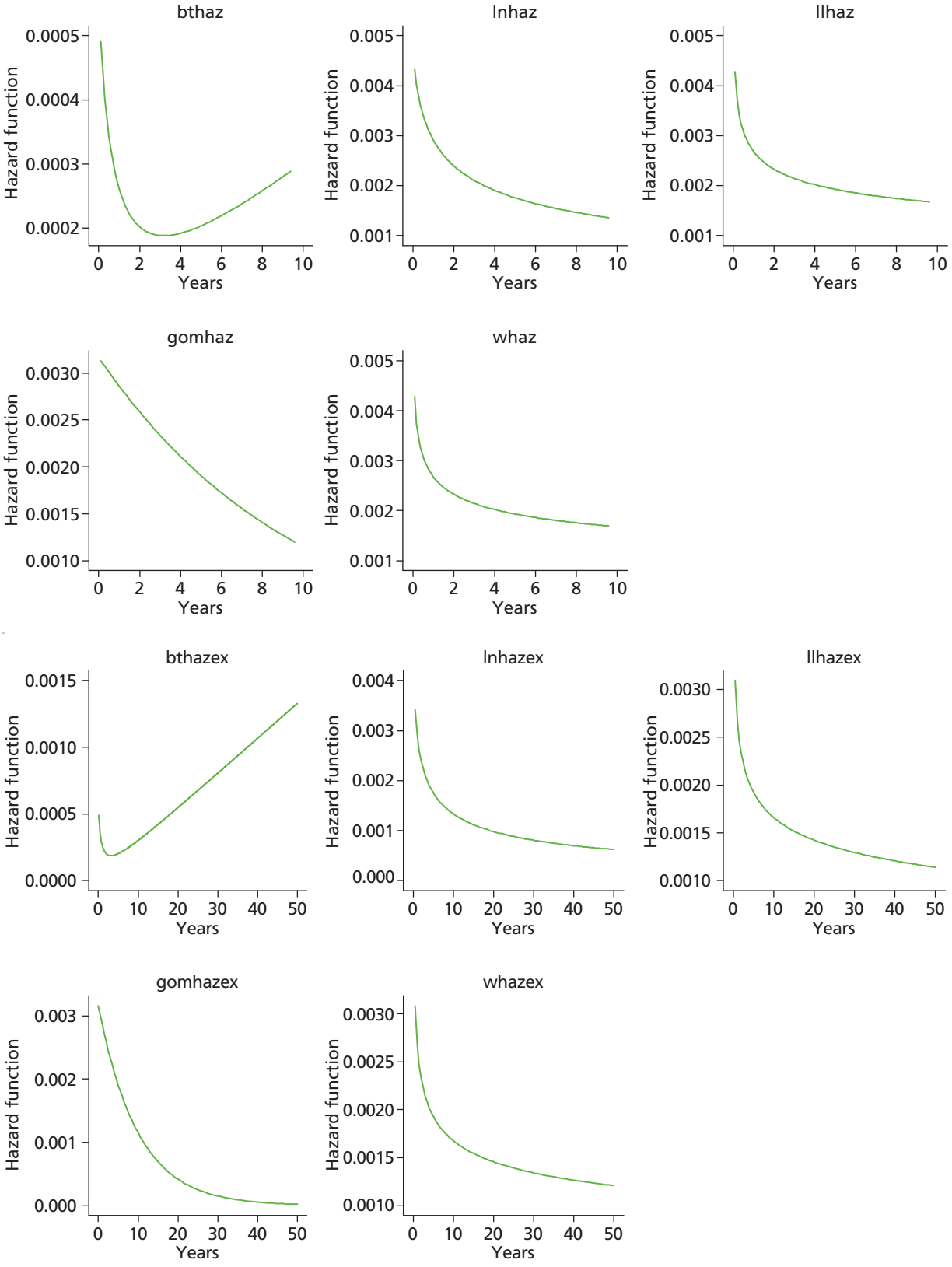
Category B [metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



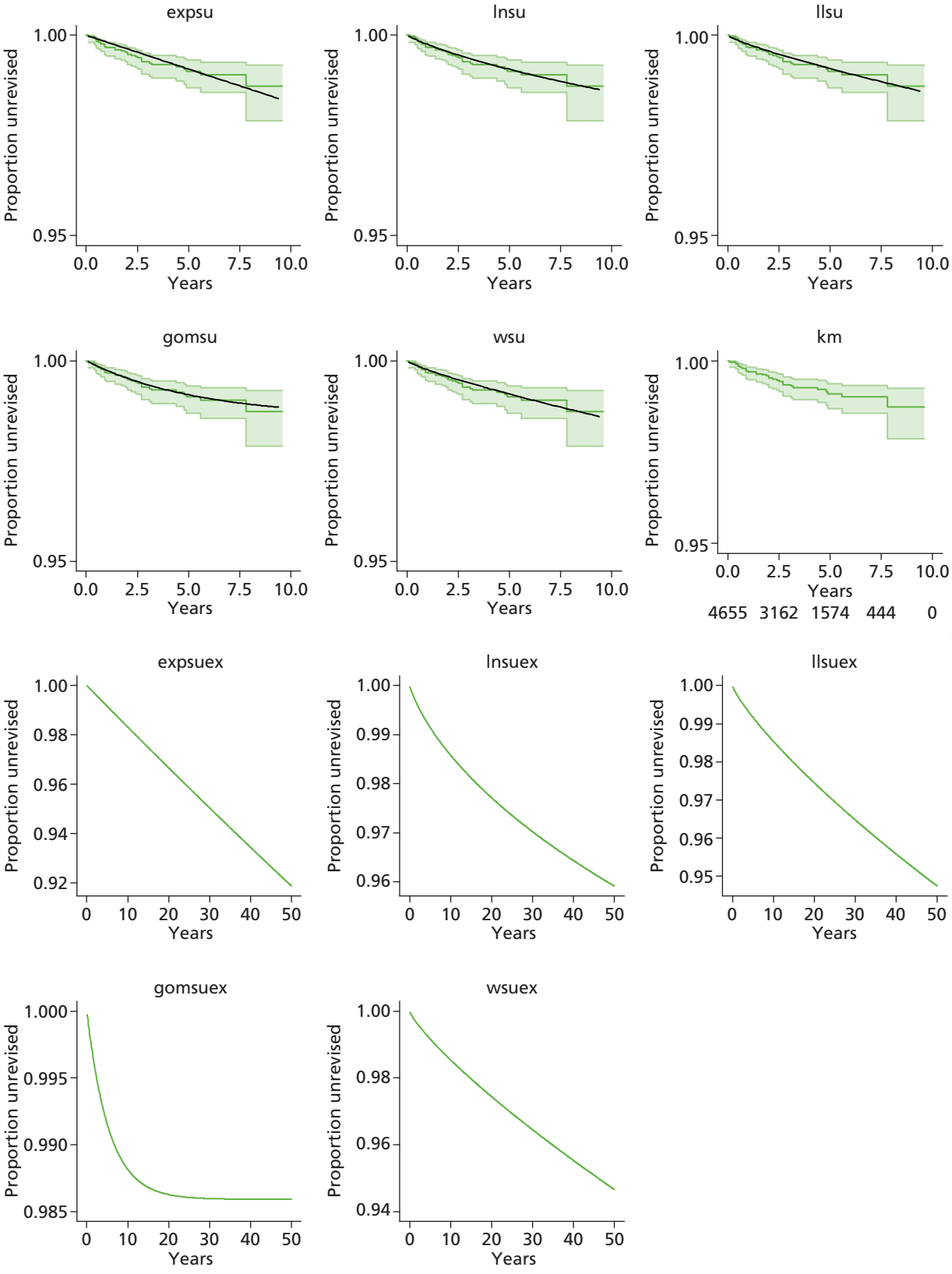


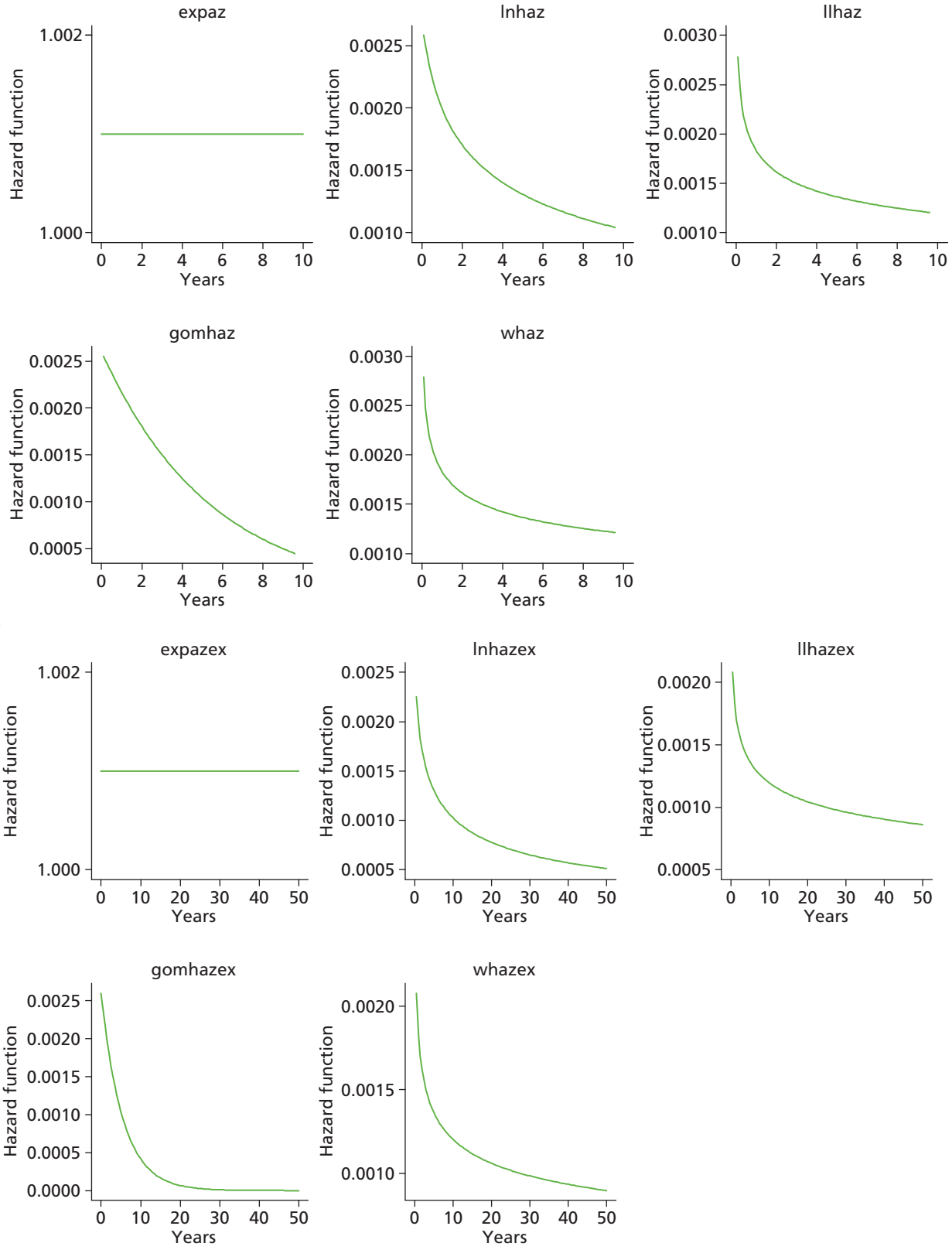
Category A [metal head (cemented stem) on cemented polyethylene cup]





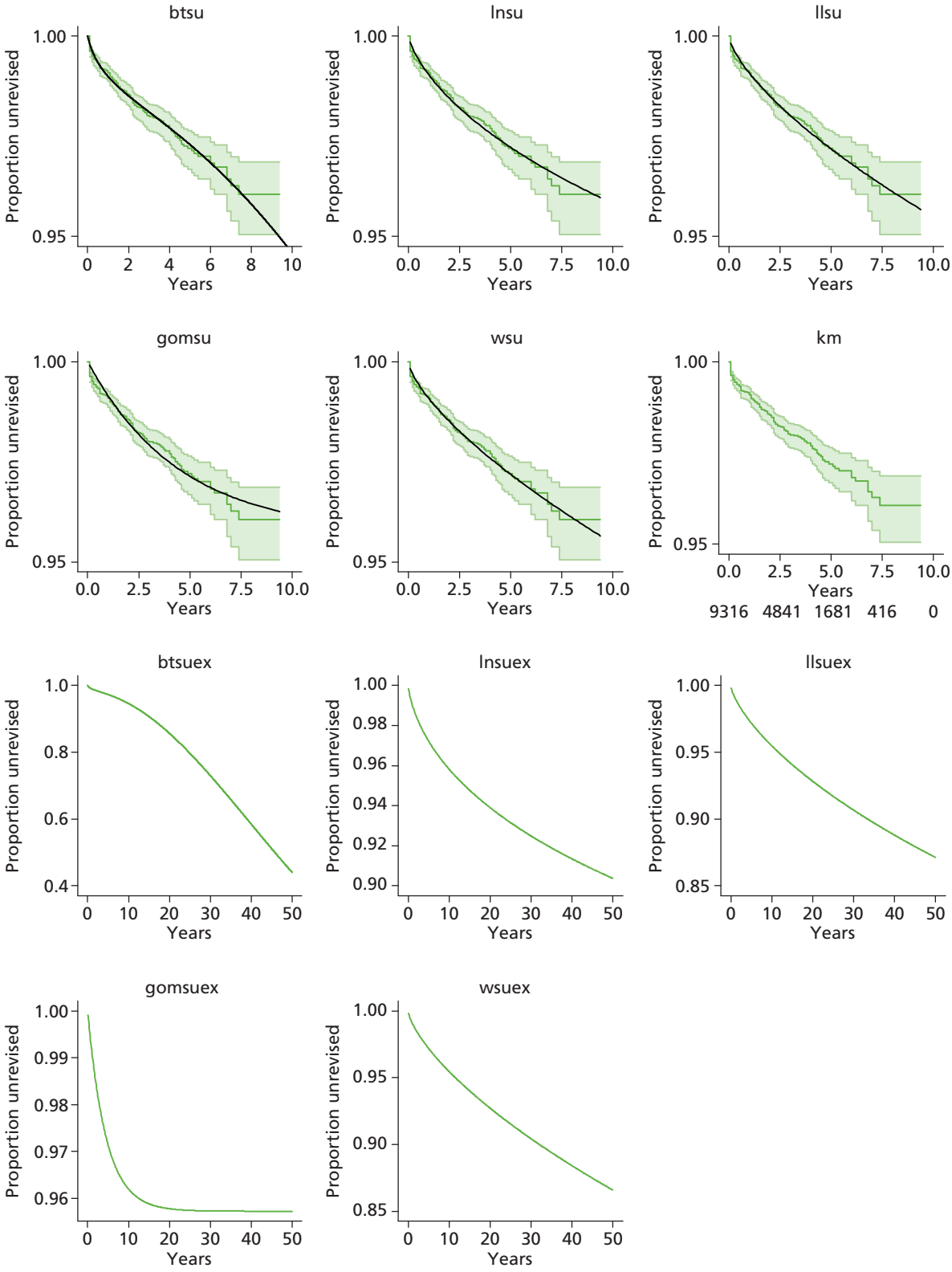
Category E [ceramic head (cemented stem) on cemented polyethylene cup]

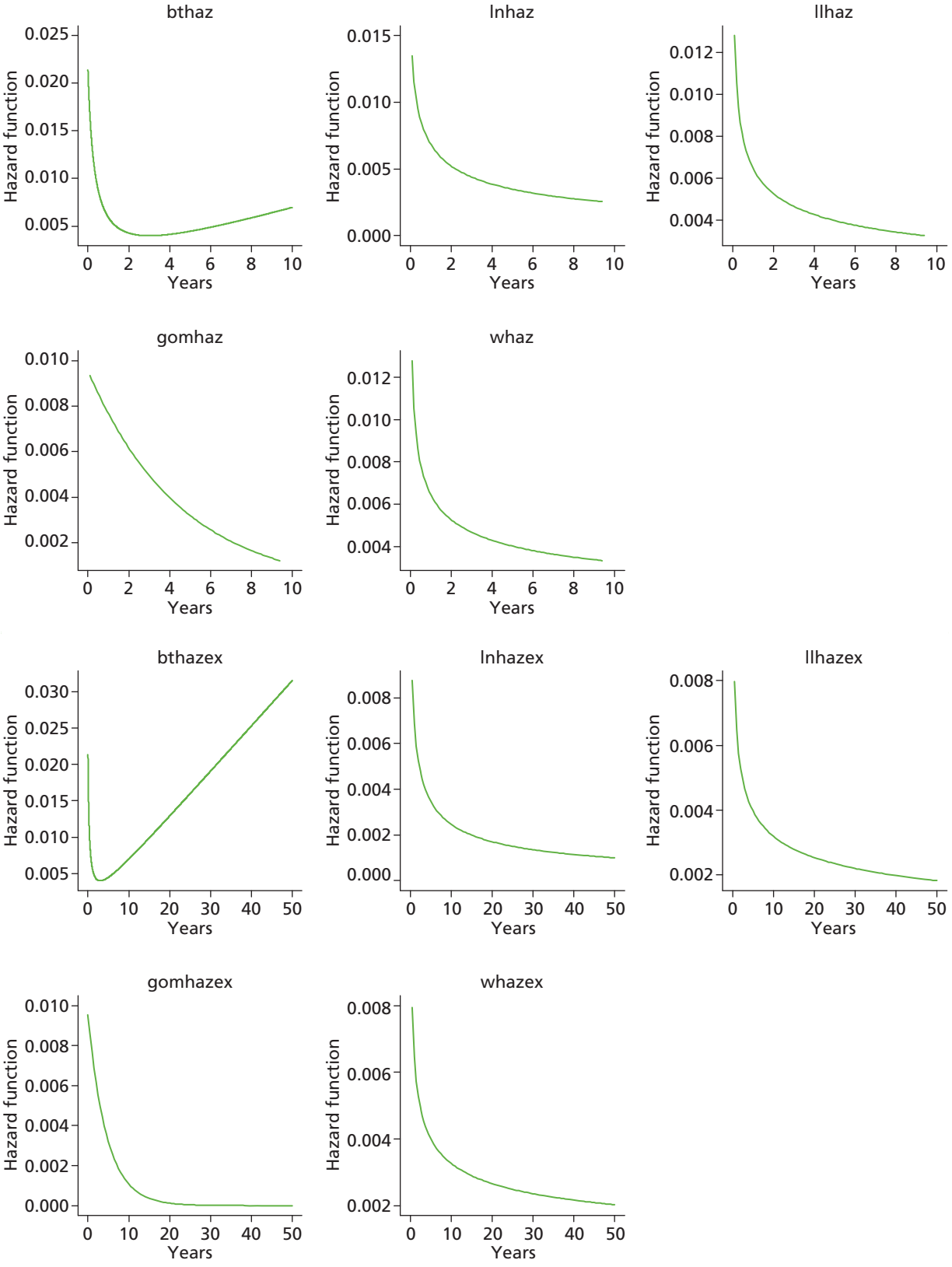




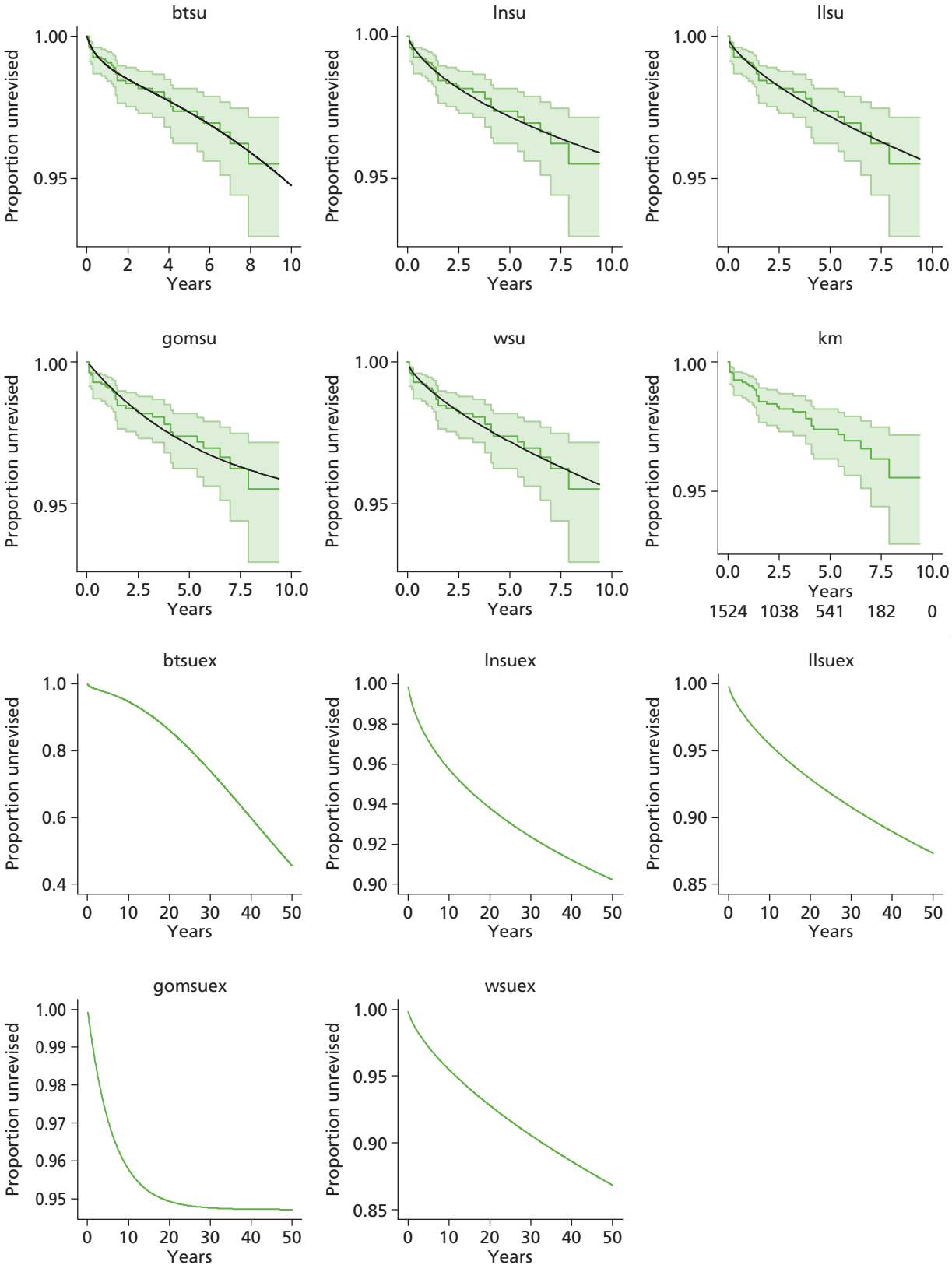
Total hip replacement categories: male aged < 65 years

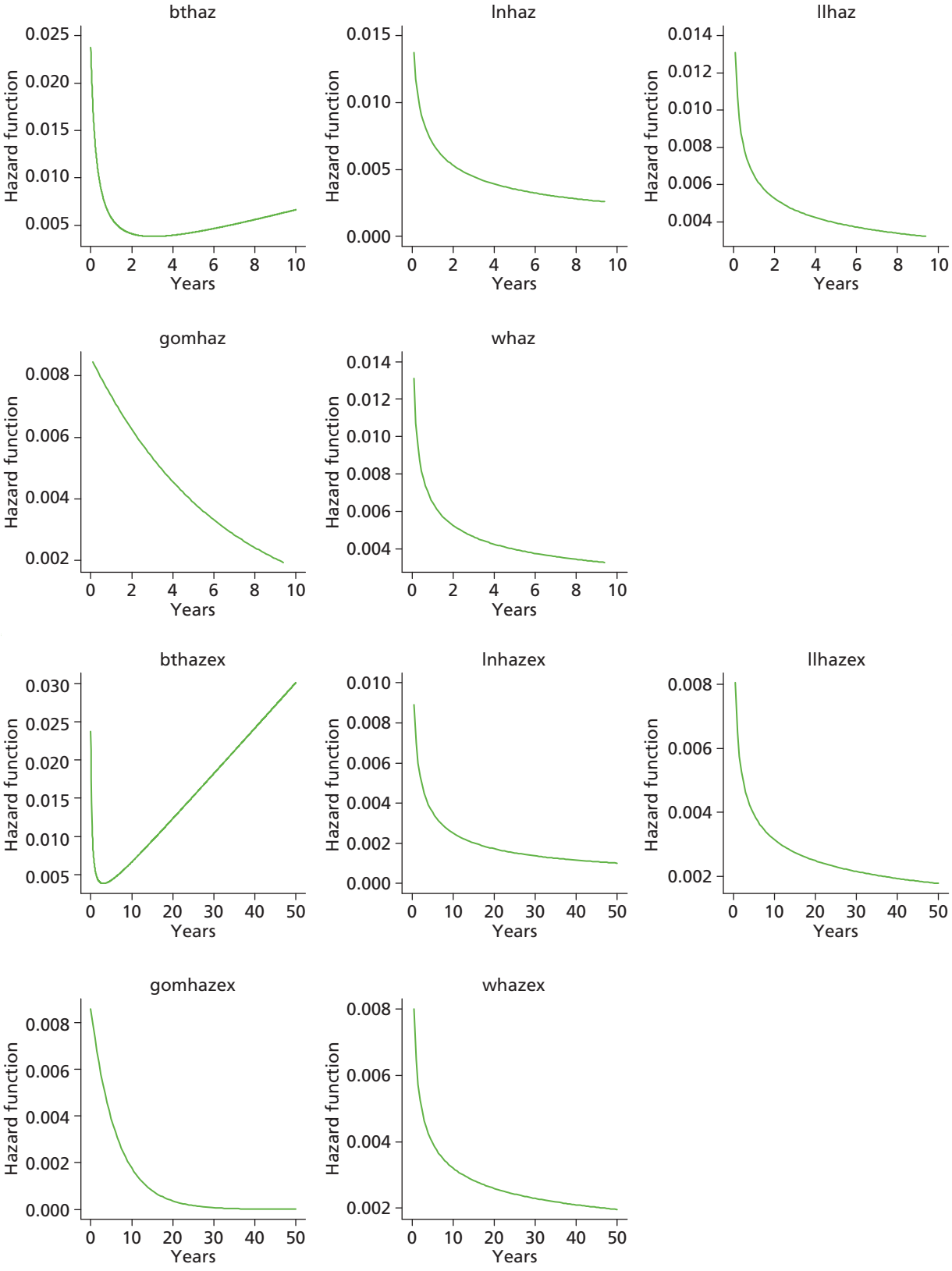
Category C [ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner]



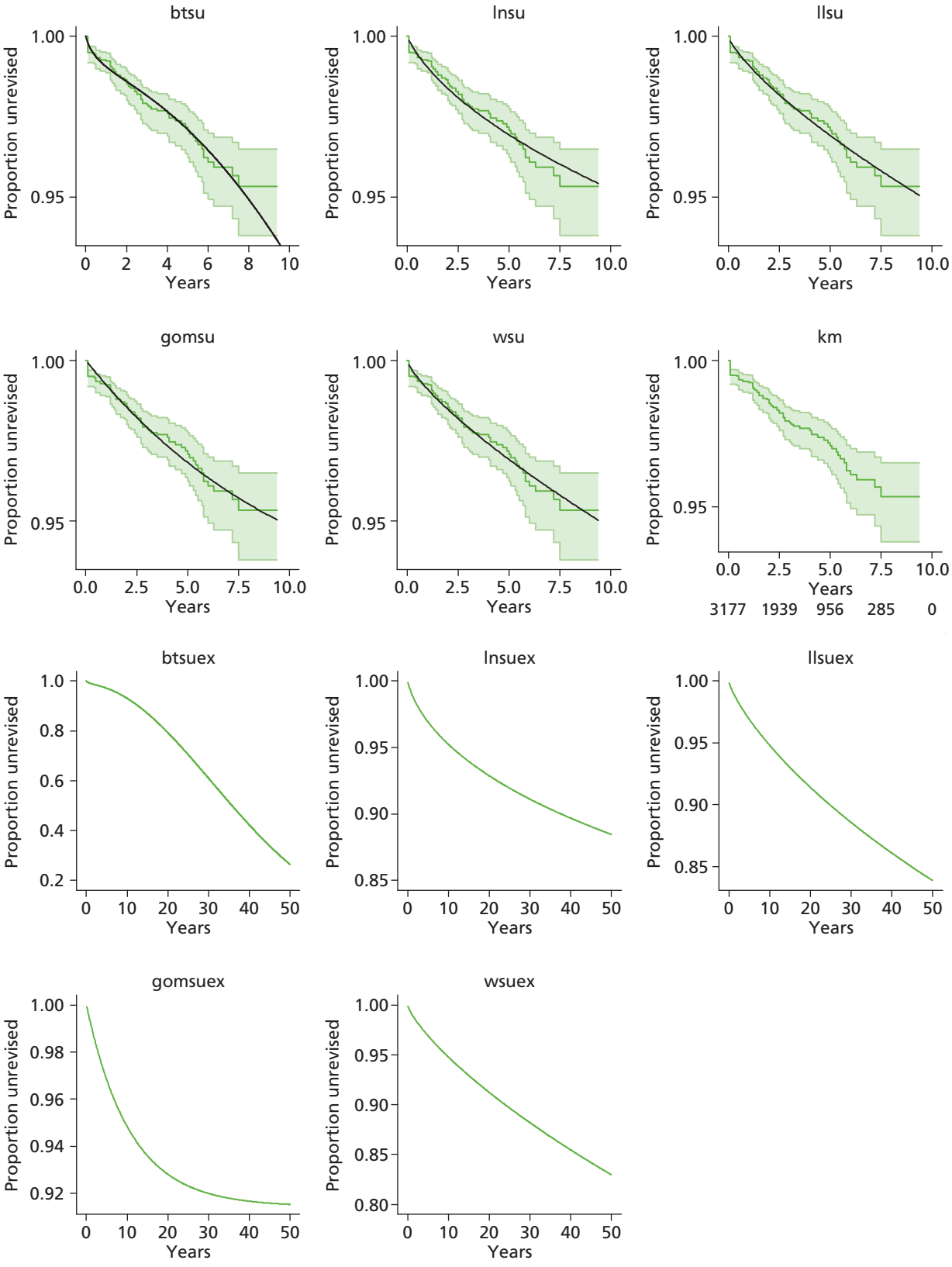


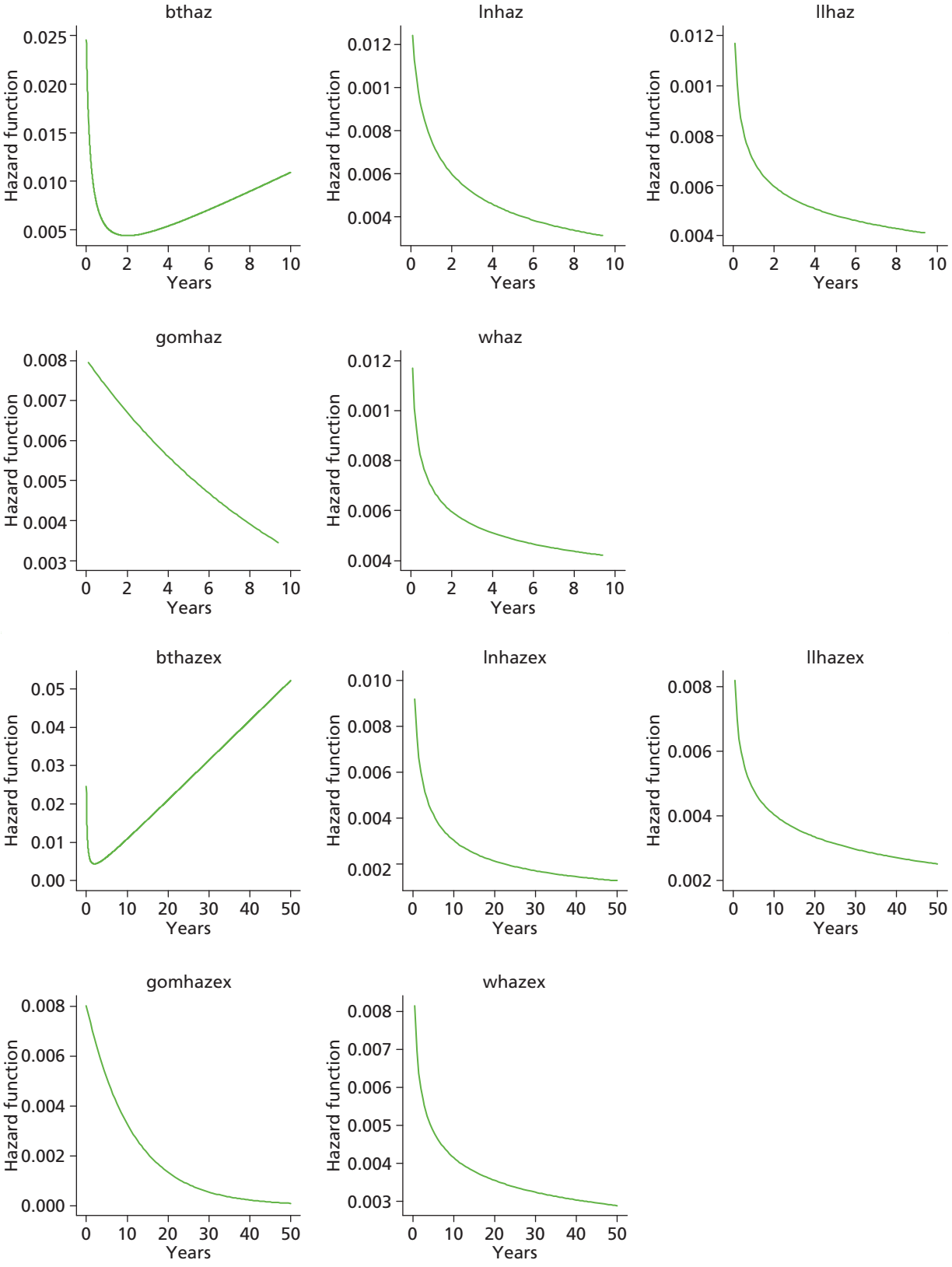
Category D [hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



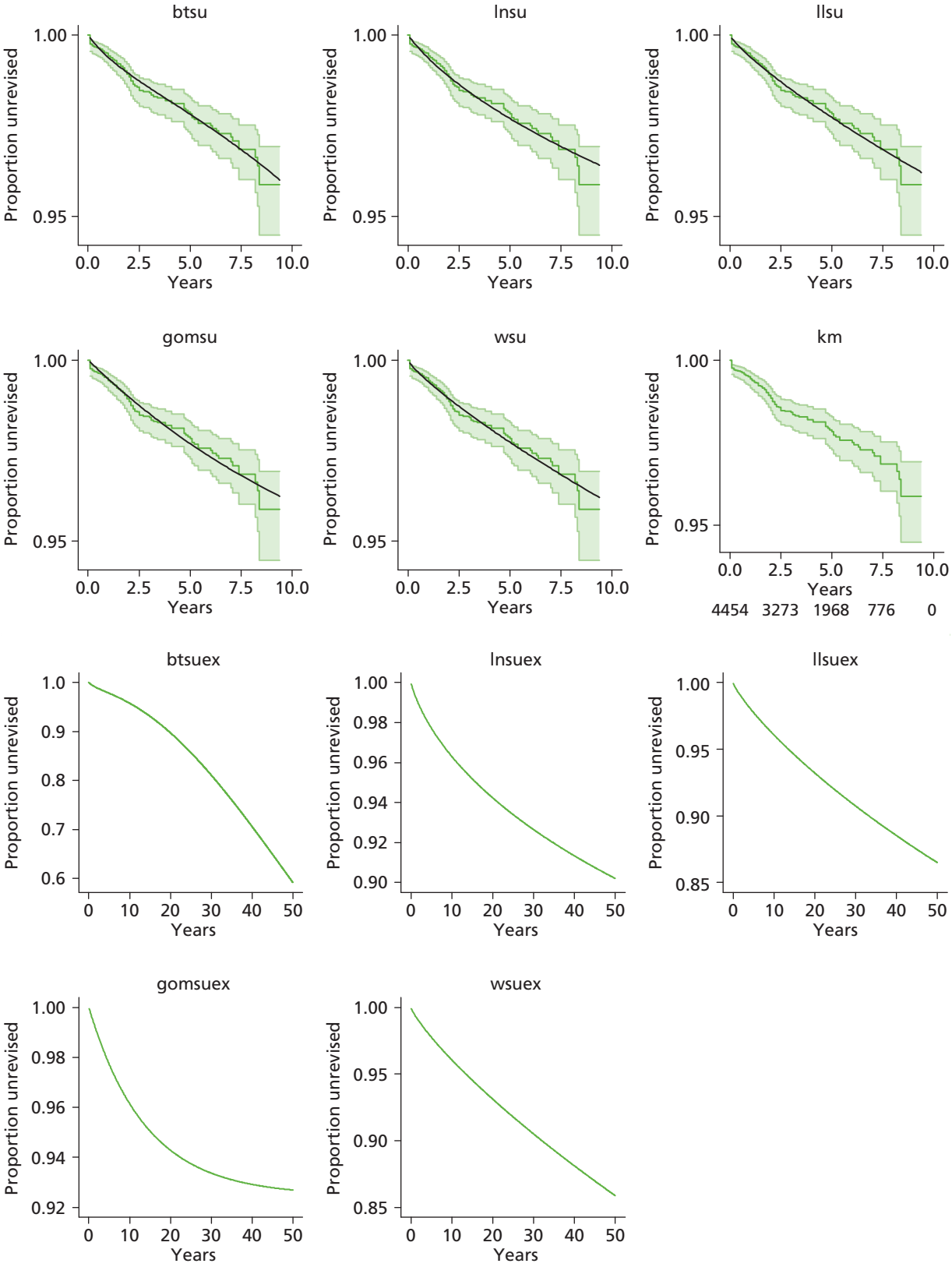


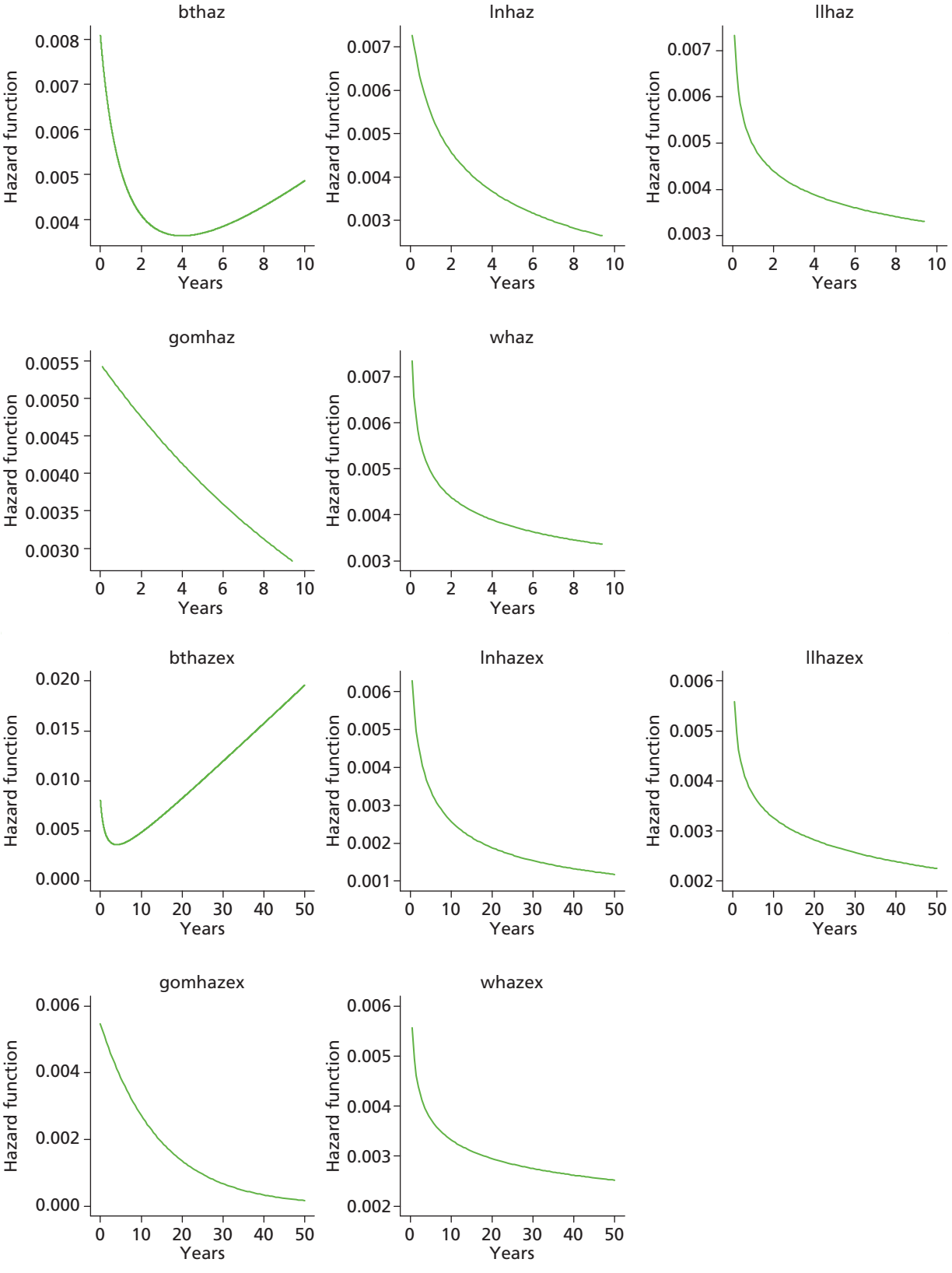
Category B [metal head (cementless stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



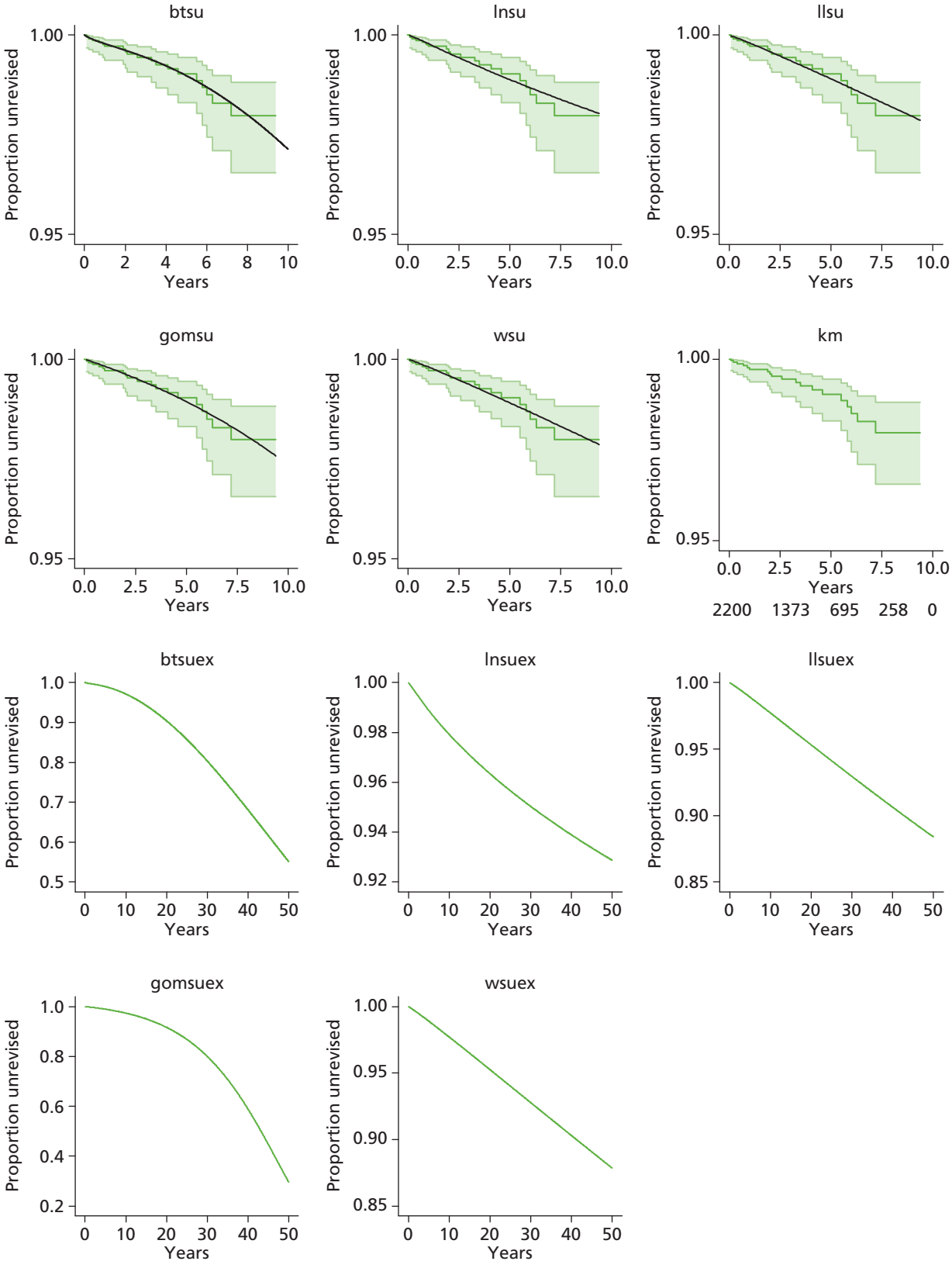


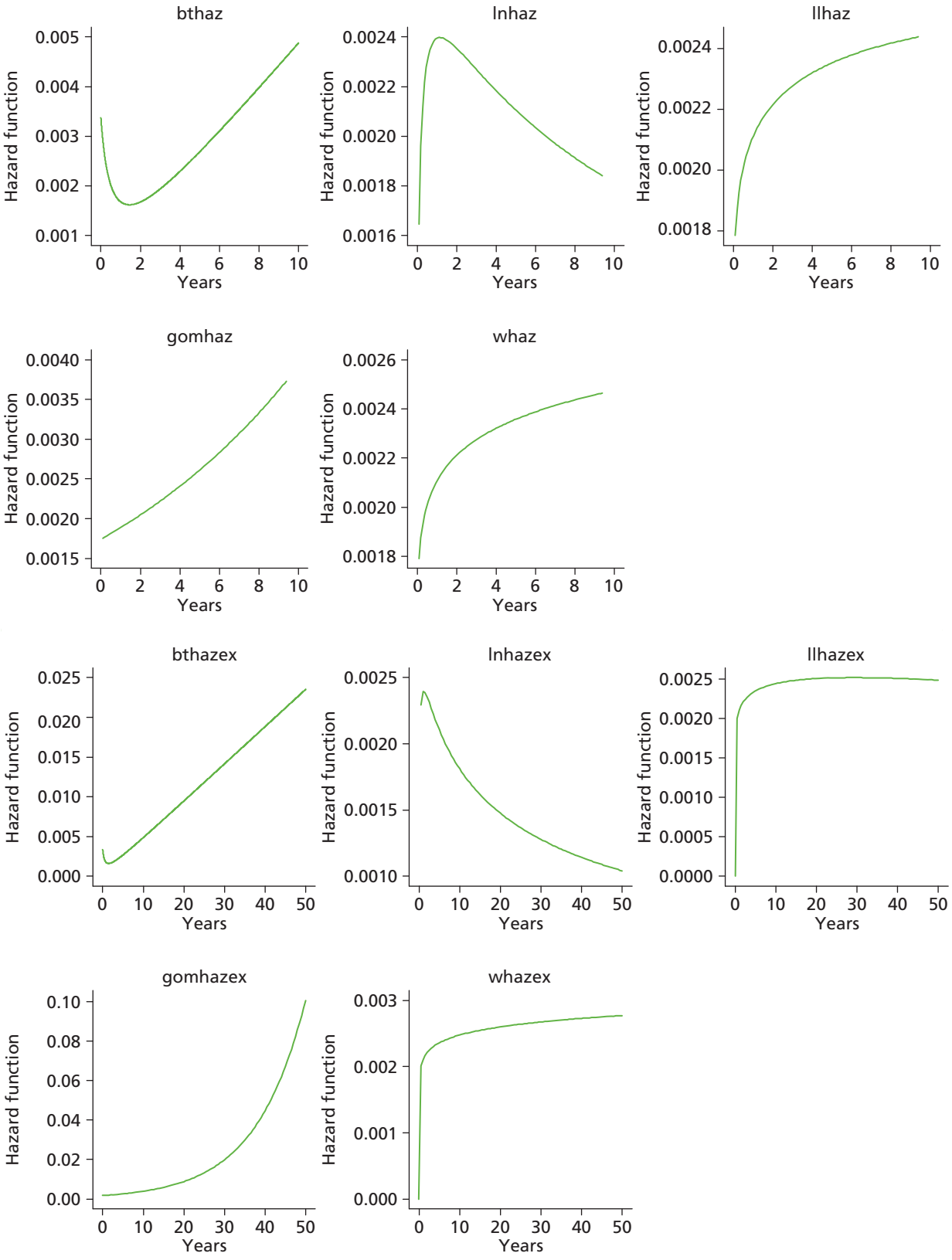
Category A [metal head (cemented stem) on cemented polyethylene cup]





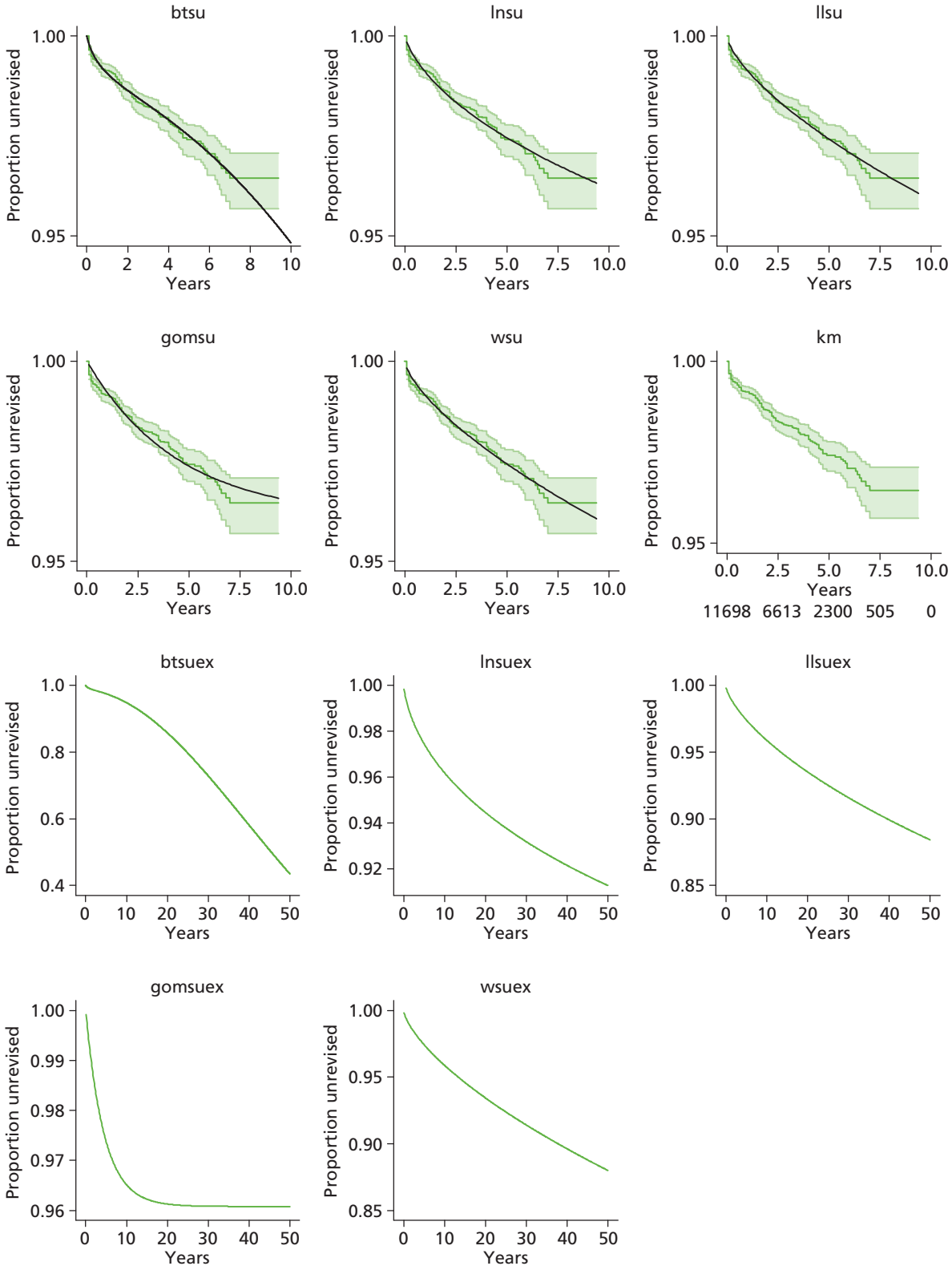
Category E [ceramic head (cemented stem) on cemented polyethylene cup]

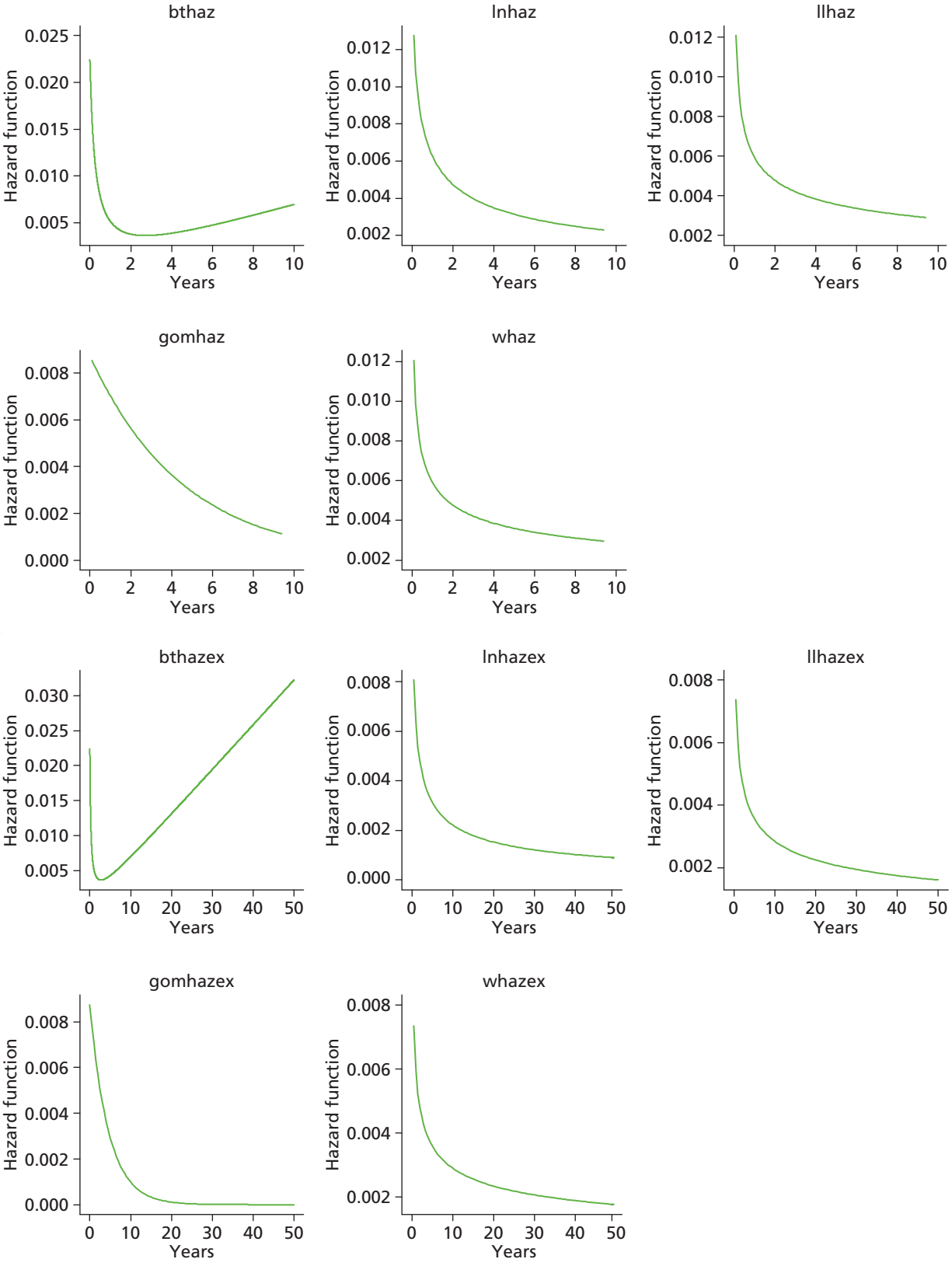




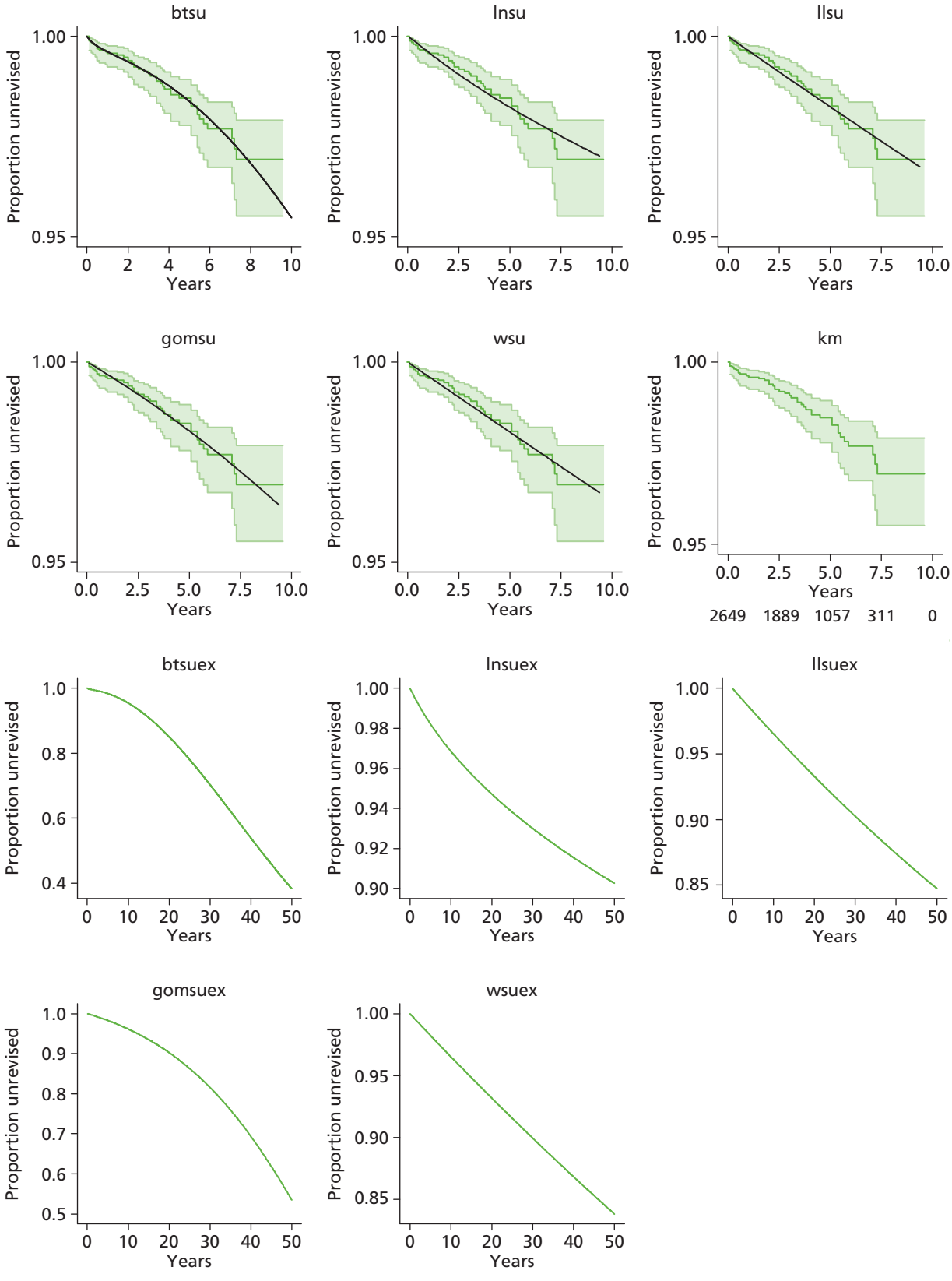
Total hip replacement categories: female aged < 65 years

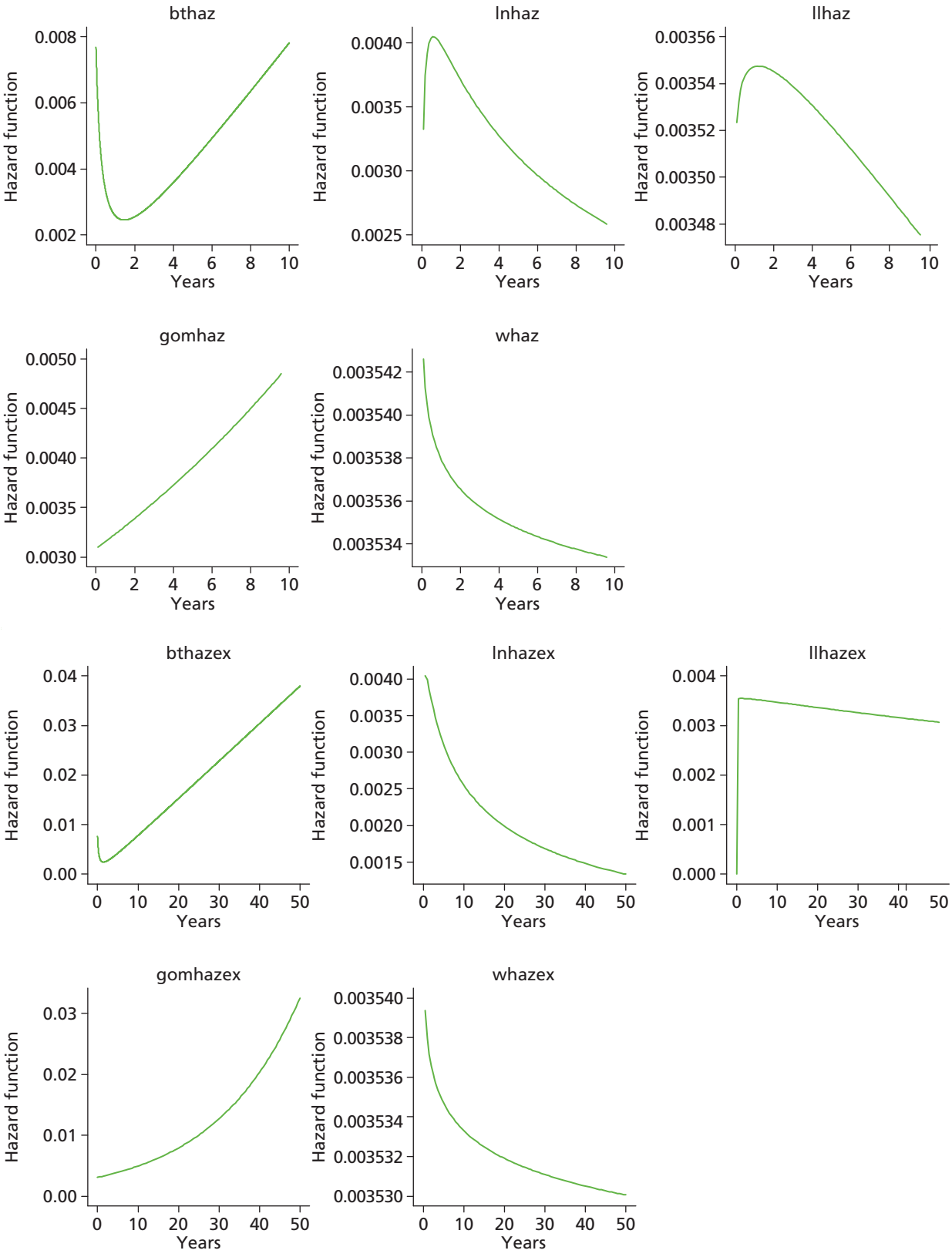
Category C [ceramic head (cementless stem) on cementless hydroxyapatite-coated metal cup (ceramic liner)]



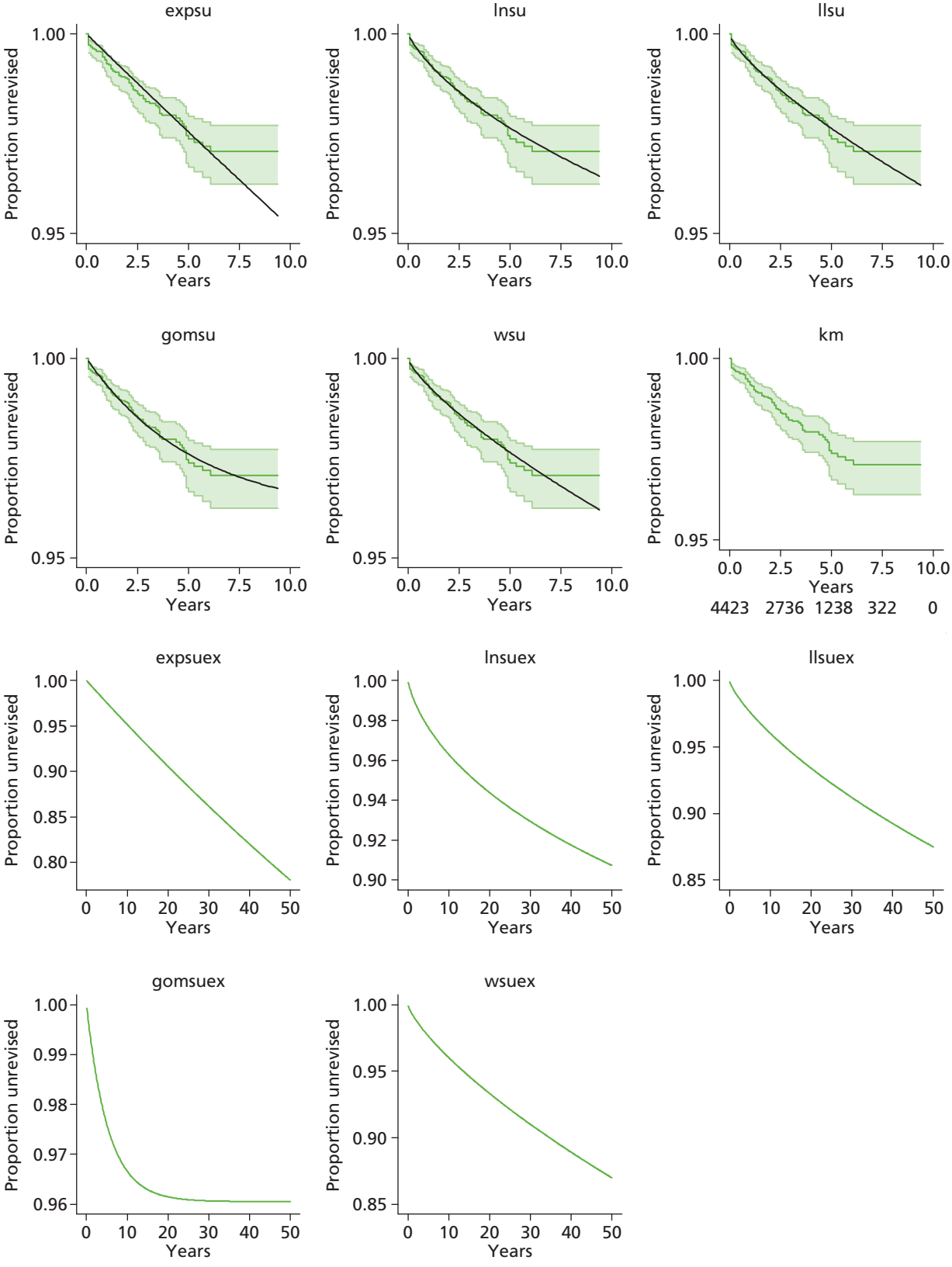


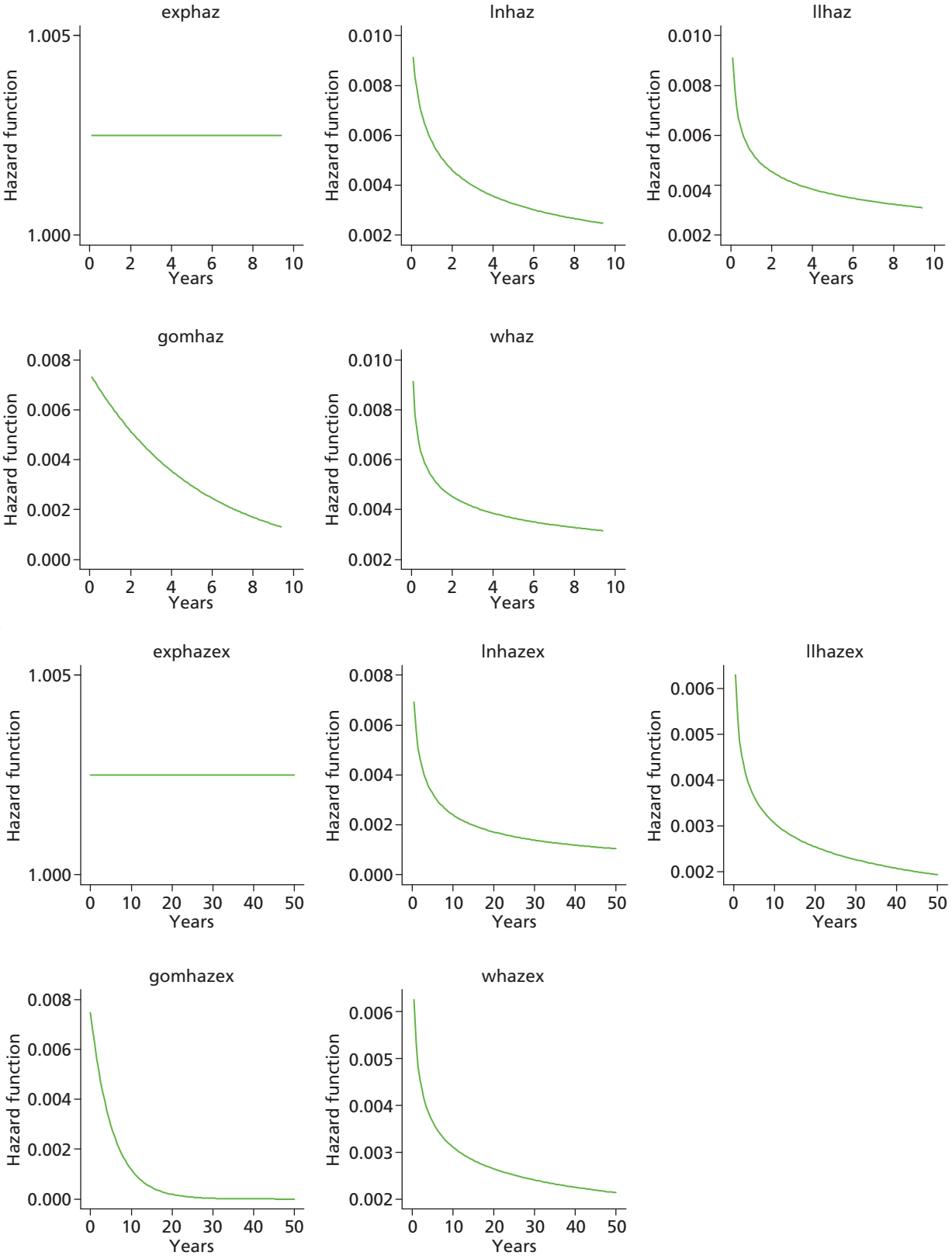
Category D [hybrid metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



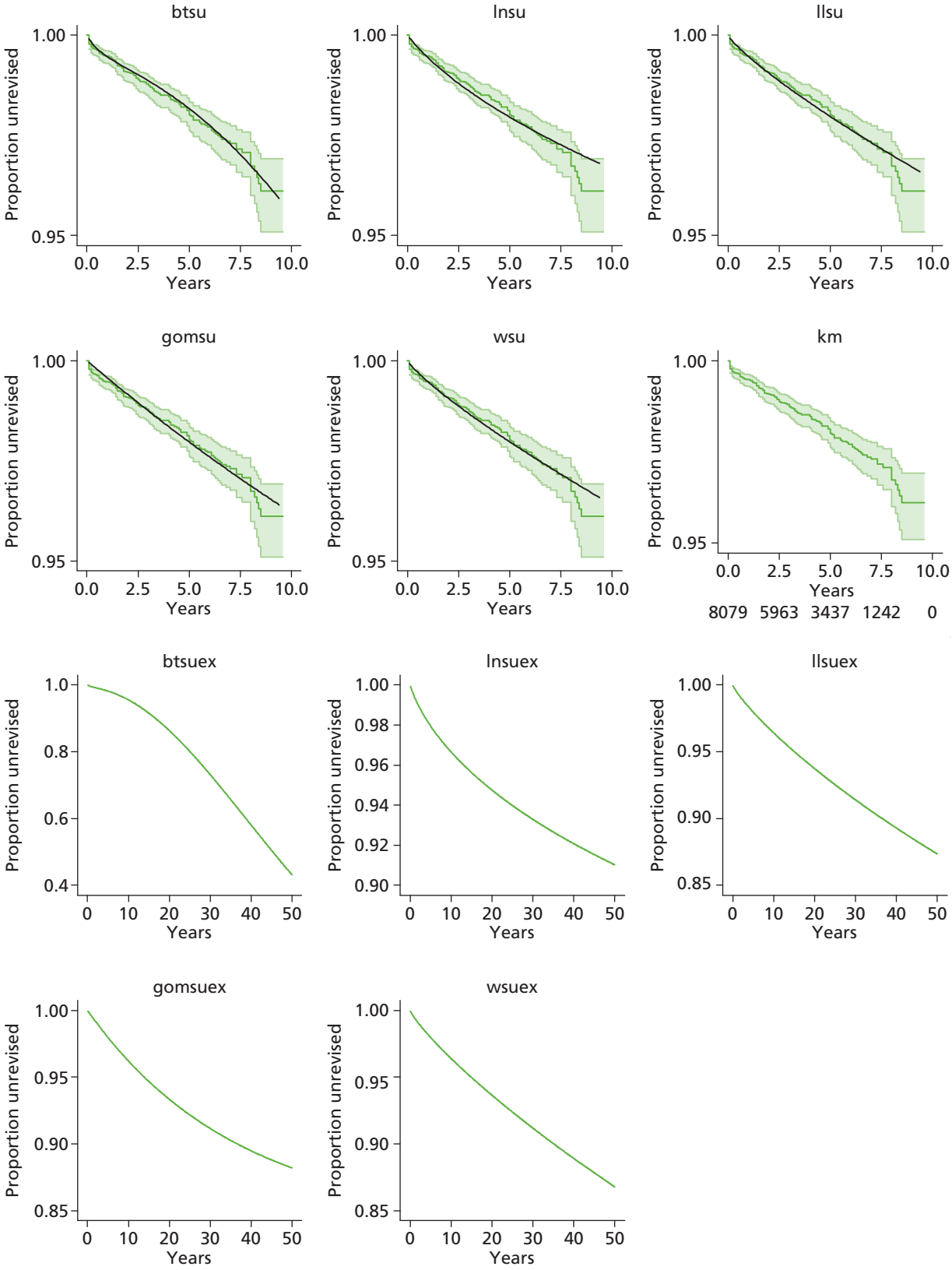


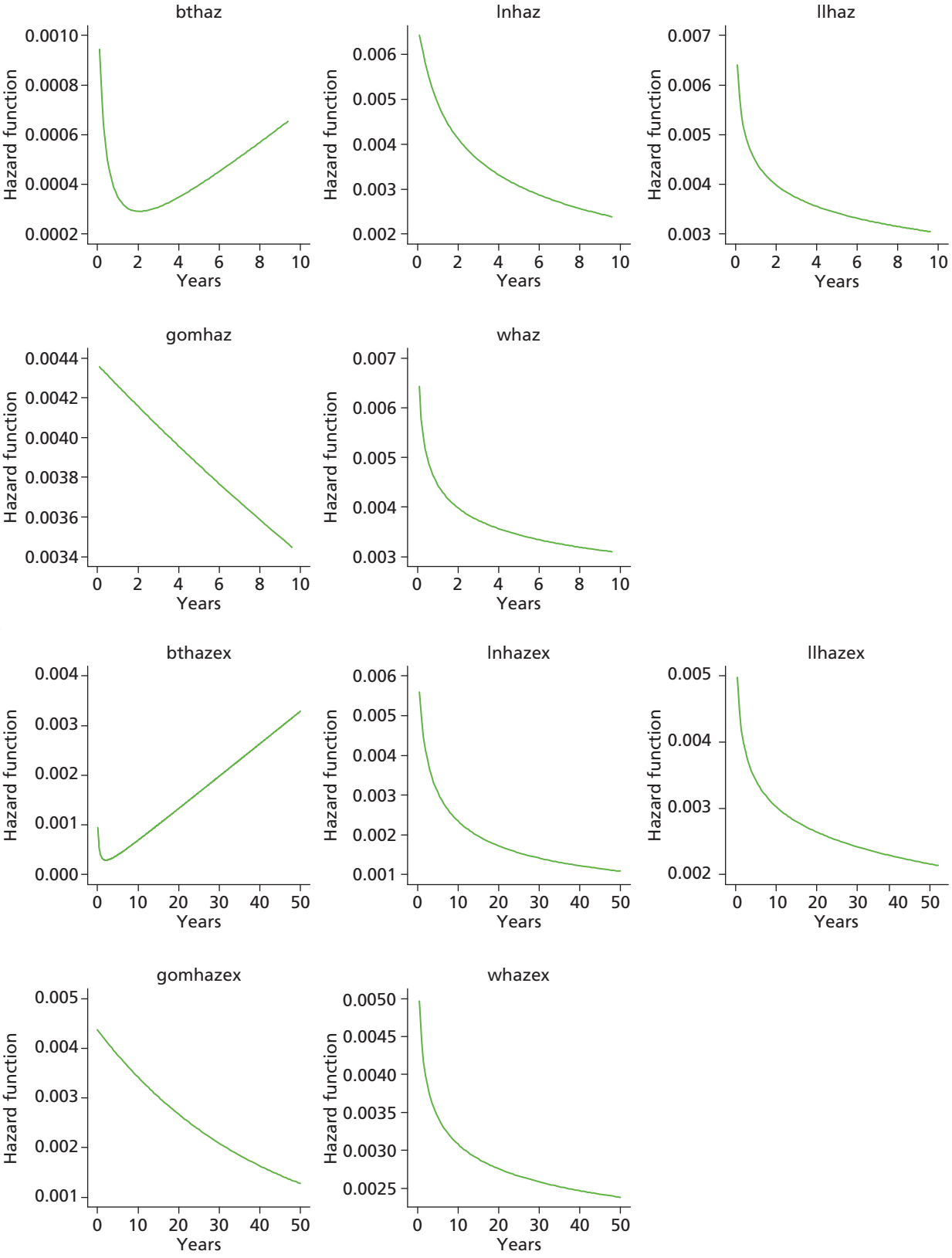
Category B [metal head (cemented stem) on cementless hydroxyapatite-coated metal cup (polyethylene liner)]



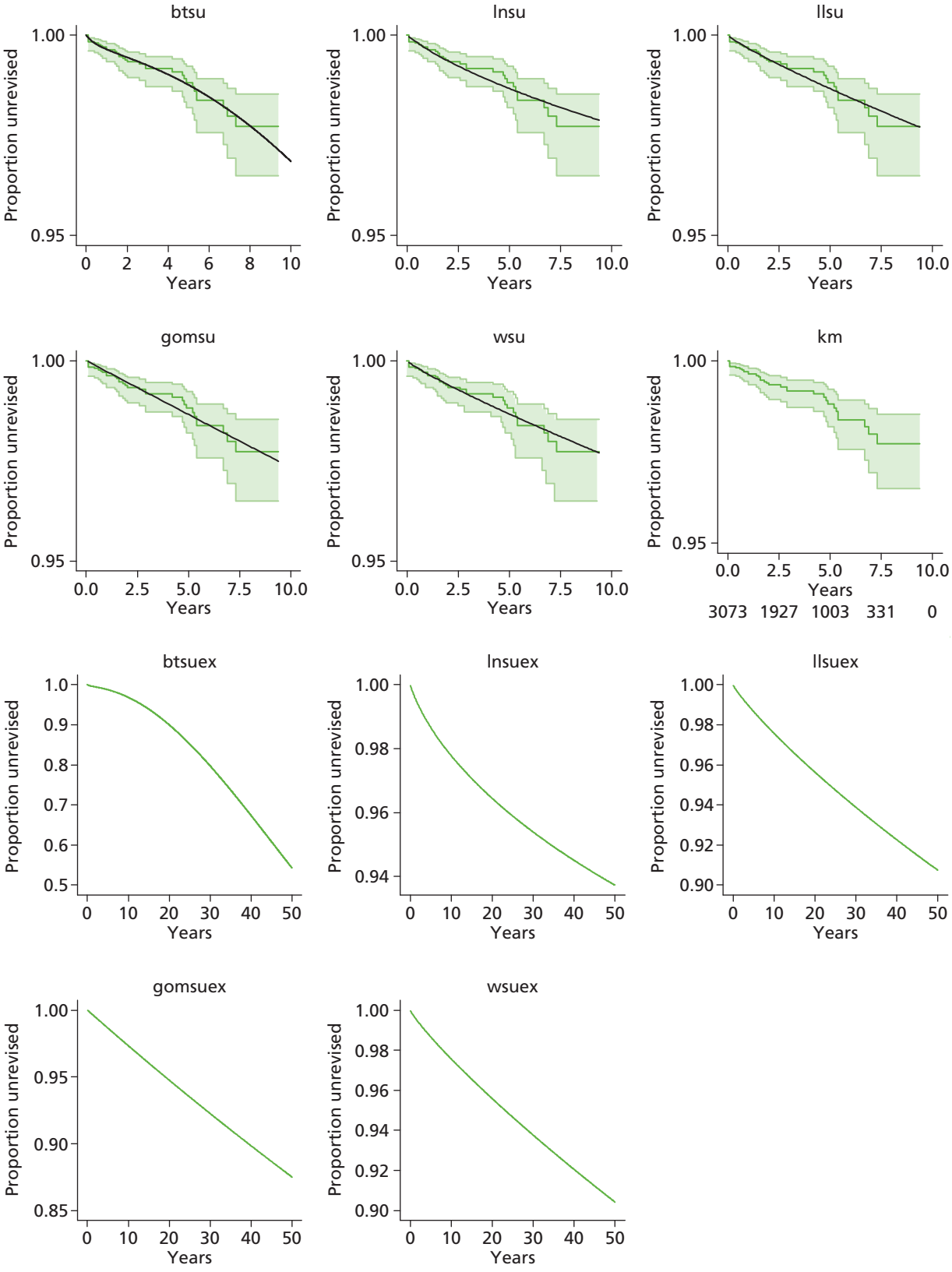


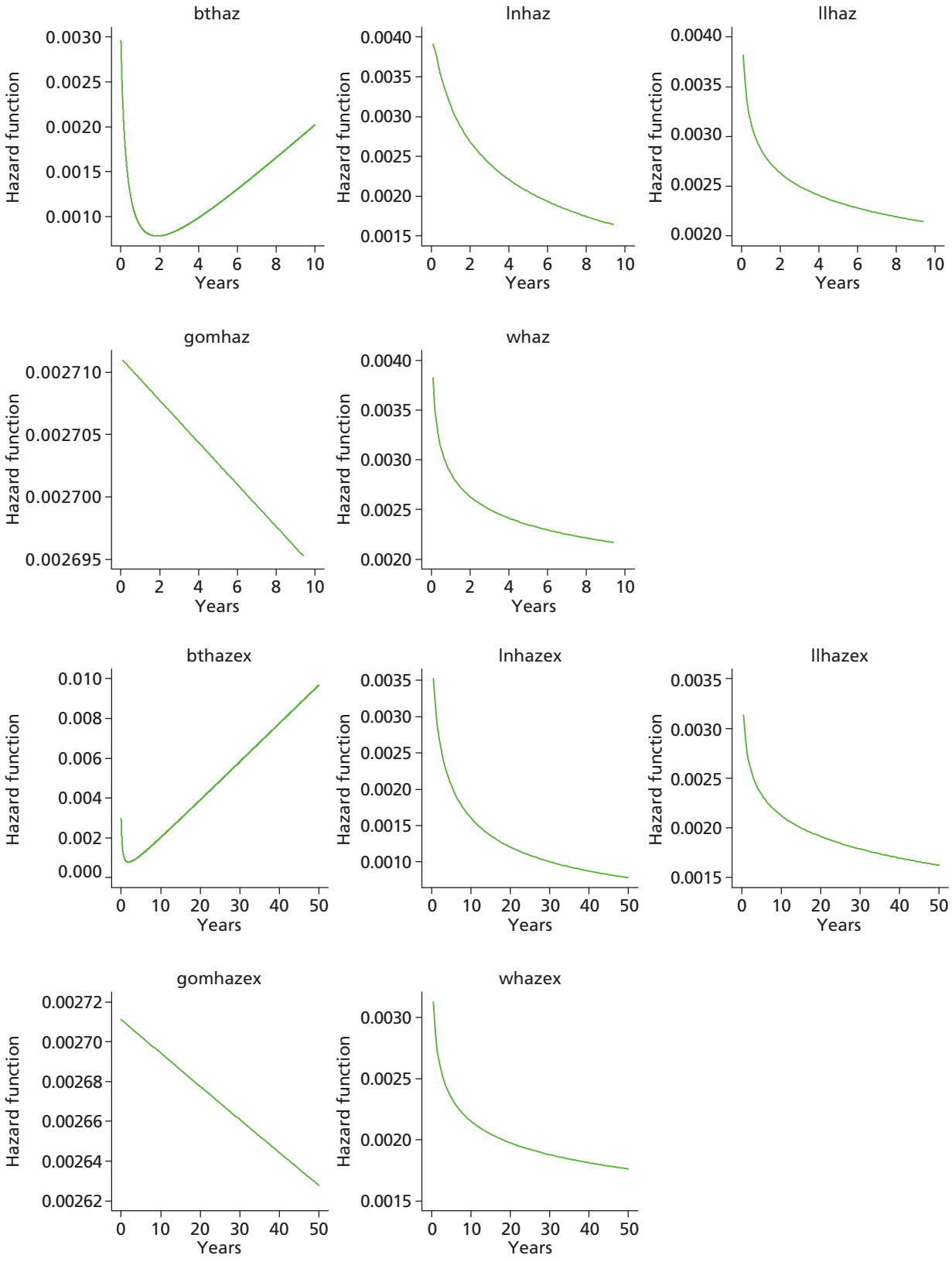
Category A [metal head (cemented stem) on cemented polyethylene cup]





Category E [ceramic head (cemented stem) on cemented polyethylene cup]





Appendix 18 Results from competing risk and Kaplan–Meier analyses of revision rates for patients receiving total hip replacement and resurfacing arthroplasty interventions

Figure 75 summarises competing risk and Kaplan–Meier analyses of revision rates for patients receiving THR or RS; the results are stratified by sex and are presented on a reversed scale for ease of comparison with NJR results published by Smith *et al.*[15] It is clear that revision rates after RS are much higher for women than for men. For RS the rates estimated from Kaplan–Meier and competing risk analyses are similar for both sexes; this is not surprising because few deaths would be likely to occur in this relatively young population during the relatively short follow-up period. For THR the Kaplan–Meier analysis generates somewhat higher rates of revision than the competing risk analysis; again, this applies for both sexes. For this older group of patients a greater proportion of deaths occur during follow-up relative to younger RS patients.

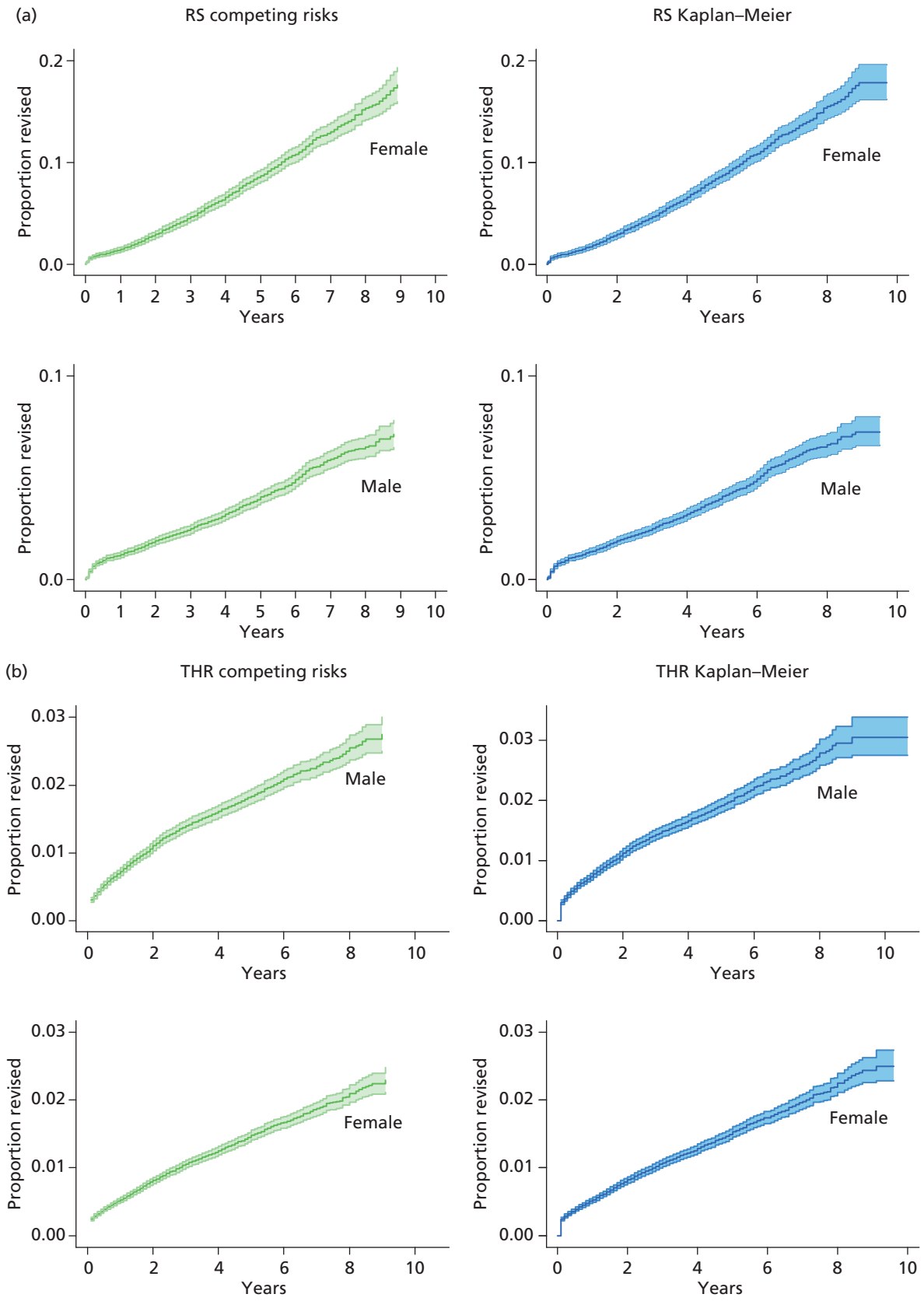


FIGURE 75 Competing risk and Kaplan–Meier analyses of revision rates after (a) RS; and (b) THR.

Appendix 19 Information criteria scores for parametric models

Resurfacing arthroplasty compared with total hip replacement

Matching status	Intervention and sex	Model	Observations	Likelihood model	Parameters	AIC	BIC
	RS female	Exponential	9321	–3832.57	2	7669.146	7683.426
	RS female	Weibull	9321	–3822.48	3	7650.955	7672.375
	RS female	Gompertz	9321	–3810.47	3	7626.936	7648.357
	RS female	Log-normal	9321	–3871.01	3	7748.02	7769.44
	RS female	Log-logistic	9321	–3826.6	3	7659.199	7680.619
	RS female	Bathtub	9321	–3804.81	4	7617.618	7646.178
Matched	THR female	Exponential	9321	–1175.68	2	2355.366	2369.646
Matched	THR female	Weibull	9321	–1171.86	3	2349.722	2371.142
Matched	THR female	Gompertz	9321	–1174.19	3	2354.382	2375.802
Matched	THR female	Log-normal	9321	–1172.61	3	2351.221	2372.641
Matched	THR female	Log-logistic	9321	–1171.9	3	2349.793	2371.213
Matched	THR female	Bathtub	9321	–1171.07	4	2350.132	2378.692
	RS male	Exponential	17,322	–4156.28	2	8316.549	8332.069
	RS male	Weibull	17,322	–4136.86	3	8279.725	8303.004
	RS male	Gompertz	17,322	–4153.83	3	8313.657	8336.936
	RS male	Log-normal	17,322	–4154.04	3	8314.077	8337.356
	RS male	Log-logistic	17,322	–4138.8	3	8283.599	8306.879
	RS male	Bathtub	17,322	–4103.65	4	8215.306	8246.345
Matched	THR male	Exponential	17,322	–2051.92	2	4107.83	4123.349
Matched	THR male	Weibull	17,322	–2038.25	3	4082.5	4105.779
Matched	THR male	Gompertz	17,322	–2046.58	3	4099.169	4122.448
Matched	THR male	Log-normal	17,322	–2040.2	3	4086.392	4109.671
Matched	THR male	Log-logistic	17,322	–2038.37	3	4082.738	4106.017
Matched	THR male	Bathtub	17,322	–2033.94	4	4075.889	4106.928

Total hip replacement categories (A–E): men aged > 65 years

THR category	Model	Observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	6186	–746.636	2	1497.273	1510.733
CeLCoC (category C)	Weibull	6186	–729.575	3	1465.15	1485.34
CeLCoC (category C)	Gompertz	6186	–732.489	3	1470.978	1491.168
CeLCoC (category C)	Log-normal	6186	–726.414	3	1458.828	1479.018
CeLCoC (category C)	Log-logistic	6186	–729.471	3	1464.942	1485.132
CeLCoC (category C)	Bathtub	6186	–723.201	4	1454.401	1481.321
HyMoP (category D)	Exponential	8657	–759.608	2	1523.211	1537.343
HyMoP (category D)	Weibull	8657	–747.974	3	1501.79	1522.989
HyMoP (category D)	Gompertz	8657	–750.925	3	1507.62	1528.818
HyMoP (category D)	Log-normal	8657	–746.515	3	1498.783	1519.982
HyMoP (category D)	Log-logistic	8657	–747.947	3	1501.732	1522.93
HyMoP (category D)	Bathtub	8657	–745.745	4	1499.49	1527.754
CeLMoP (category B)	Exponential	11,878	–1509.99	2	3017.336	3032.101
CeLMoP (category B)	Weibull	11,878	–1457.43	3	2916.575	2938.723
CeLMoP (category B)	Gompertz	11,878	–1466.86	3	2936.071	2958.218
CeLMoP (category B)	Log-normal	11,878	–1451.03	3	2902.553	2924.701
CeLMoP (category B)	Log-logistic	11,878	–1457.23	3	2916.099	2938.247
CeLMoP (category B)	Bathtub	11,878	–1438.98	4	2885.952	2915.482
CeMoP (category A)	Exponential	37,018	–3243.77	2	6491.542	6508.58
CeMoP (category A)	Weibull	37,018	–3196.01	3	6398.019	6423.577
CeMoP (category A)	Gompertz	37,018	–3212.72	3	6431.431	6456.989
CeMoP (category A)	Log-normal	37,018	–3190.82	3	6387.648	6413.205
CeMoP (category A)	Log-logistic	37,018	–3195.95	3	6397.894	6423.452
CeMoP (category A)	Bathtub	37,018	–3182.69	4	6373.389	6407.465
CeCoP (category E)	Exponential	2777	–193.422	2	390.0204	401.8786
CeCoP (category E)	Weibull	2777	–191.253	3	387.7762	405.5636
CeCoP (category E)	Gompertz	2777	–190.6	3	386.5367	404.3241
CeCoP (category E)	Log-normal	2777	–190.787	3	386.8841	404.6715
CeCoP (category E)	Log-logistic	2777	–191.24	3	387.7485	405.5359
CeCoP (category E)	Bathtub	2777	Not reported			

Total hip replacement categories (A–E): women aged > 65 years

THR category	Model	Observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	7554	–708.1475	2	1420.295	1434.155
CeLCoC (category C)	Weibull	7554	–685.9571	3	1377.914	1398.704
CeLCoC (category C)	Gompertz	7554	–692.052	3	1390.104	1410.893
CeLCoC (category C)	Log-normal	7554	–683.4155	3	1372.831	1393.62
CeLCoC (category C)	Log-logistic	7554	–685.9026	3	1377.805	1398.595
CeLCoC (category C)	Bathtub	7554	–678.7391	4	1365.478	1393.197
HyMoP (category D)	Exponential	15,641	–1200.145	2	2404.291	2419.606
HyMoP (category D)	Weibull	15,641	–1167.915	3	2341.831	2364.804
HyMoP (category D)	Gompertz	15,641	–1178.968	3	2363.936	2386.909
HyMoP (category D)	Log-normal	15,641	–1165.254	3	2336.508	2359.481
HyMoP (category D)	Log-logistic	15,641	–1167.881	3	2341.762	2364.735
HyMoP (category D)	Bathtub	15,641	–1158.393	4	2324.787	2355.417
CeLMoP (category B)	Exponential	18,396	–2076.223	2	4156.445	4172.085
CeLMoP (category B)	Weibull	18,396	–1983.403	3	3972.806	3996.265
CeLMoP (category B)	Gompertz	18,396	–2016.981	3	4039.962	4063.421
CeLMoP (category B)	Log-normal	18,396	–1975.002	3	3956.004	3979.464
CeLMoP (category B)	Log-logistic	18,396	–1983.209	3	3972.418	3995.878
CeLMoP (category B)	Bathtub	18,396	–1949.76	4	3907.519	3938.799
CeMoP (category A)	Exponential	75,734	–5258.73	2	10,521.46	10,539.93
CeMoP (category A)	Weibull	75,734	–5231.8	3	10,469.59	10,497.3
CeMoP (category A)	Gompertz	75,734	–5243.93	3	10,493.85	10,521.56
CeMoP (category A)	Log-normal	75,734	–5233.88	3	10,473.75	10,501.46
CeMoP (category A)	Log-logistic	75,734	–5231.91	3	10,469.82	10,497.52
CeMoP (category A)	Bathtub	75,734	–5228.99	4	10,465.98	10,502.92
CeCoP (category E)	Exponential	4655	–231.0568	2	466.1135	479.0049
CeCoP (category E)	Weibull	4655	–230.1712	3	466.3423	485.6794
CeCoP (category E)	Gompertz	4655	–229.2509	3	464.5019	483.839
CeCoP (category E)	Log-normal	4655	–229.665	3	465.3301	484.6672
CeCoP (category E)	Log-logistic	4655	–230.1605	3	466.321	485.6581
CeCoP (category E)	Bathtub	Not reported				

Total hip replacement categories (A–E): men aged < 65 years

THR category	Model	Observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	9316	–1127.92	2	2259.84	2274.119
CeLCoC (category C)	Weibull	9316	–1112.134	3	2230.269	2251.687
CeLCoC (category C)	Gompertz	9316	–1117.294	3	2240.588	2262.007
CeLCoC (category C)	Log-normal	9316	–1110.068	3	2226.137	2247.555
CeLCoC (category C)	Log-logistic	9316	–1112.095	3	2230.189	2251.608
CeLCoC (category C)	Bathtub	9316	–1108.026	4	2224.052	2252.61
HyMoP (category D)	Exponential	1524	–219.2187	2	442.4374	453.0956
HyMoP (category D)	Weibull	1524	–215.9177	3	437.8354	453.8226
HyMoP (category D)	Gompertz	1524	–217.6033	3	441.2066	457.1939
HyMoP (category D)	Log-normal	1524	–215.6226	3	437.2452	453.2325
HyMoP (category D)	Log-logistic	1524	–215.9186	3	437.8372	453.8245
HyMoP (category D)	Bathtub	1524	–214.6095	4	437.2189	458.5353
CeLMoP (category B)	Exponential	3177	–447.9051	2	899.8102	911.9376
CeLMoP (category B)	Weibull	3177	–444.5135	3	895.027	913.218
CeLMoP (category B)	Gompertz	3177	–446.802	3	899.6039	917.795
CeLMoP (category B)	Log-normal	3177	–445.2237	3	896.4473	914.6384
CeLMoP (category B)	Log-logistic	3177	–444.5621	3	895.1241	913.3152
CeLMoP (category B)	Bathtub	3177	–443.3329	4	894.6658	918.9206
CeMoP (category A)	Exponential	4454	–552.23	2	1108.459	1121.262
CeMoP (category A)	Weibull	4454	–550.158	3	1106.316	1125.521
CeMoP (category A)	Gompertz	4454	–551.318	3	1108.635	1127.84
CeMoP (category A)	Log-normal	4454	–550.825	3	1107.65	1126.855
CeMoP (category A)	Log-logistic	4454	–550.198	3	1106.396	1125.601
CeMoP (category A)	Bathtub	4454	–549.76	4	1107.519	1133.125
CeCoP (category E)	Exponential	2200	–121.8768	2	247.7535	259.1459
CeCoP (category E)	Weibull	2200	–121.8192	3	249.6385	266.7271
CeCoP (category E)	Gompertz	2200	–121.5913	3	249.1827	266.2713
CeCoP (category E)	Log-normal	2200	–122.5208	3	251.0416	268.1303
CeCoP (category E)	Log-logistic	2200	–121.8351	3	249.6702	266.7589
CeCoP (category E)	Bathtub	2200	–121.2367	4	250.4733	273.2582

Total hip replacement categories (A–E): women aged < 65 years

THR category	Model	Observations	Model likelihood	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	11,698	–1366.3	2	2736.599	2751.334
CeLCoC (category C)	Weibull	11,698	–1345.799	3	2697.598	2719.7
CeLCoC (category C)	Gompertz	11,698	–1353.718	3	2713.435	2735.537
CeLCoC (category C)	Log-normal	11,698	–1343.318	3	2692.636	2714.738
CeLCoC (category C)	Log-logistic	11,698	–1345.762	3	2697.523	2719.625
CeLCoC (category C)	Bathtub	11,698	–1339.205	4	2686.41	2715.879
HyMoP (category D)	Exponential	2649	–240.9111	2	485.8222	497.586
HyMoP (category D)	Weibull	2649	–240.9111	3	487.8221	505.468
HyMoP (category D)	Gompertz	2649	–240.7166	3	487.4333	505.0791
HyMoP (category D)	Log-normal	2649	–241.8498	3	489.6996	507.3454
HyMoP (category D)	Log-logistic	2649	–240.9405	3	487.8809	505.5267
HyMoP (category D)	Bathtub	2649	–239.6651	4	487.3302	510.858
CeLMoP (category B)	Exponential	4423	–495.8744	2	995.7488	1008.538
CeLMoP (category B)	Weibull	4423	–492.0127	3	990.0253	1009.209
CeLMoP (category B)	Gompertz	4423	–491.7144	3	989.4287	1008.612
CeLMoP (category B)	Log-normal	4423	–491.0348	3	988.0695	1007.253
CeLMoP (category B)	Log-logistic	4423	–491.9636	3	989.9272	1009.111
CeLMoP (category B)	Bathtub	Not reported				
CeMoP (category A)	Exponential	8079	–905.022	2	1814.045	1828.039
CeMoP (category A)	Weibull	8079	–902.087	3	1810.173	1831.164
CeMoP (category A)	Gompertz	8079	–904.823	3	1815.647	1836.638
CeMoP (category A)	Log-normal	8079	–903.623	3	1813.246	1834.237
CeMoP (category A)	Log-logistic	8079	–902.157	3	1810.314	1831.305
CeMoP (category A)	Bathtub	8079	–897.288	4	1802.577	1830.565
CeCoP (category E)	Exponential	3073	–209.3335	2	422.6671	434.7279
CeCoP (category E)	Weibull	3073	–208.9521	3	423.9041	441.9954
CeCoP (category E)	Gompertz	3073	–209.3335	3	424.667	442.7582
CeCoP (category E)	Log-normal	3073	–209.5089	3	425.0178	443.1091
CeCoP (category E)	Log-logistic	3073	–208.9722	3	423.9445	442.0357
CeCoP (category E)	Bathtub	3073	–207.8105	4	423.6209	447.7426

Appendix 20 Plots of Kaplan–Meier-estimated cumulative hazards compared with modelled cumulative hazards

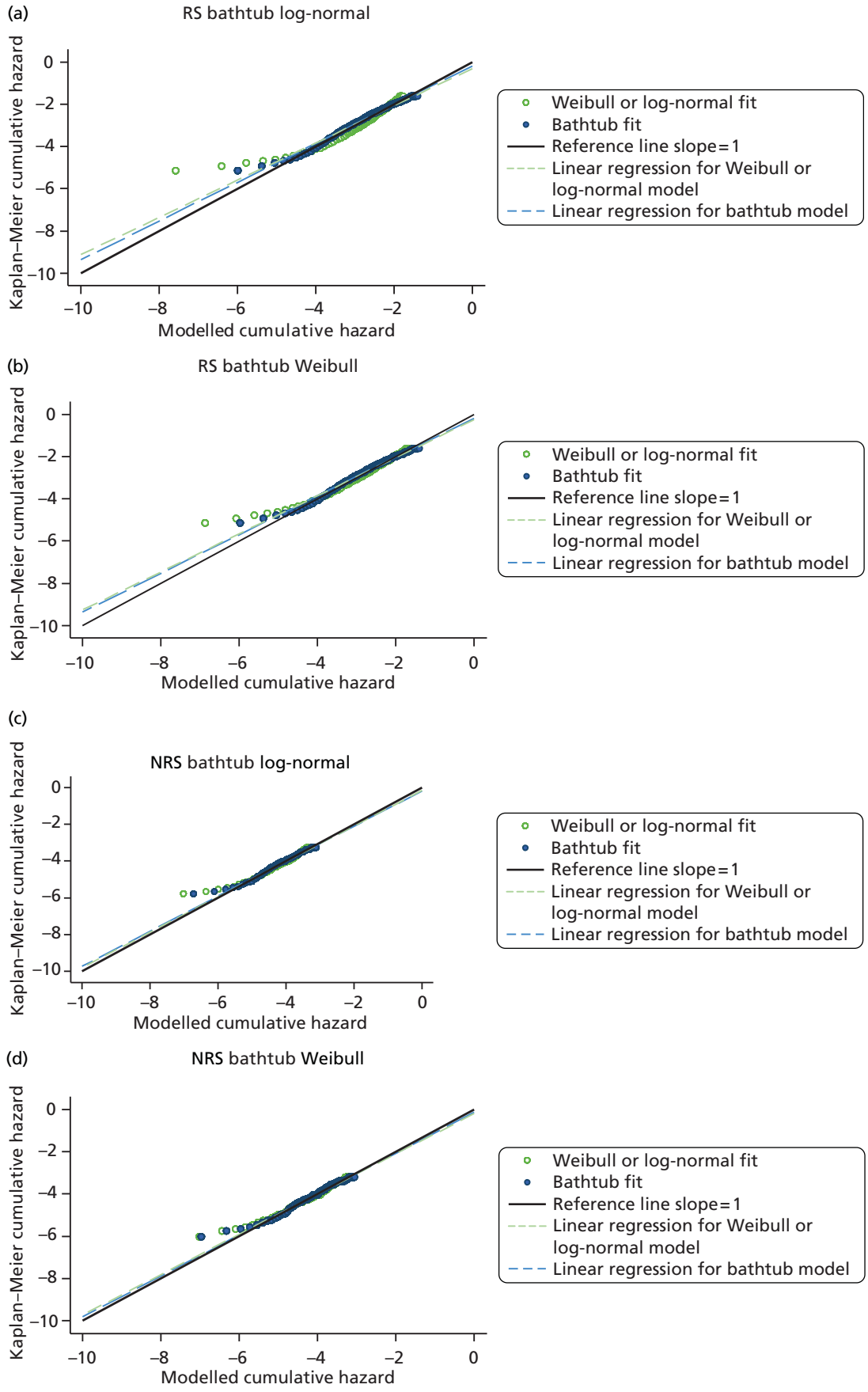


FIGURE 76 Female RS vs. non-RS (matched THR) cumulative hazard plots. (a) RS bt log-normal; (b) RS bt Weibull; (c) NRS bt log-normal; and (d) NRS bt Weibull.

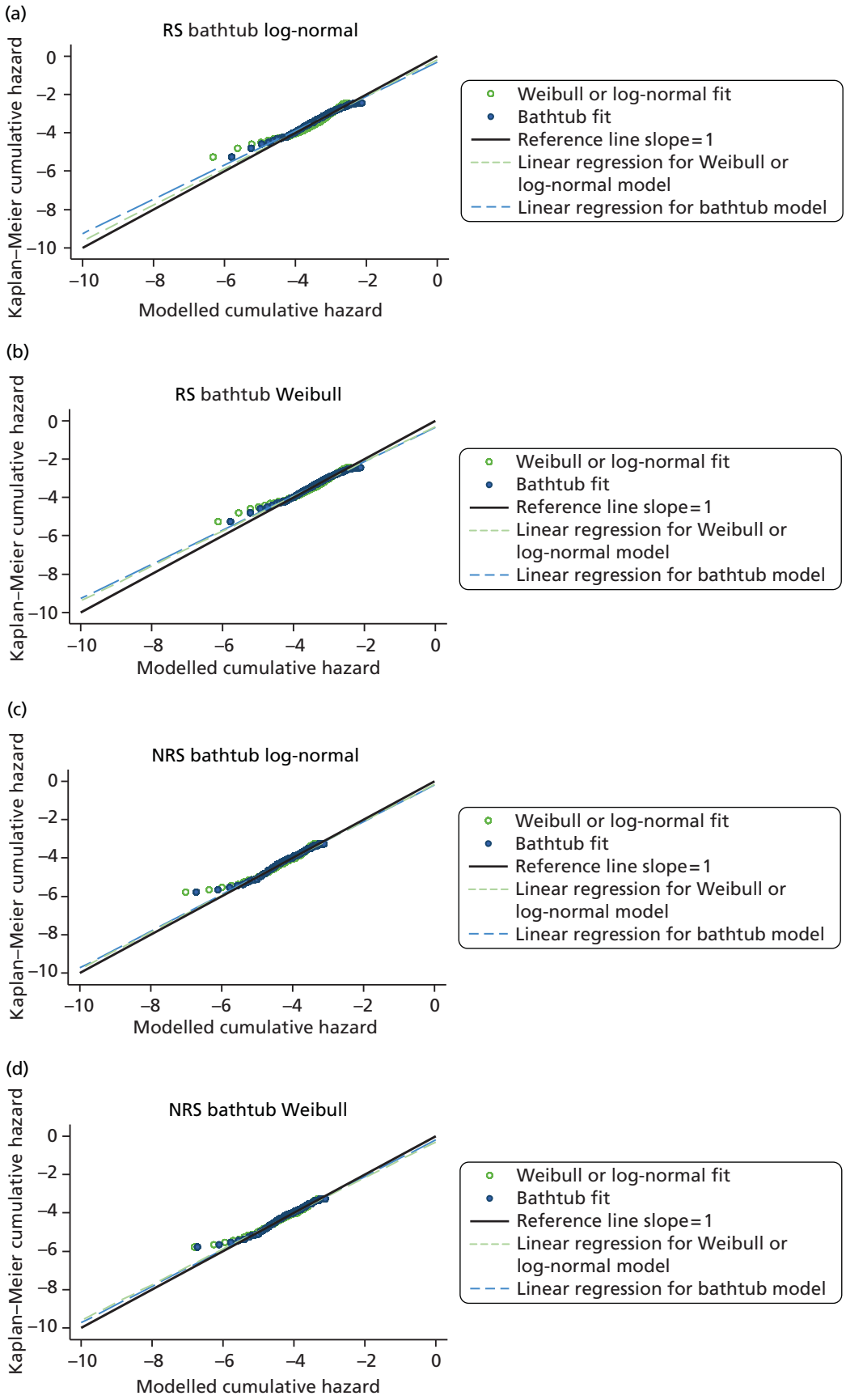


FIGURE 77 Male RS vs. non-RS (matched THR) cumulative hazard plots. (a) RS bt log-normal; (b) RS bt Weibull; (c) NRS bt log-normal; and (d) NRS bt Weibull.

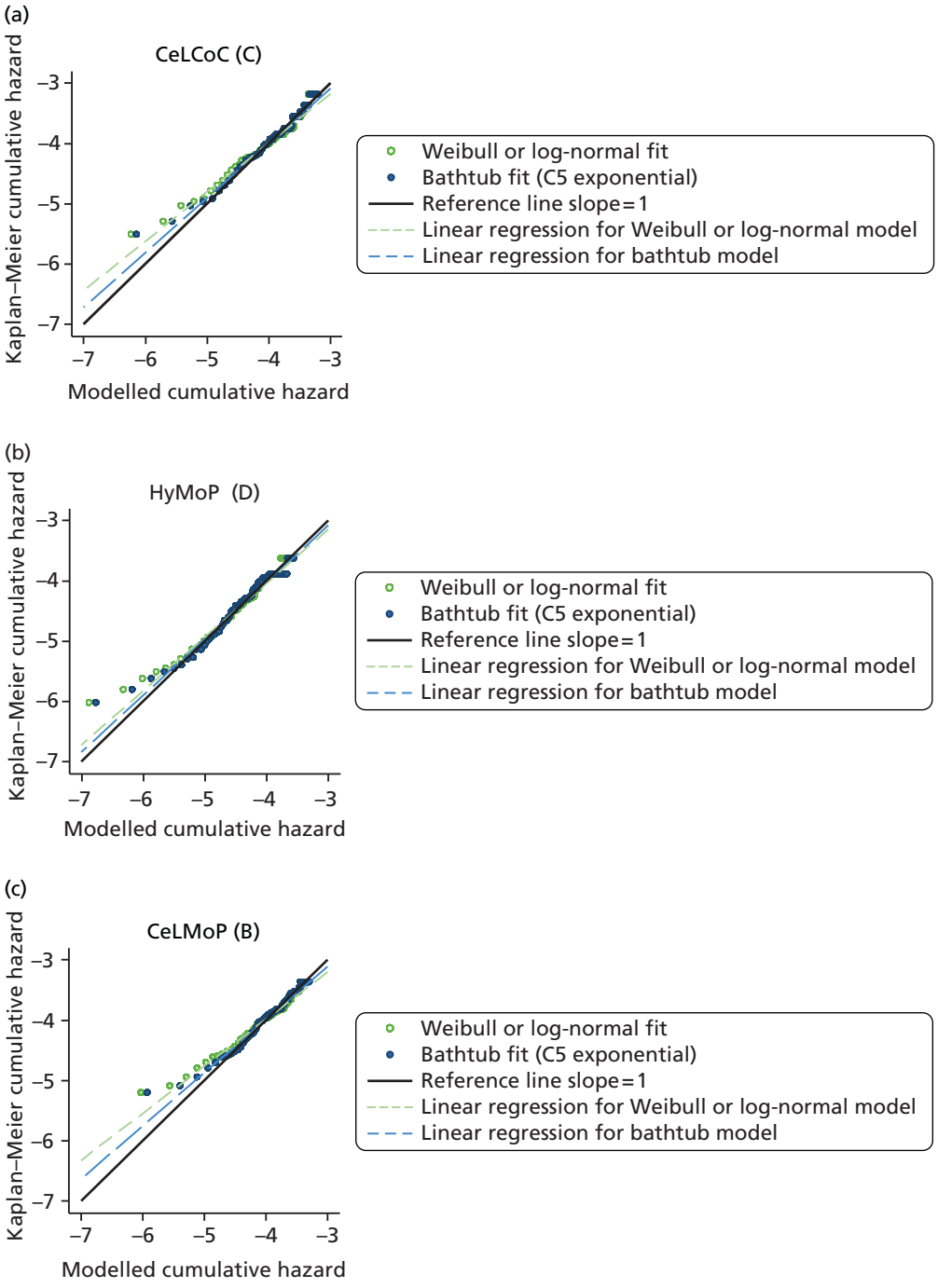


FIGURE 78 Cumulative hazard plots for men aged > 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E). (*continued*)

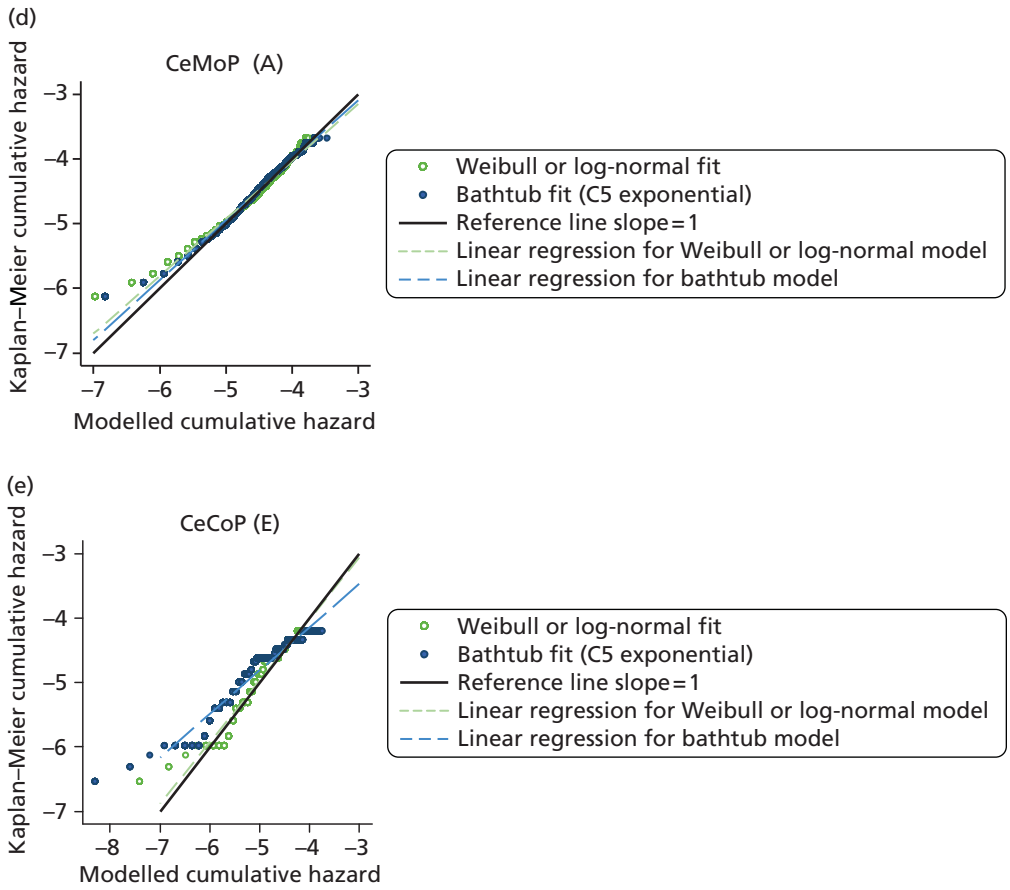


FIGURE 78 Cumulative hazard plots for men aged > 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E).

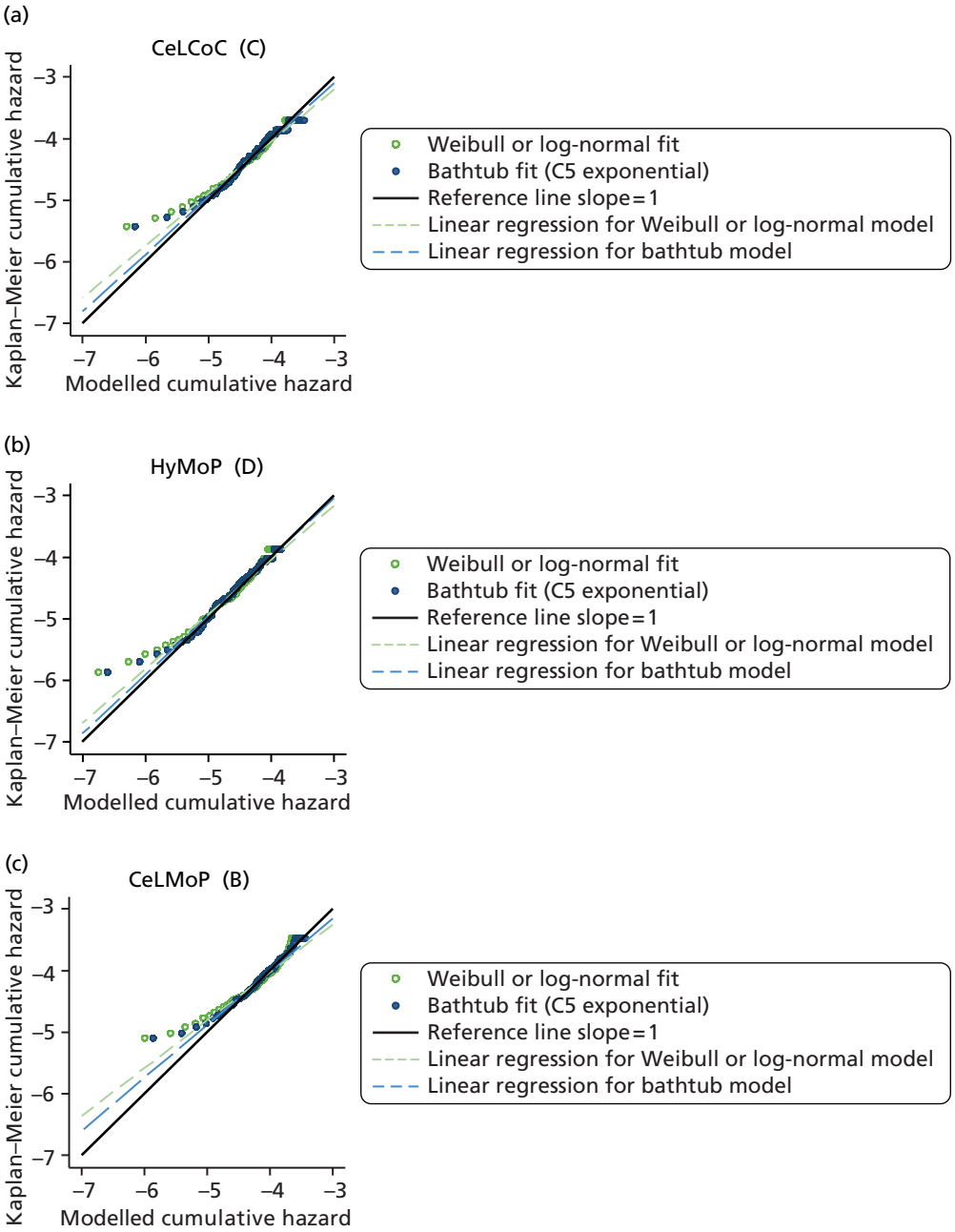


FIGURE 79 Cumulative hazard plots for women aged > 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E). (*continued*)

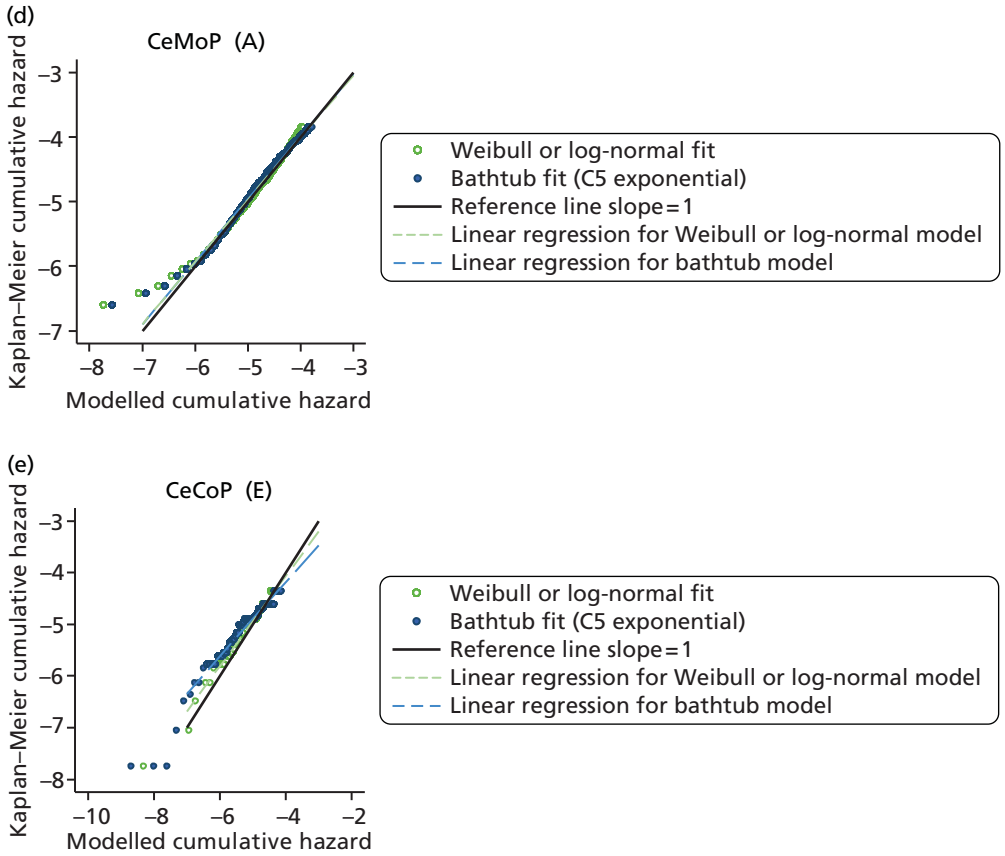


FIGURE 79 Cumulative hazard plots for women aged > 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E).

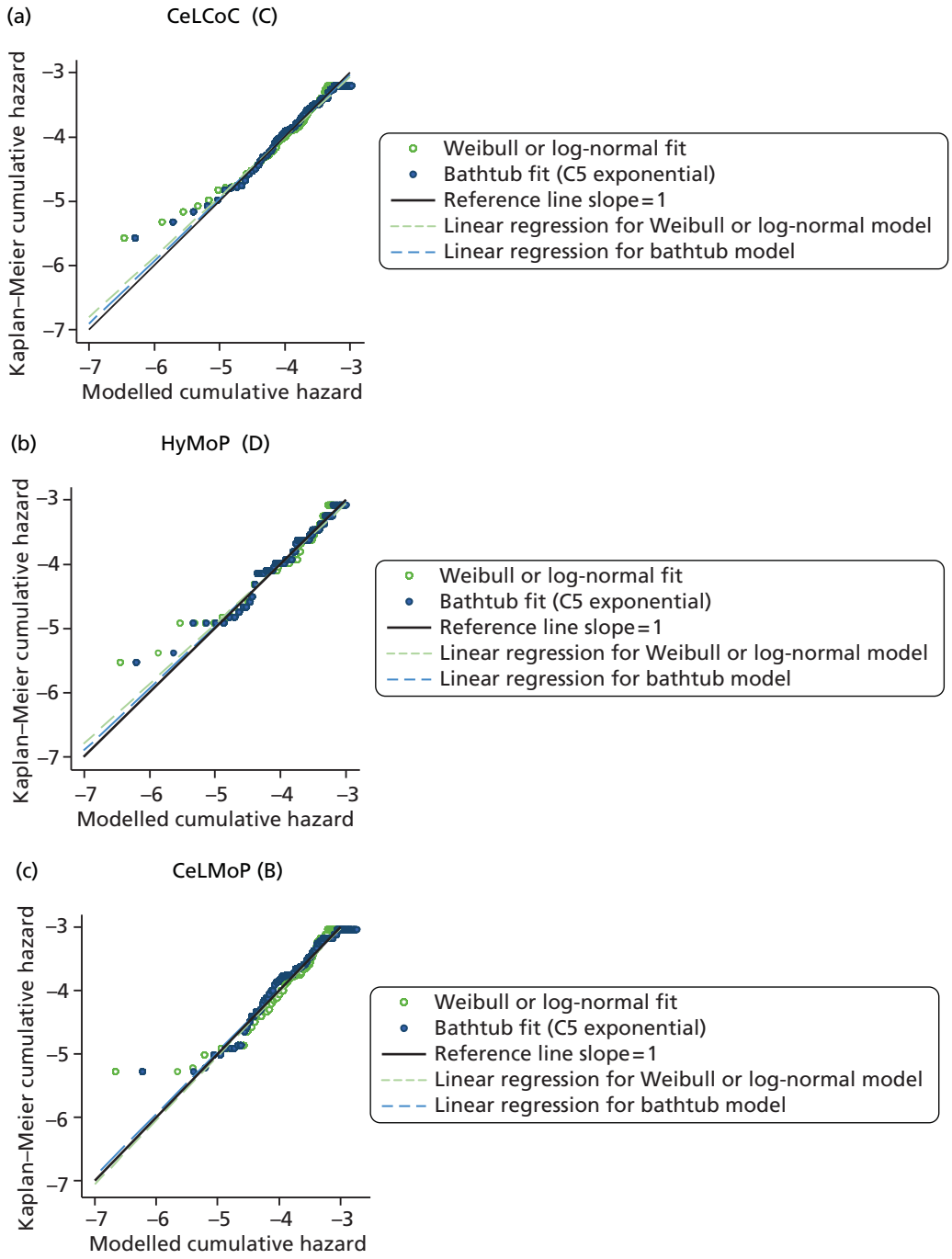


FIGURE 80 Cumulative hazard plots for men aged < 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E). (*continued*)

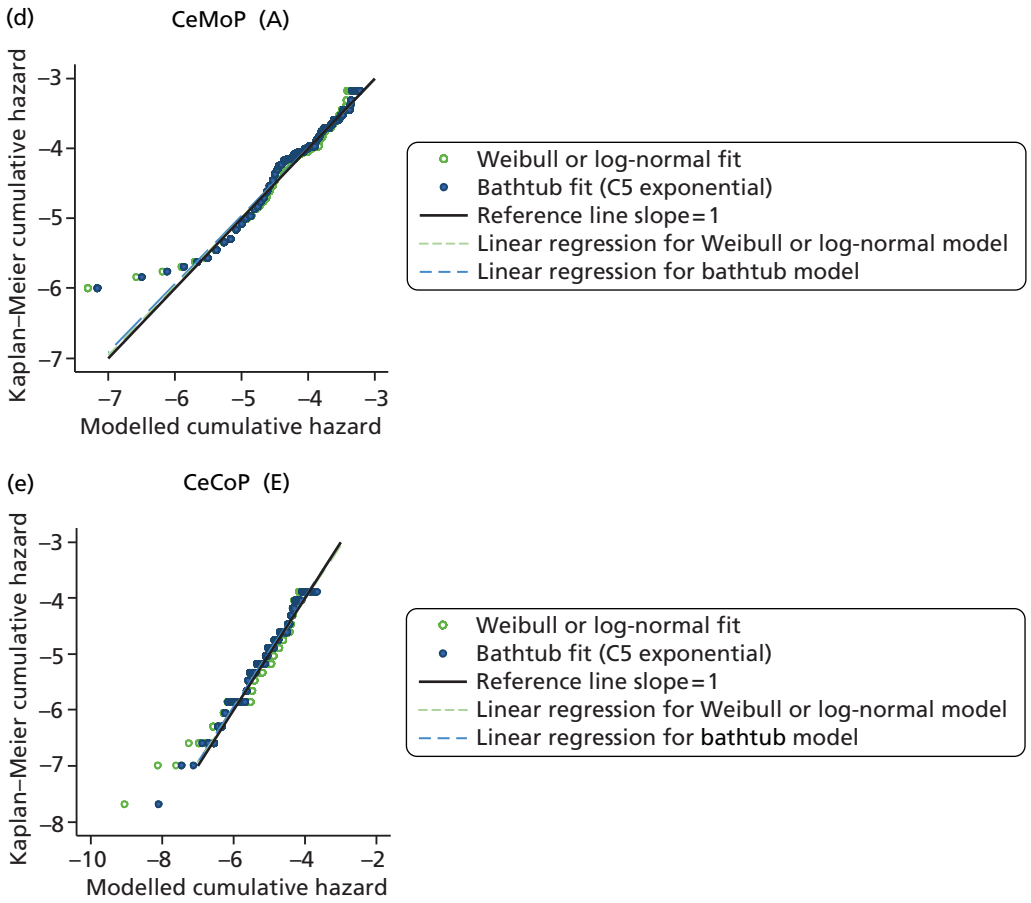


FIGURE 80 Cumulative hazard plots for men aged < 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E).

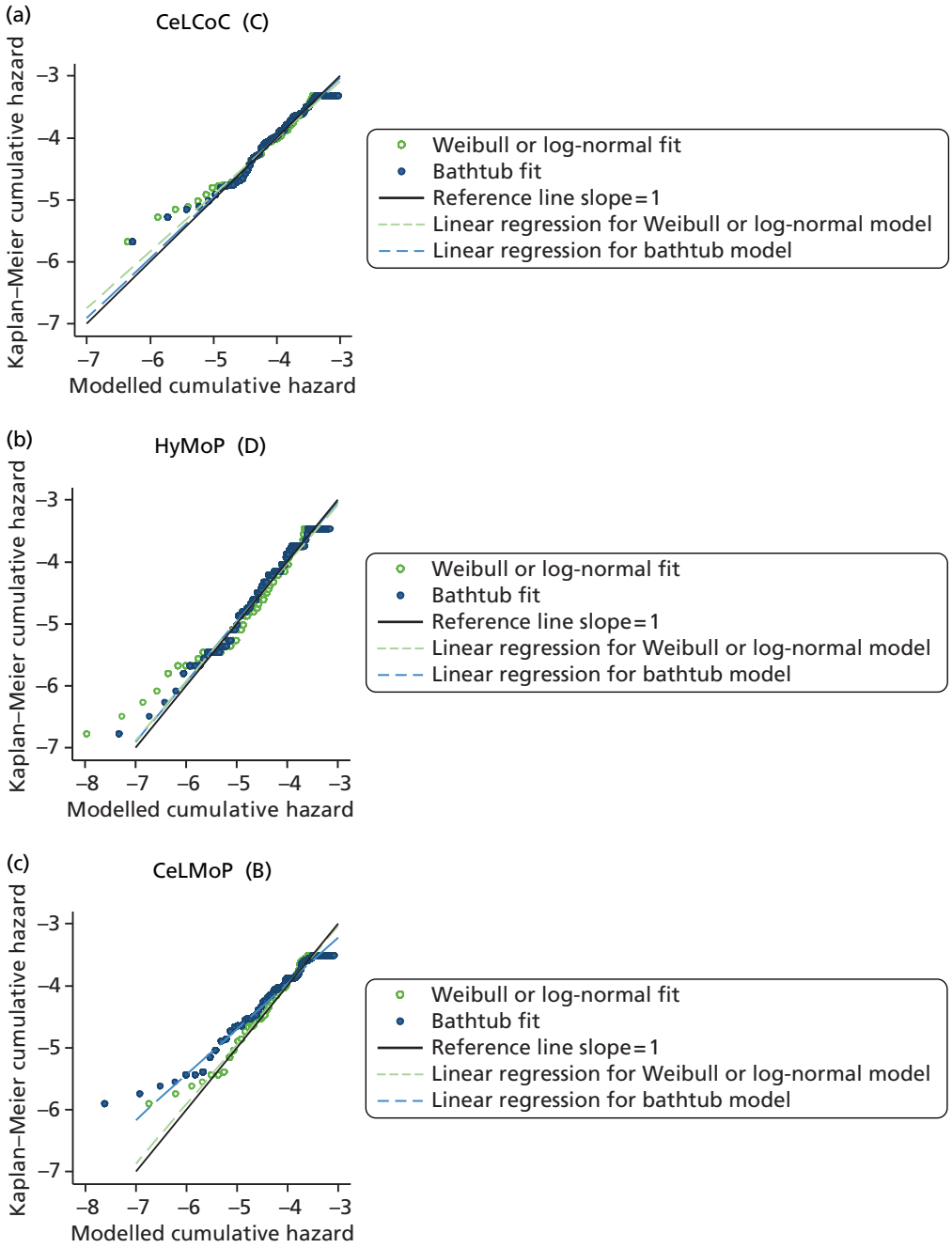


FIGURE 81 Cumulative hazard plots for women aged < 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E). (*continued*)

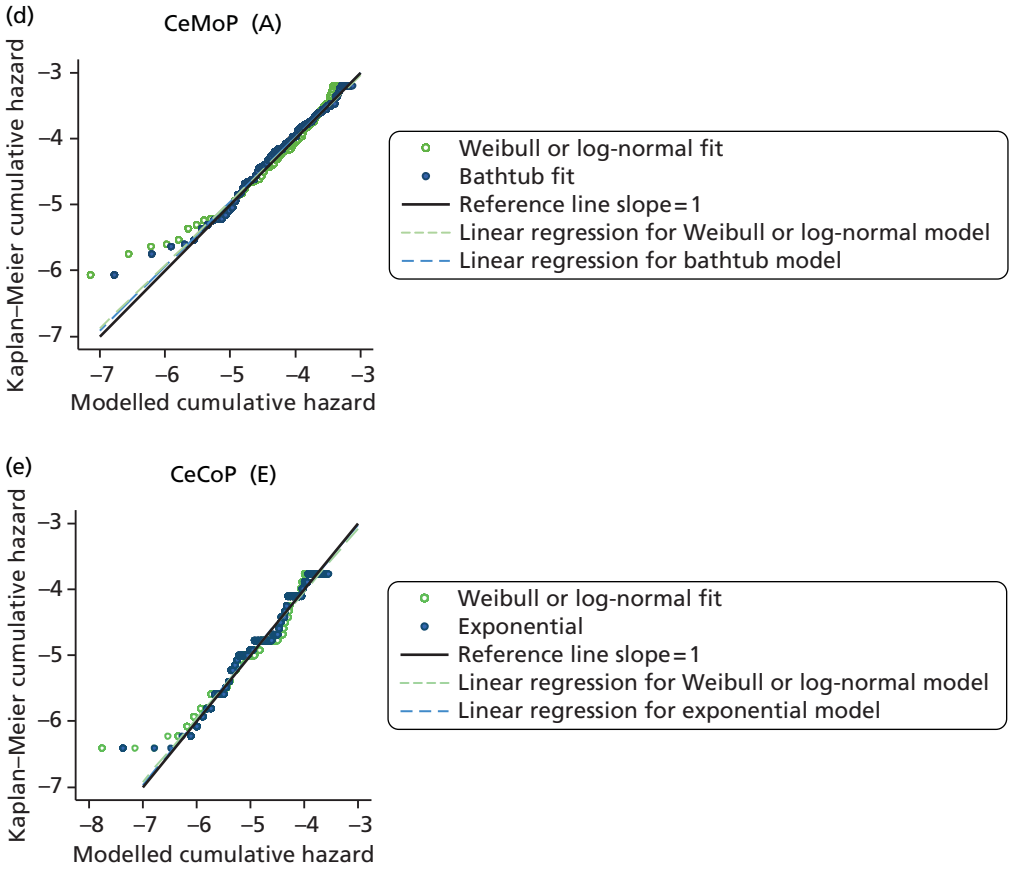


FIGURE 81 Cumulative hazard plots for women aged < 65 years. (a) CeLCoC (C); (b) HyMoP (D); (c) CeLMoP (B); (d) CeMoP (A); and (e) CeCoP (E).

Appendix 21 Comparison of resurfacing arthroplasty and total hip replacement according to sex

Women

Of 9339 female patients who received RS, 9321 were successfully matched by age with female THR patients. The identical age distributions in the two groups are shown in *Figure 82*.

Figure 83 shows that the observed revision rate is far higher for RS than for THR. Parametric fits are presented in *Appendix 18* and *Figure 84*.

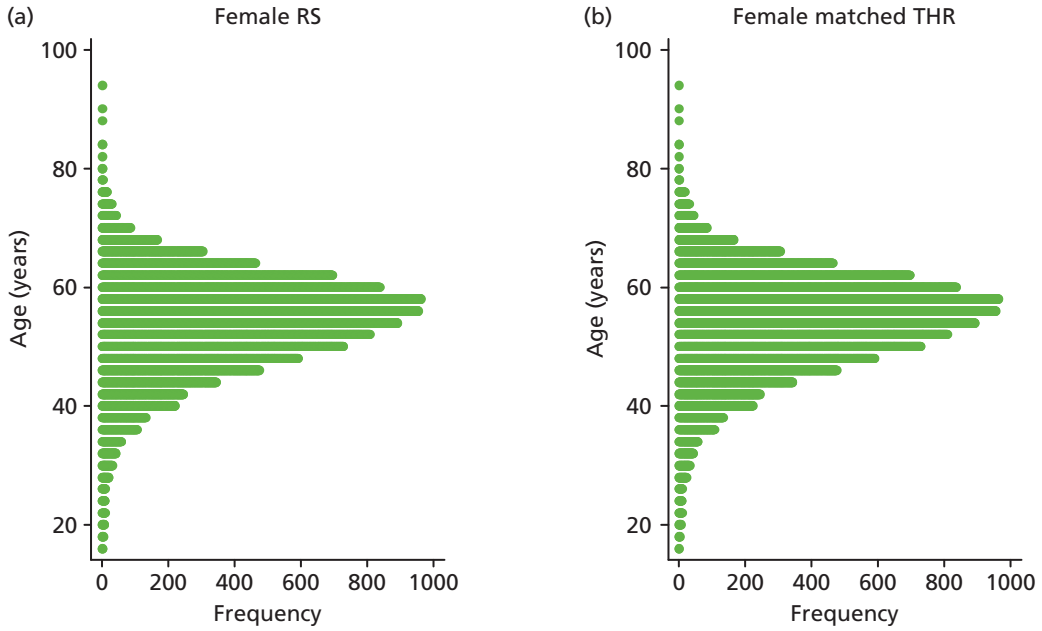


FIGURE 82 Age distribution in matched RS and THR female groups. (a) Female RS; and (b) female THR (matched).

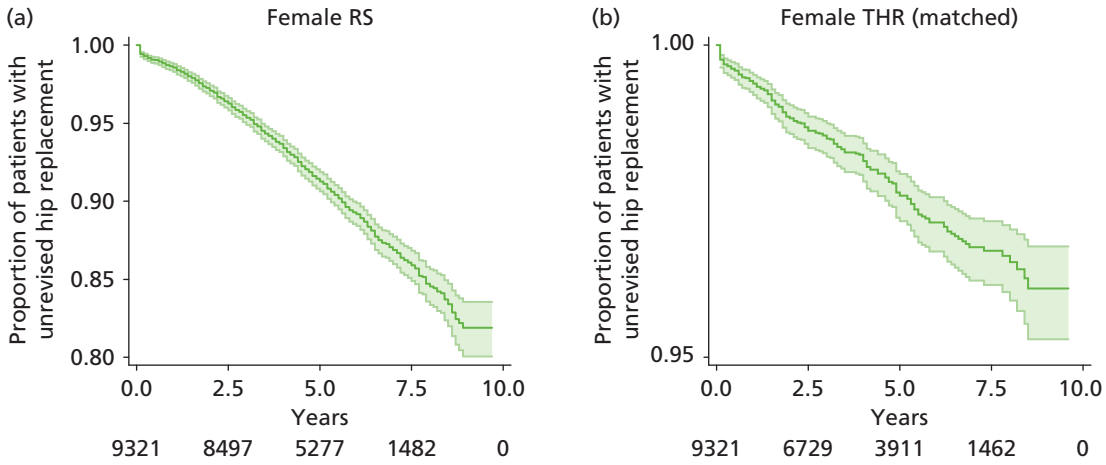


FIGURE 83 Observed revision rate for women receiving RS and THR (note difference in *y*-axis scales). (a) Female RS; and (b) female THR (matched). Numbers under *x*-axis represent patients at risk.

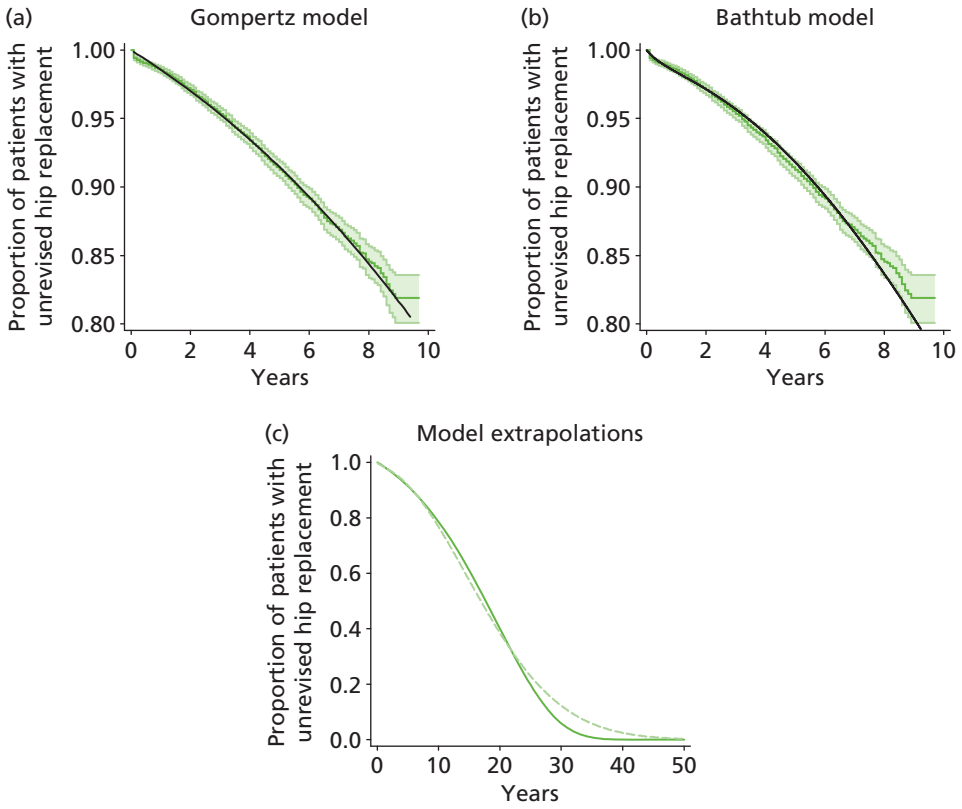


FIGURE 84 Observed revision (95% CI) for RS (women) with model fits and extrapolation. (a) Gompertz model; (b) bathtub model; and (c) model extrapolations.

For RS, AIC scores were lowest for the bathtub and Gompertz models (see *Appendix 19*); visually, there was little difference between these two. Cumulative hazard plots indicated a reasonable fit also for the Weibull and log-normal models (see *Appendix 20*). The bathtub, Gompertz and Weibull models all predicted an increasing hazard with extrapolation. The Gompertz distribution predicted that nearly all patients required revision of RS by ~30 years of follow-up. The bathtub fit was similar with a marginally lower rate of revision after about 20 years.

Table 138 lists the predicted percentage of patients requiring revision at 20 and 30 years post intervention for the Weibull, log-normal and bathtub models. All three models predict high rates of revision after RS.

TABLE 138 Predicted percentage of patients requiring revision at 20 and 30 years according to each model

Intervention	Model	20 years	30 years
RS	Log-normal	26.2	33.1
RS	Weibull	36.9	52.0
RS	Bathtub	62.3	87.5
THR	Log-normal	6.2	7.9
THR	Weibull	7.5	10.4
THR	Bathtub	14.2	26.9

Revision rates listed refer to the mean age of patients.

The revision rate for the matched THR female cohort was considerably less than that for RS (see *Figure 66*). Parametric fits to the observed revision rate for THR patients are shown in *Appendix 17*. Visually, the bathtub fit was as good as alternative models and was the only model that predicted increasing hazard beyond the observation period. According to AIC scores and cumulative hazard plots, differences were trivial between the bathtub, log-normal and Weibull models (see *Appendices 19* and *20*, respectively). *Table 138* lists the predicted percentage of patients requiring revision at 20 and 30 years according to best-fit models. The bathtub model predicted that approximately one-quarter of patients required revision after 30 years. As the bathtub model provided a good fit for both RS and THR, this model was used to estimate transition probabilities. In further sensitivity analysis log-normal models were used for each group.

Men

Of the 21,883 male patients who received a RS intervention for OA, 17,322 were successfully matched by age with male THR patients represented by categories A–E ($n = 87,271$). The age distribution in the matched groups is shown in *Figure 85*.

Figure 86 shows that the observed revision rate is far higher for RS than for THR. Parametric fits are presented in *Appendix 17*. For RS the bathtub distribution produced the lowest AIC scores and visually the superior fit (see *Appendices 17* and *19*); cumulative hazard plots indicated a good fit for the bathtub, log-normal and Weibull models (see *Appendix 20*). Apart from the bathtub model, the models predicted a decreasing hazard on extrapolation and low proportions of patients requiring revision (see *Appendix 17*). The different modelled predictions of the proportions requiring revision by 20 and 30 years are listed in *Table 139*.

For matched male THR patients the bathtub distribution provided a superior fit (see *Appendices 17* and *19*). According to cumulative hazard plots, the bathtub, Weibull and log-normal models produced good fits (see *Appendix 20*). Predicted revision rates at 20 and 30 years are listed in *Table 139*. The bathtub model predicted that about one-quarter of patients required revision of THR after 30 years (*Figure 87*) and was the only model predicting an increased hazard on extrapolation. For the economic analysis the bathtub model was adopted for both the RS and the THR groups. In sensitivity analysis log-normal models, which predicted a decreasing hazard on extrapolation, were used.

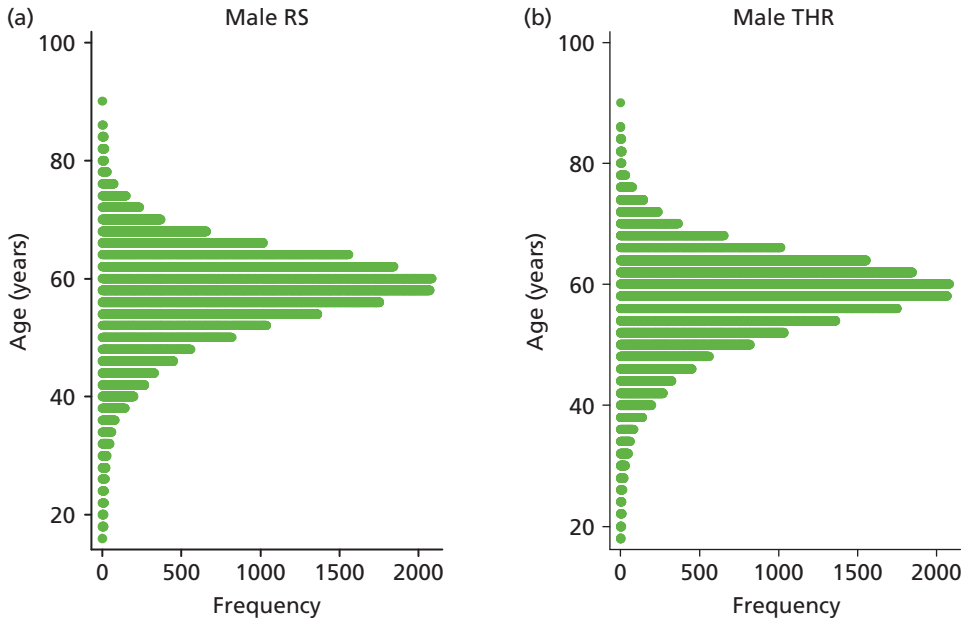


FIGURE 85 Age distribution in matched RS and THR male groups. (a) Male RS; and (b) male THR (matched).

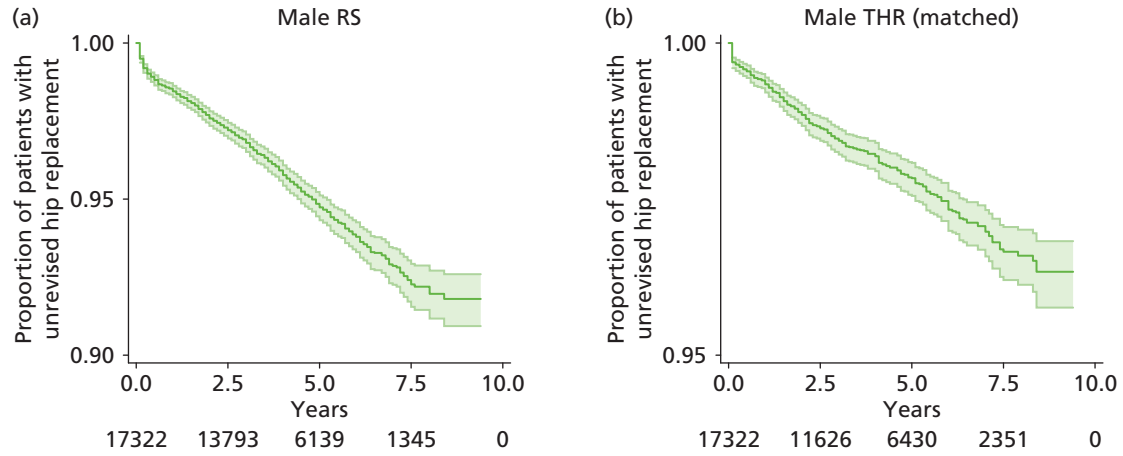


FIGURE 86 Observed revision rate for men receiving RS and THR (note the difference in the *y*-axis scales). (a) Male RS; and (b) male THR (matched).

TABLE 139 Predicted percentage of patients requiring revision at 20 and 30 years according to each model

Intervention	Model	20 years	30 years
RS	Log-normal	12.1	15.0
RS	Weibull	15.4	20.7
RS	Bathtub	35.9	61.5
THR	Log-normal	5.6	7.0
THR	Weibull	6.7	9.1
THR	Bathtub	13.4	25.6

Revision rates listed refer to the mean age of patients.

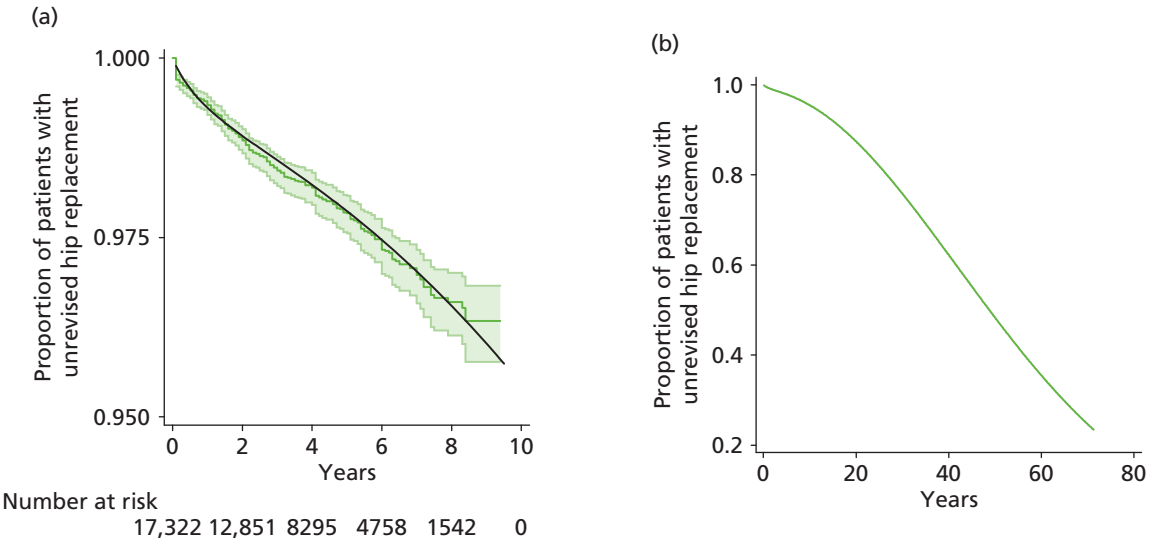


FIGURE 87 (a) Bathtub fit to observed time to revision for male THR patients; and (b) extrapolation beyond observed data.

Appendix 22 Comparison of total hip replacement revision rates according to sex and age

Comparison of total hip replacement revision rates: male patients aged > 65 years

Figure 88 shows the observed time to revision for male patients aged > 65 years according to category of THR prosthesis. Rates are broadly similar but CeCoP (category E) has a slightly lower rate and CeLMoP (category B) has a slightly higher rate. The shape of the Kaplan–Meier curve differs between categories.

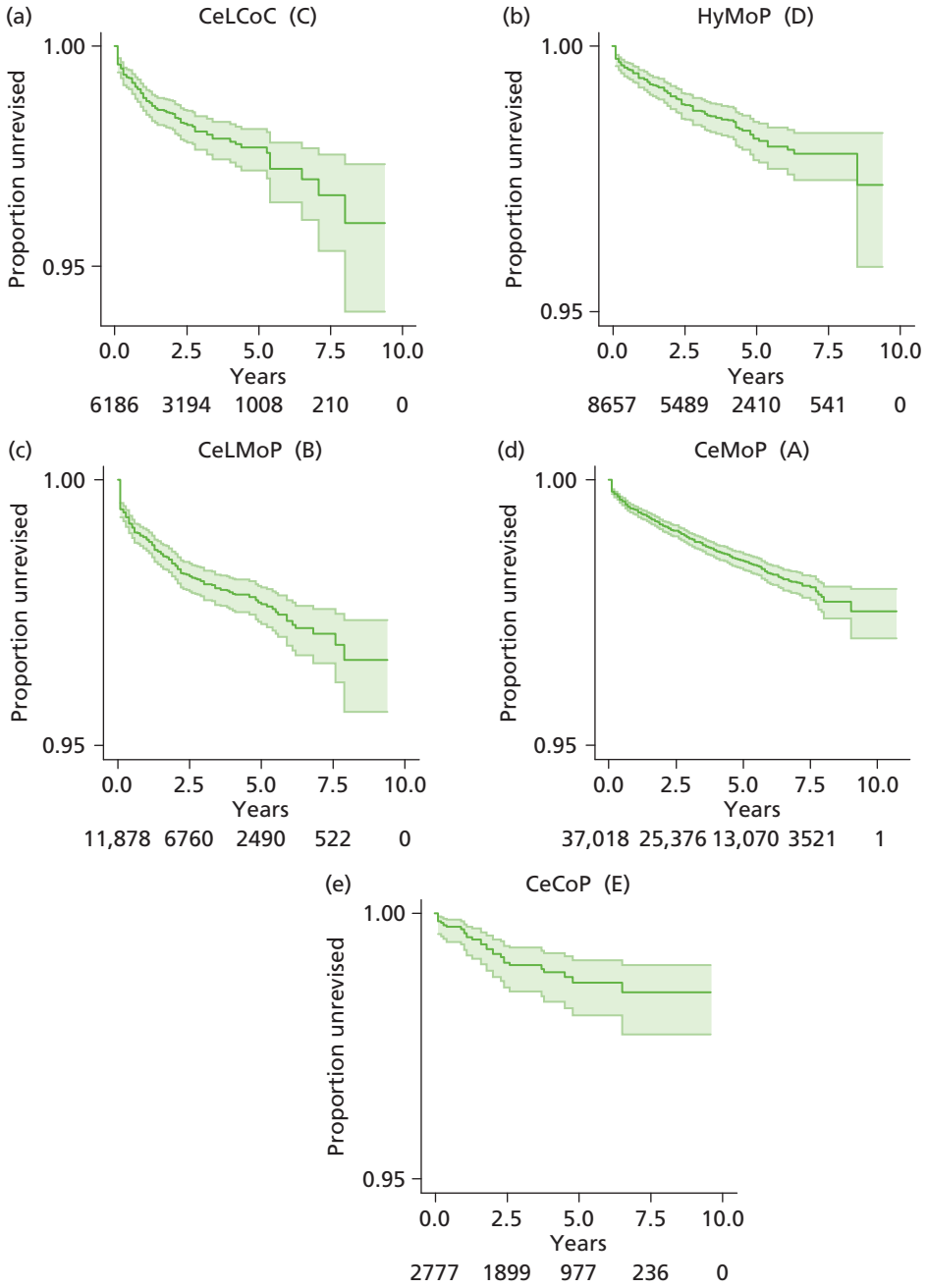


FIGURE 88 Kaplan–Meier plots: time to revision for THR categories A–E for men aged > 65 years. (a) CeLCoC (category C); (b) HyMoP (category D); CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E). Numbers under *x*-axis represents patients at risk.

Parametric fits to the observed data are shown in *Appendix 17*, AIC values for models in *Appendix 19* and diagnostic plots in *Appendix 20*.

Each parametric distribution other than the bathtub model predicted a decreasing hazard on extrapolation. Visually and by AIC scores, the bathtub and log-normal models generated best fits, except for the CeCoP (category E) prosthesis, for which the bathtub model did not resolve; the Gompertz model was marginally superior to the log-normal model for CeCoP (category E) but predicted zero revisions beyond 10 years and was considered unlikely to be realistic.

The 20- and 30-year predicted revision rates for the well-fitting models are summarised in *Table 140*. For the economic analysis, transition probabilities were based separately on the log-normal (base case) and bathtub (sensitivity analysis) models; for the latter the constant hazard model was used for CeCoP (category E) as the bathtub model did not resolve for these data.

Comparison of total hip replacement revision rates: female patients aged > 65 years

Figure 89 shows the observed time to revision for female patients aged > 65 years according to category of THR prosthesis. Rates are broadly similar with a relatively lower rate for CeCoP (category E) and a higher rate for CeLMoP (category B). The shape of the Kaplan–Meier curve differs between categories. Parametric fits to the observed data are shown in *Appendix 17*.

Other than the bathtub model, each parametric distribution predicted a decreasing hazard on extrapolation. For prosthesis categories CeLCoC (category C), HyMoP (category D), CeLMoP (category B) and CeMoP (category A) the bathtub model provided the best fit both visually and according to AIC score (see *Appendix 19*) and cumulative hazard plots (see *Appendix 20*), followed by the log-normal model for CeLCoC (category C), HyMoP (category D) and CeLMoP (category B) and the Weibull and log-logistic models for CeMoP (category A). For CeCoP (category E) the bathtub model failed to resolve and the best fit according to AIC score was provided by the Gompertz model followed by the log-normal model. The 20- and 30-year predicted revision rates for the well-fitting models are summarised in *Table 141*.

TABLE 140 Predicted revision rates for well-fitting models (% revised)

THR category	Model	20 years	30 years
CeLCoC (category C)	Log-normal	5.5	6.7
	Bathtub	9.6	17.4
HyMoP (category D)	Log-normal	3.7	4.6
	Bathtub	7.9	15.0
CeLMoP (category B)	Log-normal	4.9	5.9
	Bathtub	7.8	13.8
CeMoP (category A)	Log-normal	3.5	4.4
	Bathtub	7.2	13.9
CeCoP (category E)	Log-normal	2.9	3.6
	Gompertz	1.7	1.7
	Exponential	5.5	6.7

Revision rates listed refer to the mean age of patients in each of the THR categories.

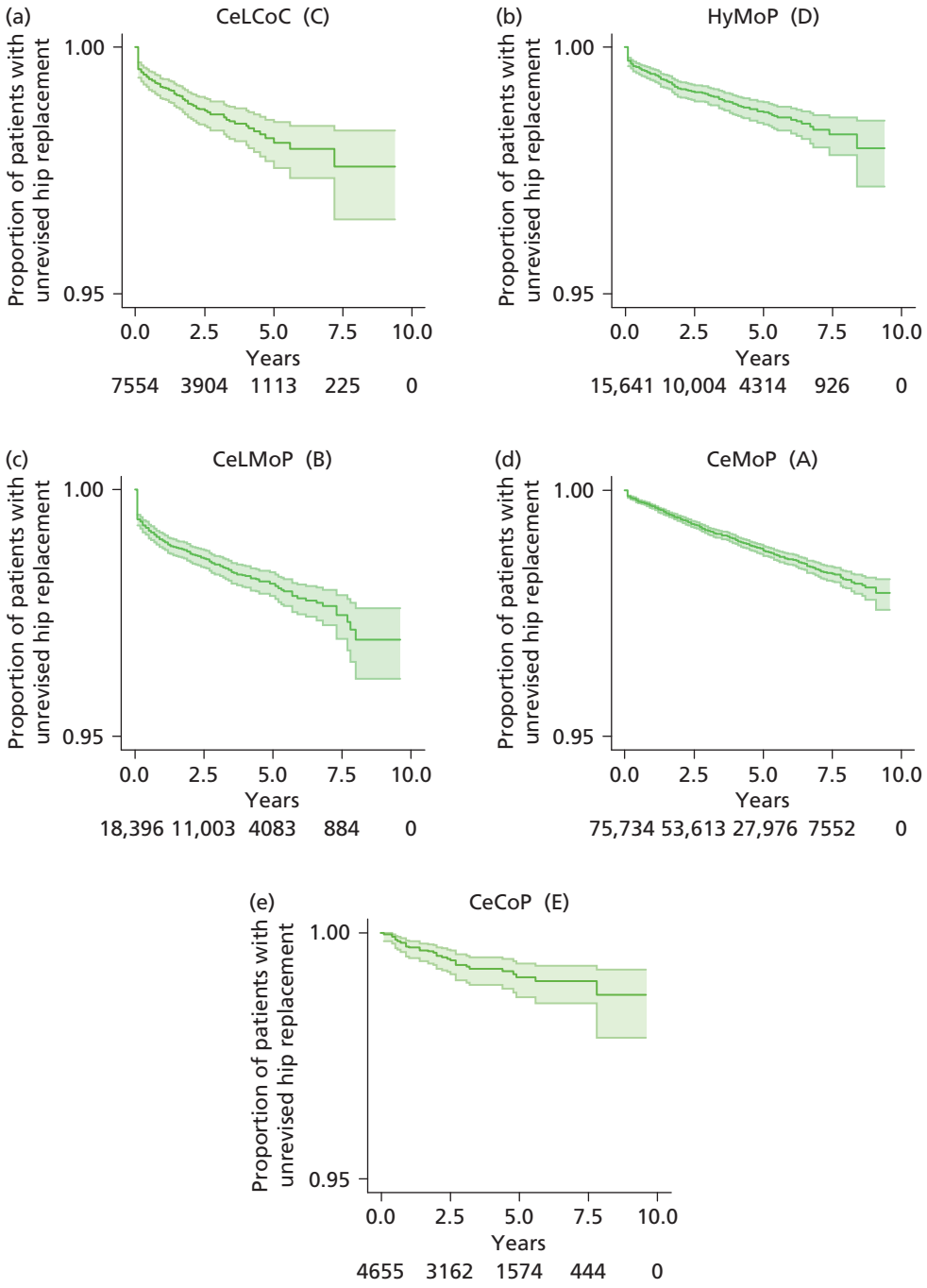


FIGURE 89 Kaplan–Meier plots: time to revision for THR categories A–E for women aged > 65 years. (a) CeLCoC (category C); (b) HyMoP (category D); CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E).

TABLE 141 Predicted revision rates for well-fitting models (% revised)

THR category	Model	20 years	30 years
CeLCoC (category C)	Log-normal	3.7	4.4
	Bathtub	8.4	15.8
HyMoP (category D)	Log-normal	2.7	3.3
	Bathtub	5.7	10.9
CeLMoP (category B)	Log-normal	3.8	4.5
	Bathtub	8.1	14.9
CeMoP (category A)	Log-normal	3.5	4.4
	Log-logistic	4.0	5.2
	Weibull	4.0	5.3
	Bathtub	6.4	12.5
CeCoP (category E)	Log-normal	2.3	3.0
	Gompertz	1.4	1.4

Revision rates listed refer to the mean age of patients in each of the THR categories.

For the economic analysis, transition probabilities were based separately on the log-normal (base case) and bathtub (sensitivity analysis) models; for the latter analysis the constant hazard model was used for CeCoP (category E) as the bathtub model did not resolve for this group.

Comparison of total hip replacement revision rates: male patients aged < 65 years

Figure 90 shows the observed time to revision for male patients aged < 65 years according to category of THR prosthesis. The number of patients in this group was much smaller than the number in the group aged > 65 years. Rates are broadly similar with a relatively lower rate for CeCoP (category E) and a higher rate for CeLMoP (category B). The shape of the Kaplan–Meier curve differs between categories. Parametric fits to the observed data are shown in *Appendix 17* and AIC values for models are summarised in *Appendix 19*. Cumulative hazard plots are shown in *Appendix 20*.

According to AIC values (and visually), the bathtub model provided a superior fit for CeLCoC (category C), HyMoP (category D) and CeLMoP (category B), followed by the log-normal model for CeLCoC (category C) and HyMoP (category D). AIC scores for the log-logistic and log-normal models for CeLMoP (category B) were almost equal and each model predicted a similar proportion of revisions after 20 and 30 years' follow-up. For CeMoP (category A) there were only trivial differences in AIC scores between the Weibull, log-normal, log-logistic and bathtub models. Only the bathtub model predicted an increasing hazard for CeLCoC (category C), HyMoP (category D), CeLMoP (category B) and CeMoP (category A) on extrapolation. For CeCoP (category E) the bathtub, Weibull and Gompertz models each predicted an increasing hazard on extrapolation whereas the exponential model provided the lowest AIC score. The 20- and 30-year predicted revision rates for the well-fitting models are summarised in *Table 142*.

Transition probabilities for economic analysis were based separately on the bathtub (base case) and log-normal (sensitivity analysis) models. In sensitivity analysis the exponential distribution was used for CeCoP (category E) in combination with the log-normal model for the other categories.

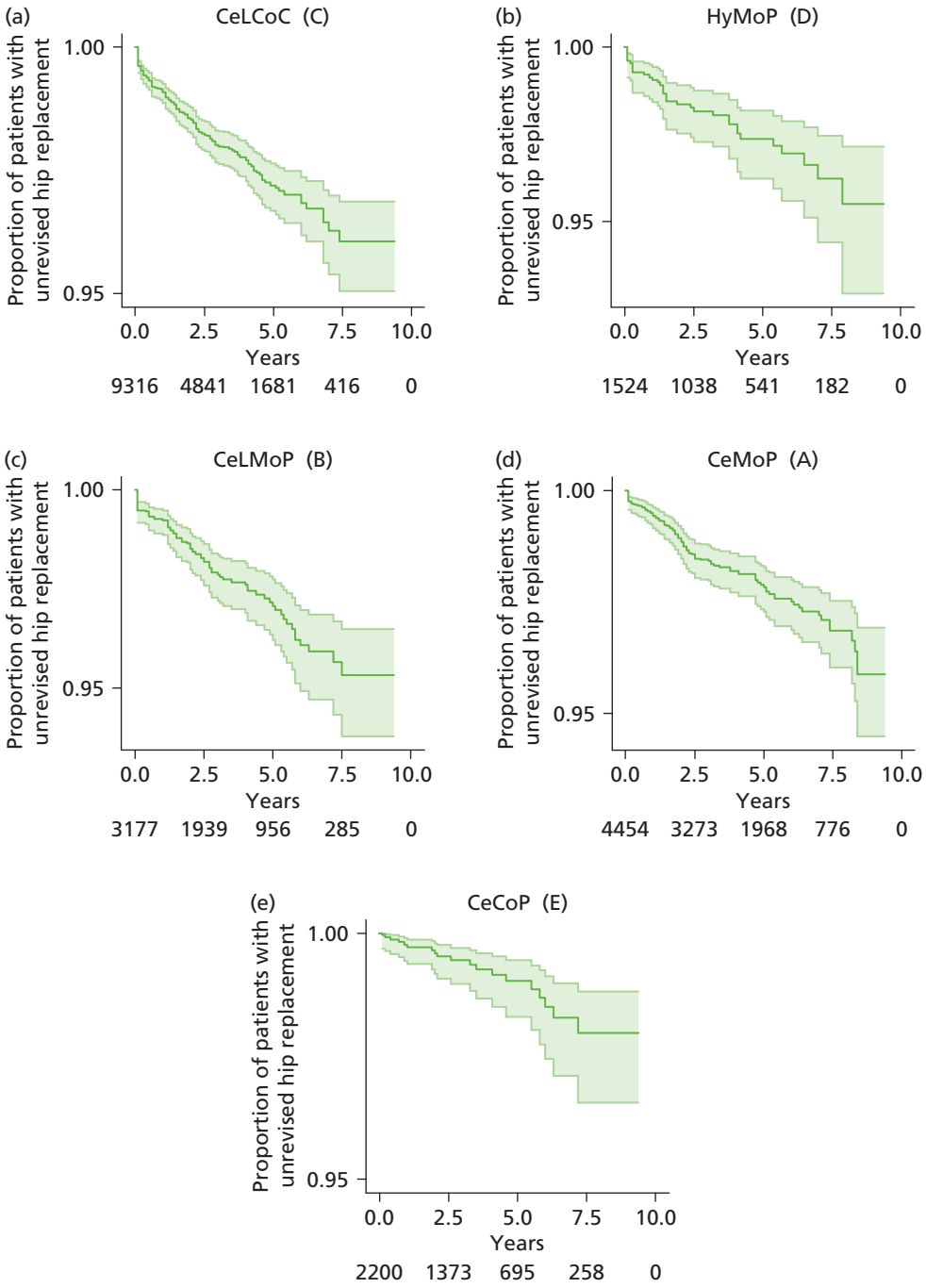


FIGURE 90 Kaplan–Meier plots: time to revision for THR categories A–E for men aged < 65 years. (a) CeLCoC (category C); (b) HyMoP (category D); CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E).

TABLE 142 Predicted revision rates for well-fitting models (% revised)

THR category	Model	20 years	30 years
CeLCoC (category C)	Log-normal	6.1	7.5
	Bathtub	14.4	27.0
HyMoP (category D)	Log-normal	6.2	7.6
	Bathtub	14.0	26.2
CeLMoP (category B)	Log-normal	7.1	8.9
	Bathtub	20.9	39.2
CeMoP (category A)	Log-normal	5.8	7.3
	Bathtub	10.3	19.0
CeCoP (category E)	Log-normal	3.7	5.0
	Bathtub	9.6	19.7
	Exponential	3.7	5.0

Revision rates listed refer to the mean age of patients in each of the THR categories.

Comparison of total hip replacement revision rates: female patients aged < 65 years

Figure 91 shows the observed time to revision for female patients aged < 65 years according to category of THR prosthesis. The number of patients was small relative to the number in the group aged > 65 years. Rates are broadly similar with a somewhat lower rate for CeCoP (category E); however, after 5 years, when the data are associated with considerable uncertainty, the CeCoP (category E) revision rate increases.

According to AIC scores the bathtub model provided the better fit for CeLCoC (category C), HyMoP (category D) and CeMoP (category A) but failed to resolve for CeLMoP (category B) (see *Appendix 19*). For CeCoP (category E) the exponential and bathtub models produced the lowest AIC scores. For the models that generated a decreasing hazard on extrapolation, the log-normal model was superior with regard to AIC values for CeLCoC (category C), CeLMoP (category B) and CeCoP (category E) whereas the Weibull model was better for HyMoP (category D) and CeMoP (category A). The 20- and 30-year predicted revision rates for the well-fitting models are summarised in *Table 143*. Transition probabilities for economic analysis were based separately on the bathtub (base case) and Weibull (sensitivity analysis) models; for the former the exponential model was used for CeLMoP (category B) as the bathtub model did not resolve for this category.

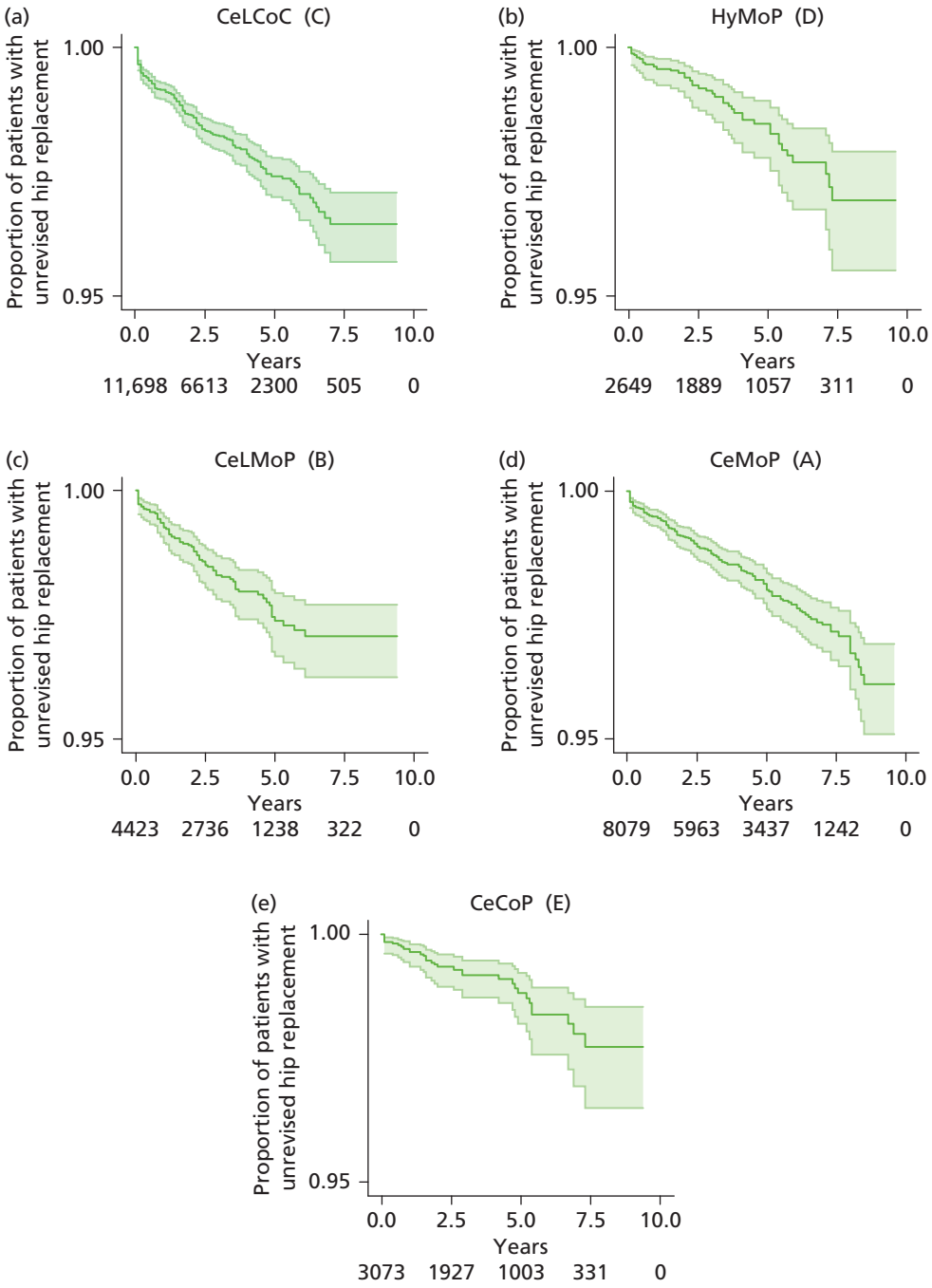


FIGURE 91 Kaplan–Meier plots: time to revision for THR categories A–E for women aged < 65 years. (a) CeLCoC (category C); (b) HyMoP (category D); CeLMoP (category B); (d) CeMoP (category A); and (e) CeCoP (category E).

TABLE 143 Predicted revision rates for well-fitting models (% revised)

THR category	Model	20 years	30 years
CeLCoC (category C)	Log-normal	5.5	6.8
	Bathtub	13.6	26.1
HyMoP (category D)	Log-normal	5.3	7.0
	Weibull	6.8	10.1
	Bathtub	15.0	29.8
CeLMoP (category B)	Log-normal	5.6	7.1
	Exponential	9.4	13.8
CeMoP (category A)	Log-normal	5.4	6.8
	Weibull	6.5	9.0
	Bathtub	13.9	27.3
CeCoP (category E)	Log-normal	3.6	4.6
	Gompertz	5.2	7.7
	Bathtub	9.8	19.9

Revision rates listed refer to the mean age of patients in each of the THR categories.

Appendix 23 Flexible parametric fits to observed time to revision

The following figures compare the extrapolation of flexible parametric fits to observed time to revision data with other parametric models.

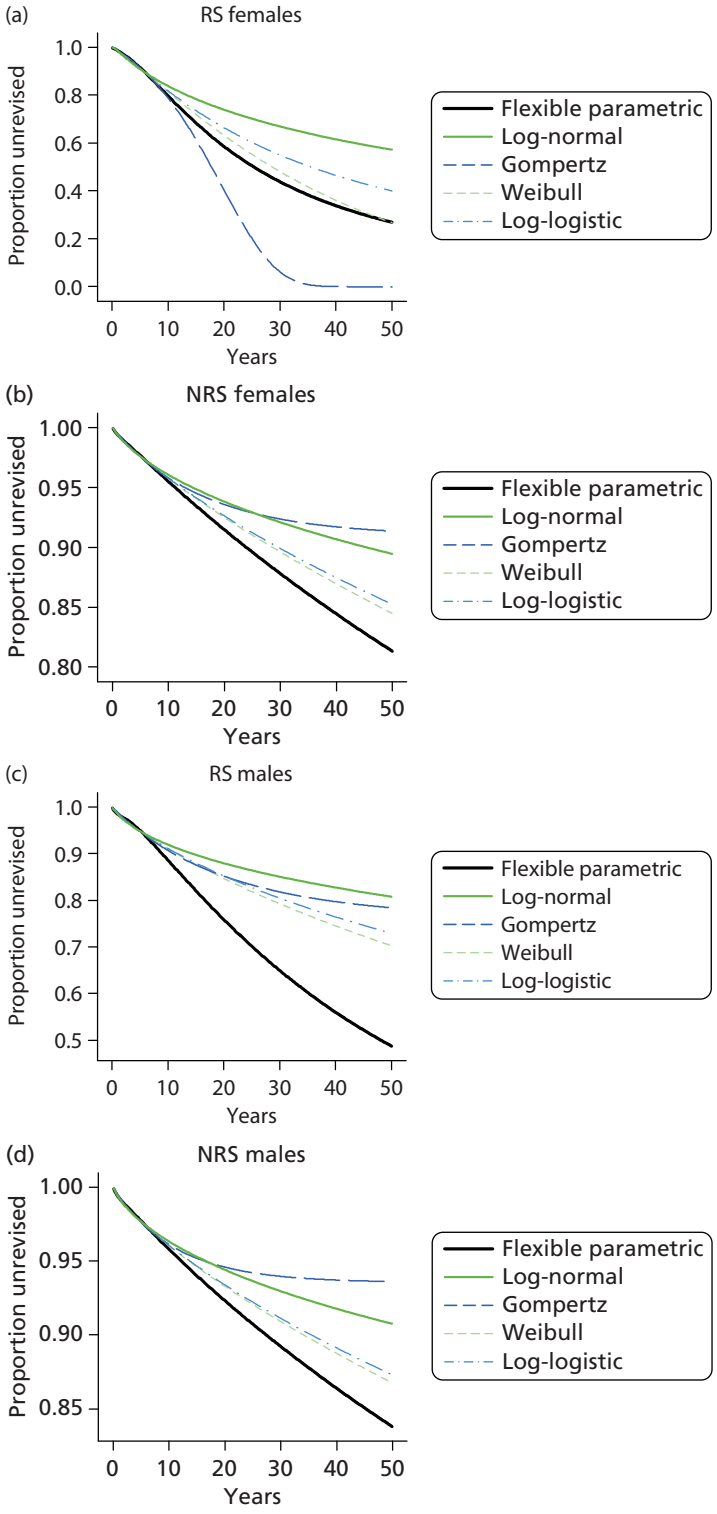


FIGURE 92 Resurfacing arthroplasty and non-RS, men and women. (a) RS females; (b) non-RS females; (c) RS males; and (d) non-RS males.

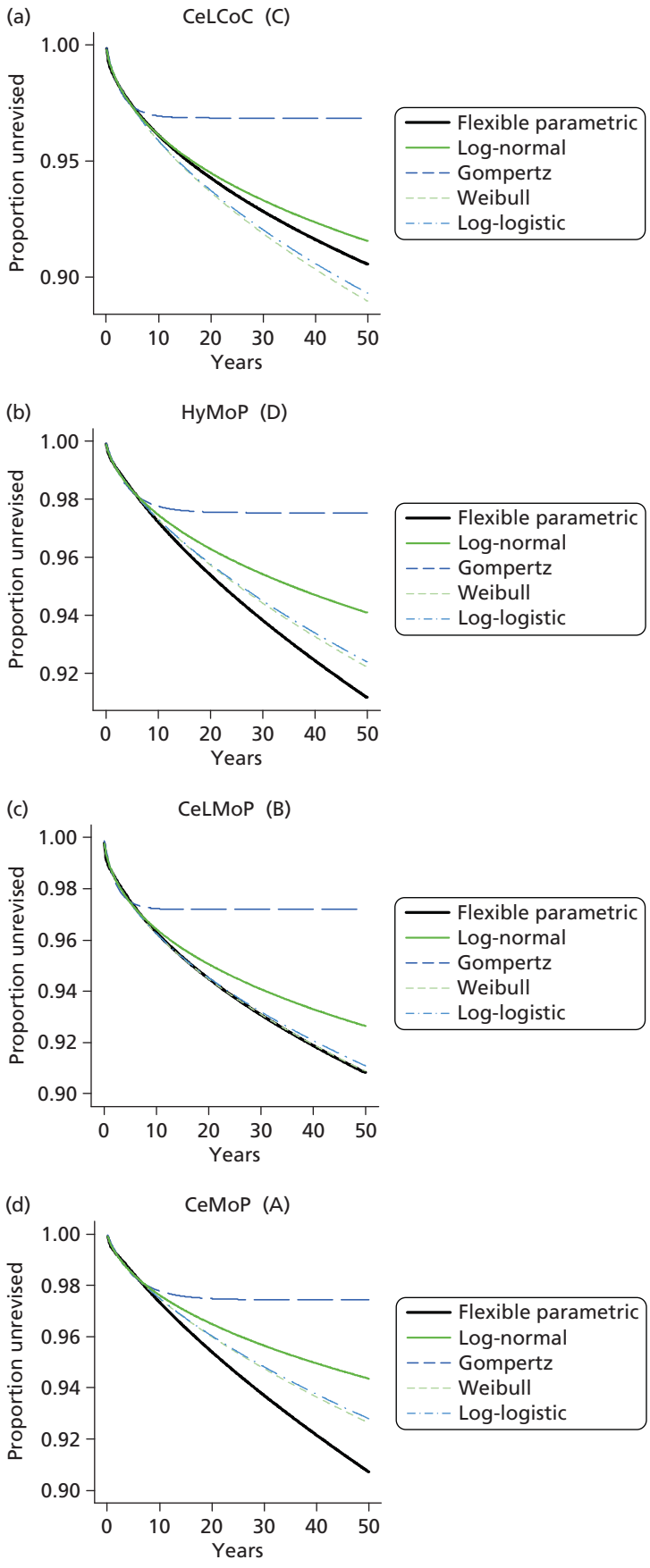


FIGURE 93 Total hip replacement comparisons: men (a–e) and women (f–j) aged > 65 years. (*continued*)

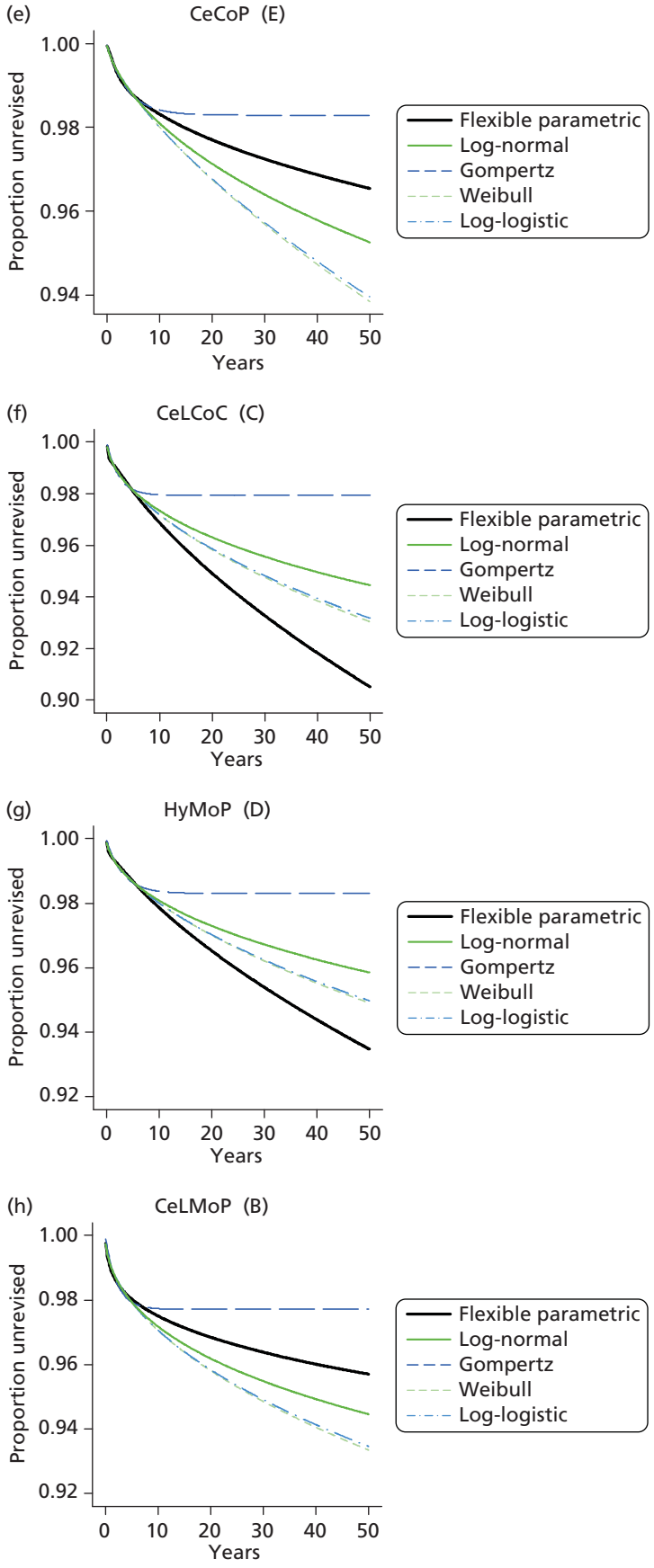


FIGURE 93 Total hip replacement comparisons: men (a–e) and women (f–j) aged > 65 years. (*continued*)

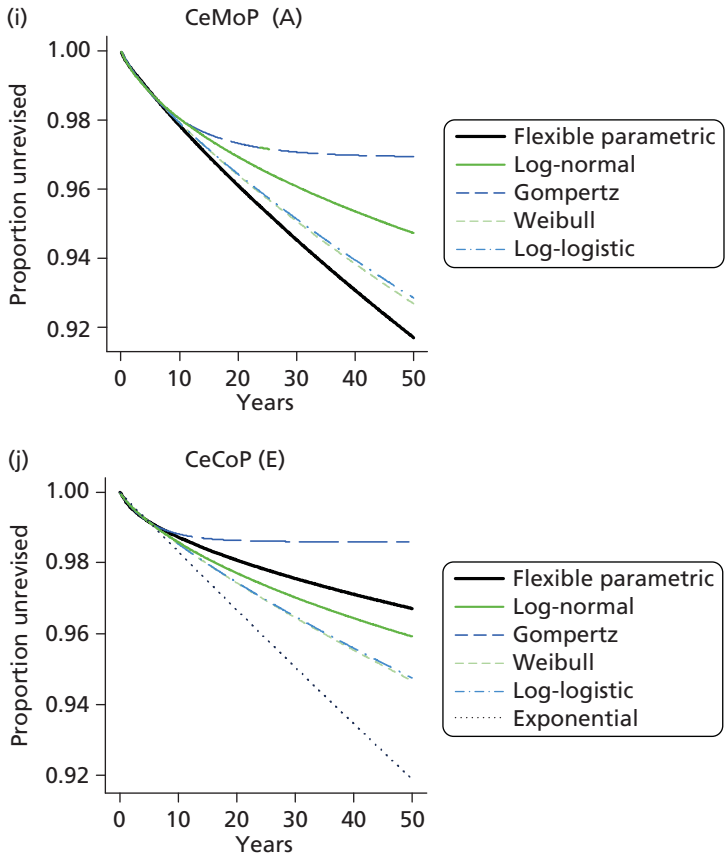


FIGURE 93 Total hip replacement comparisons: men (a–e) and women (f–j) aged > 65 years.

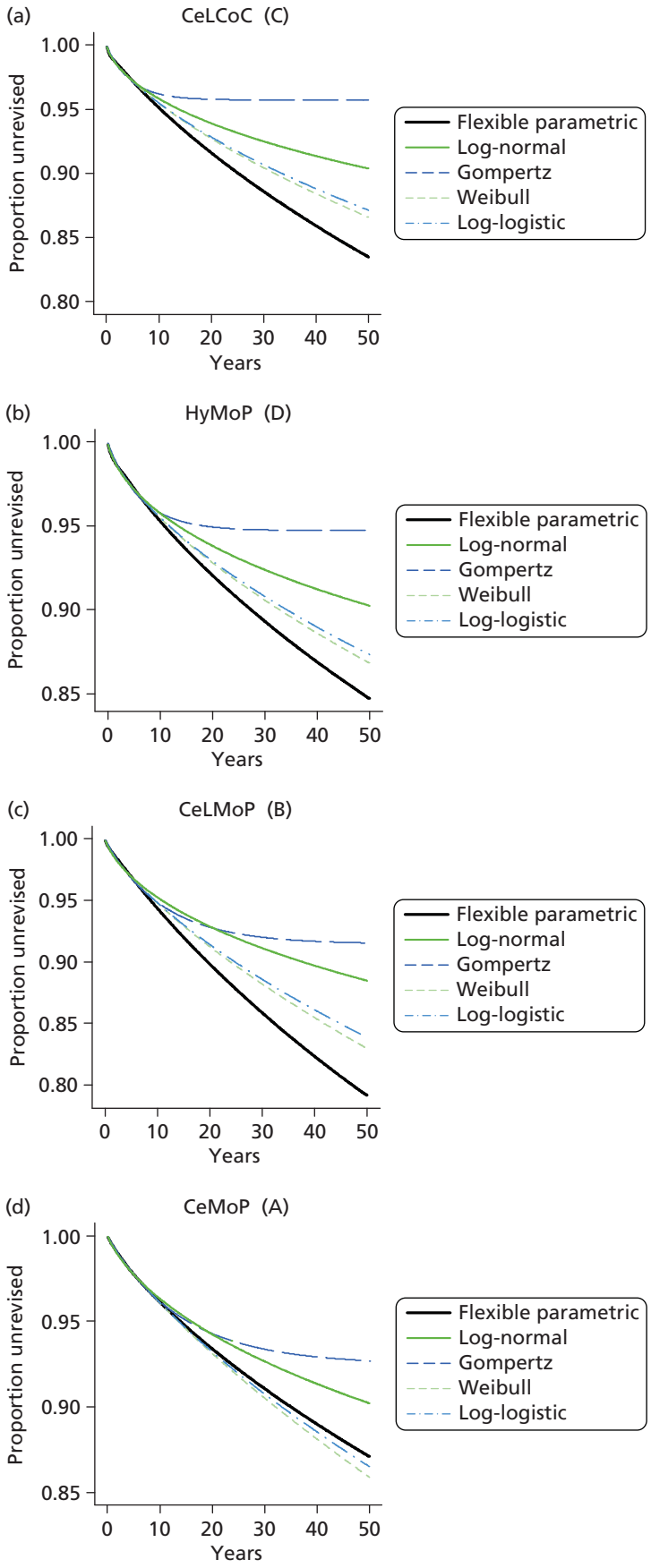


FIGURE 94 Total hip replacement comparisons: men (a–e) and women (f–j) aged < 65 years. (*continued*)

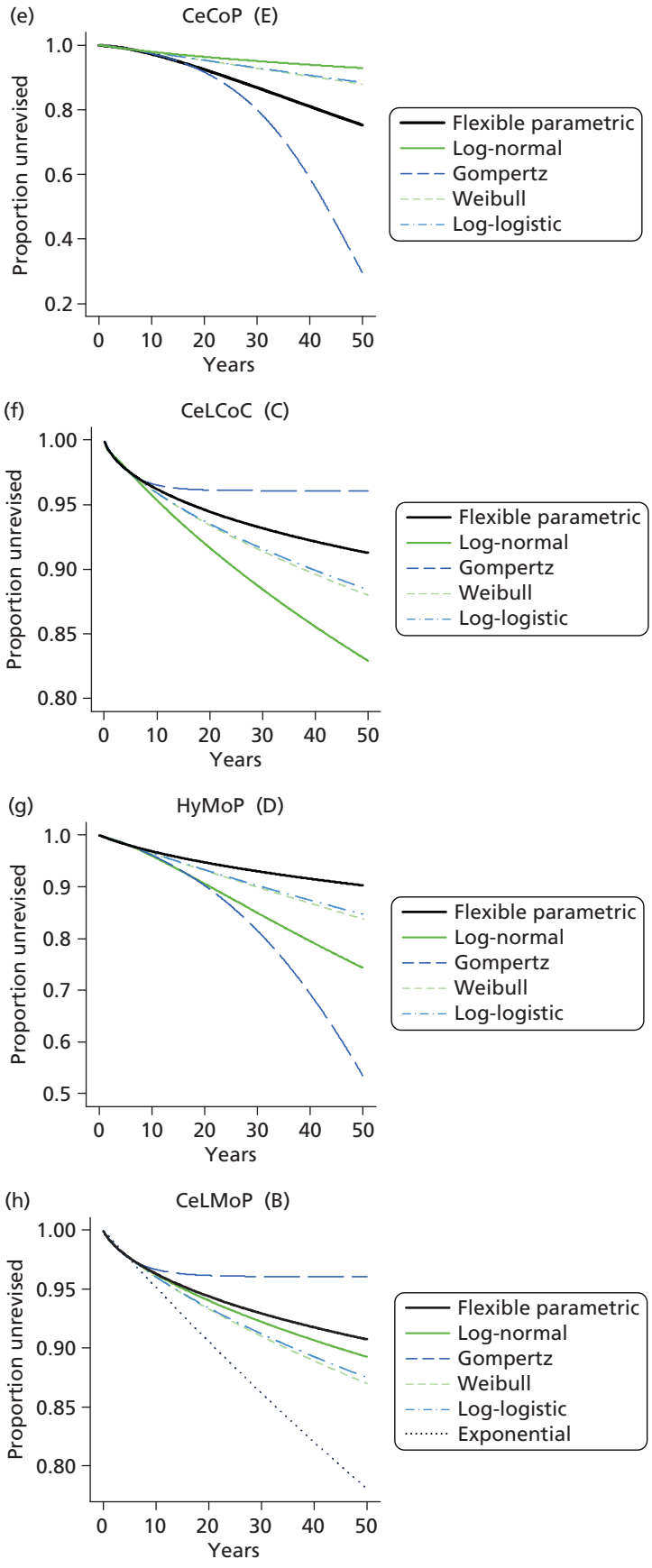


FIGURE 94 Total hip replacement comparisons: men (a–e) and women (f–j) aged < 65 years. (*continued*)

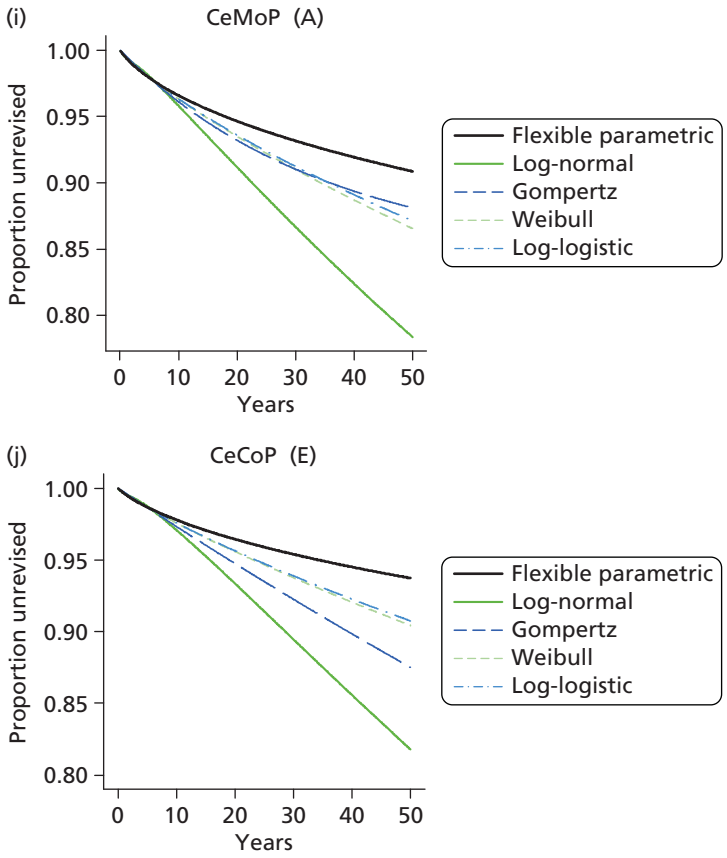


FIGURE 94 Total hip replacement comparisons: men (a–e) and women (f–j) aged < 65 years.

Akaike information criterion scores for flexible parametric fits: resurfacing arthroplasty compared with total hip replacement

Category	Model	Observations	Likelihood model	Parameters	AIC	BIC
RS female	Exponential	9321	–3832.57	2	7669.146	7683.426
RS female	Weibull	9321	–3822.48	3	7650.955	7672.375
RS female	Gompertz	9321	–3810.47	3	7626.936	7648.357
RS female	Log-normal	9321	–3871.01	3	7748.02	7769.44
RS female	Log-logistic	9321	–3826.6	3	7659.199	7680.619
RS female	Flexible	9321	–3805.76	5	7621.52	7657.22
THR female	Exponential	9321	–1175.68	2	2355.366	2369.646
THR female	Weibull	9321	–1171.86	3	2349.722	2371.142
THR female	Gompertz	9321	–1174.19	3	2354.382	2375.802
THR female	Log-normal	9321	–1172.61	3	2351.221	2372.641
THR female	Log-logistic	9321	–1171.9	3	2349.793	2371.213
THR female	Flexible	9321	–1171.13	5	2352.257	2387.957
RS male	Exponential	17,322	–4156.28	2	8316.549	8332.069
RS male	Weibull	17,322	–4136.86	3	8279.725	8303.004
RS male	Gompertz	17,322	–4153.83	3	8313.657	8336.936
RS male	Log-normal	17,322	–4154.04	3	8314.077	8337.356
RS male	Log-logistic	17,322	–4138.8	3	8283.599	8306.879
RS male	Flexible	17,322	–4107.94	5	8225.879	8264.678
THR male	Exponential	17,322	–2051.92	2	4107.83	4123.349
THR male	Weibull	17,322	–2038.25	3	4082.5	4105.779
THR male	Gompertz	17,322	–2046.58	3	4099.169	4122.448
THR male	Log-normal	17,322	–2040.2	3	4086.392	4109.671
THR male	Log-logistic	17,322	–2038.37	3	4082.738	4106.017
THR male	Flexible	17,322	–2036.62	5	4083.234	4122.032

Men aged > 65 years

Category	Model	Observations	Likelihood model	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	6186	–746.636	2	1497.273	1510.733
CeLCoC (category C)	Weibull	6186	–729.575	3	1465.15	1485.34
CeLCoC (category C)	Gompertz	6186	–732.489	3	1470.978	1491.168
CeLCoC (category C)	Log-normal	6186	–726.414	3	1458.828	1479.018
CeLCoC (category C)	Log-logistic	6186	–729.471	3	1464.942	1485.132
CeLCoC (category C)	Flexible	6186	–722.322	5	1454.645	1488.295
HyMoP (category D)	Exponential	8657	–759.606	2	1523.211	1537.343
HyMoP (category D)	Weibull	8657	–747.895	3	1501.79	1522.989
HyMoP (category D)	Gompertz	8657	–750.81	3	1507.62	1528.818
HyMoP (category D)	Log-normal	8657	–746.392	3	1498.783	1519.982
HyMoP (category D)	Log-logistic	8657	–747.866	3	1501.732	1522.93
HyMoP (category D)	Flexible	8657	–745.547	5	1501.094	1536.425
CeLMoP (category B)	Exponential	11,878	–1506.67	2	3017.336	3032.101
CeLMoP (category B)	Weibull	11,878	–1455.29	3	2916.575	2938.723
CeLMoP (category B)	Gompertz	11,878	–1465.04	3	2936.071	2958.218
CeLMoP (category B)	Log-normal	11,878	–1448.28	3	2902.553	2924.701
CeLMoP (category B)	Log-logistic	11,878	–1455.05	3	2916.099	2938.247
CeLMoP (category B)	Flexible	11,878	–1434.02	5	2878.038	2914.951
CeMoP (category A)	Exponential	37,018	–3248.63	2	6501.262	6518.303
CeMoP (category A)	Weibull	37,018	–3201.27	3	6408.546	6434.107
CeMoP (category A)	Gompertz	37,018	–3218.14	3	6442.275	6467.836
CeMoP (category A)	Log-normal	37,018	–3196.16	3	6398.314	6423.874
CeMoP (category A)	Log-logistic	37,018	–3201.21	3	6408.426	6433.987
CeMoP (category A)	Flexible	37,018	–3187.21	5	6384.421	6427.022
CeCoP (category E)	Exponential	2777	–193.01	2	390.0204	401.8786
CeCoP (category E)	Weibull	2777	–190.888	3	387.7762	405.5636
CeCoP (category E)	Gompertz	2777	–190.268	3	386.5367	404.3241
CeCoP (category E)	Log-normal	2777	–190.442	3	386.8841	404.6715
CeCoP (category E)	Log-logistic	2777	–190.874	3	387.7485	405.5359
CeCoP (category E)	Flexible	2777	–190.082	5	390.163	419.8087

Men aged < 65 years

Category	Model	Observations	Likelihood model	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	9316	–1127.92	2	2259.84	2274.119
CeLCoC (category C)	Weibull	9316	–1112.13	3	2230.269	2251.687
CeLCoC (category C)	Gompertz	9316	–1117.29	3	2240.588	2262.007
CeLCoC (category C)	Log-normal	9316	–1110.07	3	2226.137	2247.555
CeLCoC (category C)	Log-logistic	9316	–1112.1	3	2230.189	2251.608
CeLCoC (category C)	Flexible	9316	–1106.75	5	2223.497	2259.194
HyMoP (category D)	Exponential	1524	–219.219	2	442.4374	453.0956
HyMoP (category D)	Weibull	1524	–215.918	3	437.8354	453.8226
HyMoP (category D)	Gompertz	1524	–217.603	3	441.2066	457.1939
HyMoP (category D)	Log-normal	1524	–215.623	3	437.2452	453.2325
HyMoP (category D)	Log-logistic	1524	–215.919	3	437.8372	453.8245
HyMoP (category D)	Flexible	1524	–215.273	5	440.5452	467.1906
CeLMoP (category B)	Exponential	3177	–447.905	2	899.8102	911.9376
CeLMoP (category B)	Weibull	3177	–444.514	3	895.027	913.218
CeLMoP (category B)	Gompertz	3177	–446.802	3	899.6039	917.795
CeLMoP (category B)	Log-normal	3177	–445.224	3	896.4473	914.6384
CeLMoP (category B)	Log-logistic	3177	–444.562	3	895.1241	913.3152
CeLMoP (category B)	Flexible	3177	–444.095	5	898.1896	928.5081
CeMoP (category A)	Exponential	4454	–553.723	2	1111.446	1124.268
CeMoP (category A)	Weibull	4454	–551.57	3	1109.14	1128.373
CeMoP (category A)	Gompertz	4454	–552.736	3	1111.471	1130.704
CeMoP (category A)	Log-normal	4454	–552.155	3	1110.309	1129.542
CeMoP (category A)	Log-logistic	4454	–551.605	3	1109.209	1128.442
CeMoP (category A)	Flexible	4454	–551.55	5	1113.099	1145.154
CeCoP (category E)	Exponential	2200	–121.877	2	247.7535	259.1459
CeCoP (category E)	Weibull	2200	–121.819	3	249.6385	266.7271
CeCoP (category E)	Gompertz	2200	–121.591	3	249.1827	266.2713
CeCoP (category E)	Log-normal	2200	–122.521	3	251.0416	268.1303
CeCoP (category E)	Log-logistic	2200	–121.835	3	249.6702	266.7589
CeCoP (category E)	Flexible	2200	–121.274	5	252.5475	281.0285

Women aged > 65 years

Category	Model	Observations	Likelihood model	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	7554	–708.148	2	1420.295	1434.155
CeLCoC (category C)	Weibull	7554	–685.957	3	1377.914	1398.704
CeLCoC (category C)	Gompertz	7554	–692.052	3	1390.104	1410.893
CeLCoC (category C)	Log-normal	7554	–683.416	3	1372.831	1393.62
CeLCoC (category C)	Log-logistic	7554	–685.903	3	1377.805	1398.595
CeLCoC (category C)	Flexible	7554	–671.673	5	1353.345	1387.995
HyMoP (category D)	Exponential	15,641	–1200.15	2	2404.291	2419.606
HyMoP (category D)	Weibull	15,641	–1167.92	3	2341.831	2364.804
HyMoP (category D)	Gompertz	15,641	–1178.97	3	2363.936	2386.909
HyMoP (category D)	Log-normal	15,641	–1165.25	3	2336.508	2359.481
HyMoP (category D)	Log-logistic	15,641	–1167.88	3	2341.762	2364.735
HyMoP (category D)	Flexible	15,641	–1155.5	5	2320.993	2359.282
CeLMoP (category B)	Exponential	18,396	–2076.22	2	4156.445	4172.085
CeLMoP (category B)	Weibull	18,396	–1983.4	3	3972.806	3996.265
CeLMoP (category B)	Gompertz	18,396	–2016.98	3	4039.962	4063.421
CeLMoP (category B)	Log-normal	18,396	–1975	3	3956.004	3979.464
CeLMoP (category B)	Log-logistic	18,396	–1983.21	3	3972.418	3995.878
CeLMoP (category B)	Flexible	18,396	–1970.01	4	3948.018	3979.298
CeMoP (category A)	Exponential	75,734	–5258.73	2	10,521.46	10,539.93
CeMoP (category A)	Weibull	75,734	–5231.8	3	10,469.59	10,497.3
CeMoP (category A)	Gompertz	75,734	–5243.93	3	10,493.85	10,521.56
CeMoP (category A)	Log-normal	75,734	–5233.88	3	10,473.75	10,501.46
CeMoP (category A)	Log-logistic	75,734	–5231.91	3	10,469.82	10,497.52
CeMoP (category A)	Flexible	75,734	–5230.44	5	10,470.89	10,517.06
CeCoP (category E)	Exponential	4655	–231.057	2	466.1135	479.0049
CeCoP (category E)	Weibull	4655	–230.171	3	466.3423	485.6794
CeCoP (category E)	Gompertz	4655	–229.251	3	464.5019	483.839
CeCoP (category E)	Log-normal	4655	–229.665	3	465.3301	484.6672
CeCoP (category E)	Log-logistic	4655	–230.161	3	466.321	485.6581
CeCoP (category E)	Flexible	4655	–228.751	5	467.5019	499.7303

Women aged < 65 years

Category	Model	Observations	Likelihood model	Parameters	AIC	BIC
CeLCoC (category C)	Exponential	11,698	–1366.3	2	2736.599	2751.334
CeLCoC (category C)	Weibull	11,698	–1345.8	3	2697.598	2719.7
CeLCoC (category C)	Gompertz	11,698	–1353.72	3	2713.435	2735.537
CeLCoC (category C)	Log-normal	11,698	–1343.32	3	2692.636	2714.738
CeLCoC (category C)	Log-logistic	11,698	–1345.76	3	2697.523	2719.625
CeLCoC (category C)	Flexible	11,698	–1336.36	5	2682.723	2719.559
HyMoP (category D)	Exponential	2649	–240.911	2	485.8222	497.586
HyMoP (category D)	Weibull	2649	–240.911	3	487.8221	505.468
HyMoP (category D)	Gompertz	2649	–240.717	3	487.4333	505.0791
HyMoP (category D)	Log-normal	2649	–241.85	3	489.6996	507.3454
HyMoP (category D)	Log-logistic	2649	–240.941	3	487.8809	505.5267
HyMoP (category D)	Flexible	2649	–240.189	5	490.377	519.7867
CeLMoP (category B)	Exponential	4423	–495.874	2	995.7488	1008.538
CeLMoP (category B)	Weibull	4423	–492.013	3	990.0253	1009.209
CeLMoP (category B)	Gompertz	4423	–491.714	3	989.4287	1008.612
CeLMoP (category B)	Log-normal	4423	–491.035	3	988.0695	1007.253
CeLMoP (category B)	Log-logistic	4423	–491.964	3	989.9272	1009.111
CeLMoP (category B)	Flexible	4423	–491.749	5	993.4976	1025.47
CeMoP (category A)	Exponential	8079	–949.689	2	1903.377	1917.409
CeMoP (category A)	Weibull	8079	–946.435	3	1898.87	1919.918
CeMoP (category A)	Gompertz	8079	–949.457	3	1904.915	1925.962
CeMoP (category A)	Log-normal	8079	–948.231	3	1902.463	1923.51
CeMoP (category A)	Log-logistic	8079	–946.524	3	1899.048	1920.096
CeMoP (category A)	Flexible	8079	–899.444	5	1808.888	1843.873
CeCoP (category E)	Exponential	3073	–209.334	2	422.6671	434.7279
CeCoP (category E)	Weibull	3073	–208.952	3	423.9041	441.9954
CeCoP (category E)	Gompertz	3073	–209.334	3	424.667	442.7582
CeCoP (category E)	Log-normal	3073	–209.509	3	425.0178	443.1091
CeCoP (category E)	Log-logistic	3073	–208.972	3	423.9445	442.0357
CeCoP (category E)	Flexible	3073	–208.059	5	426.1171	456.2692

Appendix 24 Revision rates in studies with extended follow-up

The RCT by Kim *et al.*[126] reported an extended follow-up to about 20 years. The reported revision rates were higher between 15 and 20 years than between 10 and 15 years.

The bathtub model of hazard for revision implies that revision rates will gradually increase at some time after plateauing. The trajectory of revision during substantial follow-up beyond the plateau is required to see if this is reasonable; however, it should be borne in mind that long follow-up times (e.g. up to 20 years) necessitate looking at devices and practices that may now no longer be widely used. The NJR data provided observed rates to between 9 and 10 years only; we looked elsewhere for relevant information on the trajectory of revision rates over the long term.

The SHAR provides relevant information for up to 19 years of follow-up (*Figure 95*). This shows increasing rates of revision from about 5 to 15 years for age groups other than 60- to 75-year-old patients. For these age groups these data are consistent with a bathtub hazard. For the oldest age group revision rates are relatively low and are not consistent with the bathtub model.

Some small observational studies[368–370] with long follow-up times also support the bathtub model. The results from these are summarised in *Figure 96*.

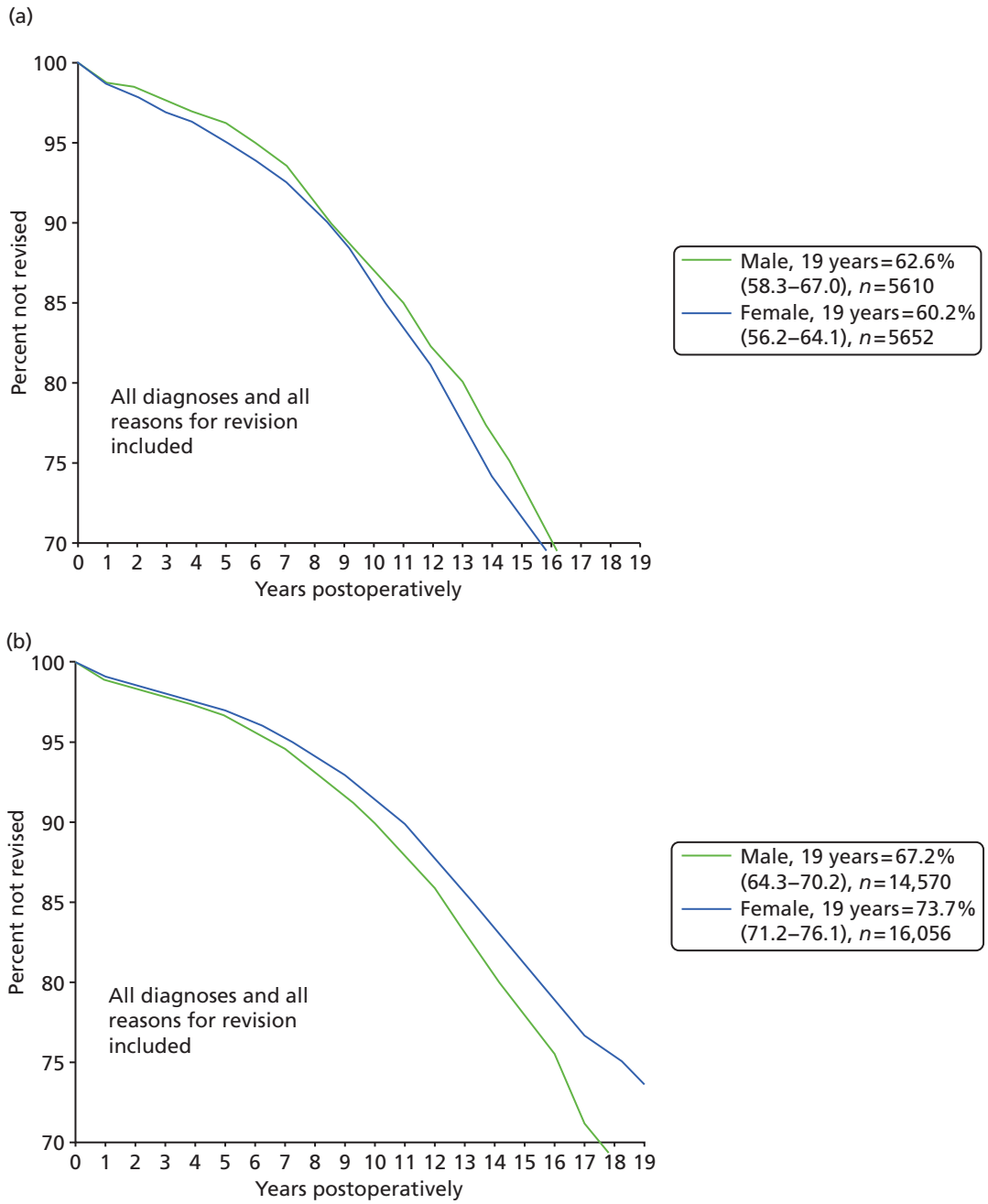


FIGURE 95 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010). (a) Patients aged < 50 years; (b) patients aged between 50 and 59 years; (c) patients aged between 60 and 75 years; and (d) patients aged > 75 years. Reproduced from Garellick *et al.*[96] and permission granted by the Swedish Hip Arthroplasty Register. (*continued*)

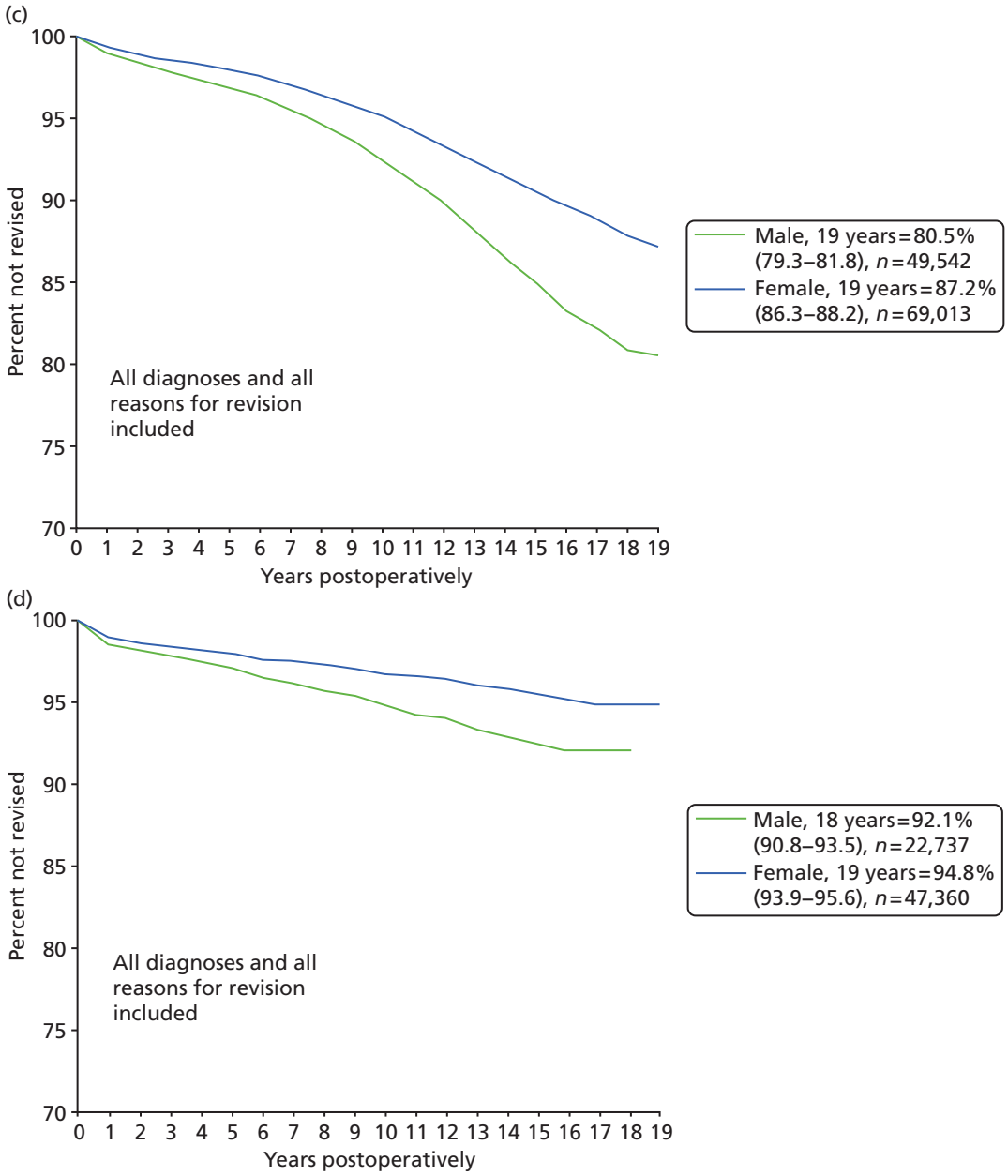


FIGURE 95 Swedish registry data for time to revision up to 19 years of follow-up (1992–2010). (a) Patients aged < 50 years; (b) patients aged between 50 and 59 years; (c) patients aged between 60 and 75 years; and (d) patients aged > 75 years. Reproduced from Garellick *et al.*[96] and permission granted by the Swedish Hip Arthroplasty Register.

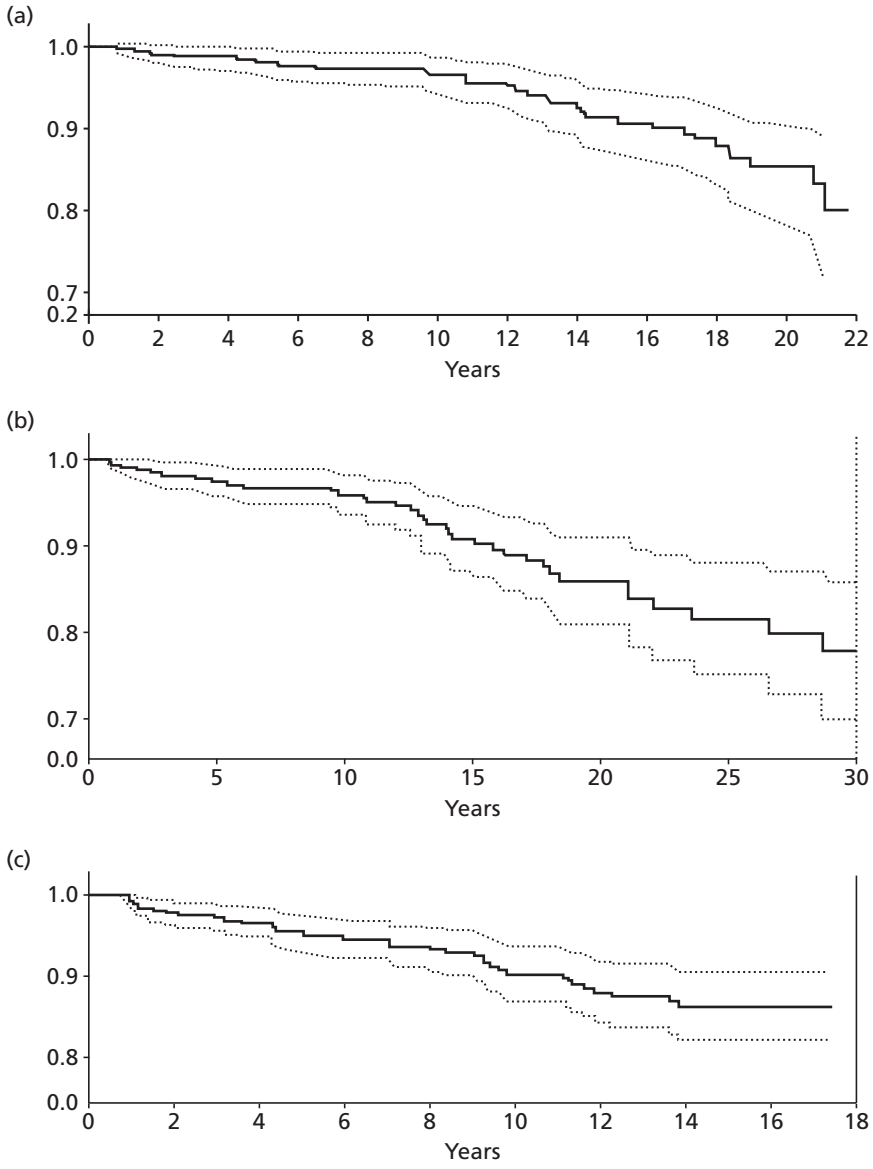


FIGURE 96 Time to revision results reproduced with permission from (a) Schulte *et al.*;[368] (b) Madey *et al.*;[369] and (c) Callaghan *et al.*[370]

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library